

ACADEMIC PRESS ADVANCED FINANCE SERIES



SECOND EDITION

INTERMEDIATE FINANCIAL THEORY

**Jean-Pierre Danthine
and John B. Donaldson**

List of Frequently Used Symbols and Notation

A text such as *Intermediate Financial Theory* is, by nature, relatively notation intensive. We have adopted a strategy to minimize the notational burden within each individual chapter at the cost of being, at times, inconsistent in our use of symbols across chapters. We list here a set of symbols regularly used with their specific meaning. At times, however, we have found it more practical to use some of the listed symbols to represent a different concept. In other instances, clarity required making the symbolic representation more precise (e.g., by being more specific as to the time dimension of an interest rate).

Roman Alphabet

a	Amount invested in the risky asset; in Chapter 14, fraction of wealth invested in the risky asset or portfolio
A^T	Transpose of the matrix (or vector) A
c	Consumption; in Chapter 14 only, consumption is represented by C , while c represents $\ln C$
c_θ^k	Consumption of agent k in state of nature θ
CE	Certainty equivalent
C_A	Price of an American call option
C_E	Price of a European call option
d	Dividend rate or amount
Δ	Number of shares in the replicating portfolio (Chapter xx)
E	The expectations operator
e_θ^k	Endowment of agent k in state of nature θ
f	Futures position (Chapter 16);
p^f	Price of a futures contract (Chapter 16)
F, G	Cumulative distribution functions associated with densities:
f, g	Probability density functions
K	The strike or exercise price of an option
$K(\tilde{x})$	Kurtosis of the random variable \tilde{x}
L	A lottery
\mathcal{L}	Lagrangian
m	Pricing kernel
M	The market portfolio
MU_θ^k	Marginal utility of agent k in state θ
p	Price of an arbitrary asset
\mathbf{P}	Measure of Absolute Prudence
q	Arrow-Debreu price
q^b	Price of risk-free discount bond, occasionally denoted p^{rf}
q^e	Price of equity
r_f	Rate of return on a risk-free asset
R_f	Gross rate of return on a risk-free asset
\tilde{r}	Rate of return on a risky asset
\tilde{R}	Gross rate of return on a risky asset

R_A	Absolute risk aversion coefficient
R_R	Relative risk aversion coefficient
s	Usually denotes the amount saved
S	In the context of discussing options, used to denote the price of the underlying stock
$S(\tilde{x})$	Skewness of the random variable \tilde{x}
\mathbf{T}	Transition matrix
U	Utility function
\mathcal{U}	von Neuman-Morgenstern utility function
V	Usually denotes variance-covariance matrix of asset returns; occasionally is used as another utility function symbol; may also signify value as in
V_P	The value of portfolio P or
V_F	The value of the firm
w_i	Portfolio weight of asset i in a given portfolio
Y_0	Initial wealth

Greek Alphabet

α	Intercept coefficient in the market model (alpha)
β	The slope coefficient in the market model (beta)
δ	Time discount factor
η	Elasticity
λ	Lagrange multiplier
μ	Mean
π_θ	State probability of state θ
π_θ^{RN}	Risk-neutral probability of state θ
Π	Risk premium
$\rho(\tilde{x}, \tilde{y})$	Correlation of random variables \tilde{x} and \tilde{y}
ρ	Elasticity of intertemporal substitution (Chapter 14)
σ	Standard deviation
σ_{ij}	Covariance between random variables i and j
θ	Index for state of nature
Ω	Rate of depreciation of physical capital
ψ	Compensating precautionary premium

Numerals and Other Terms

$\mathbf{1}$	Vector of ones
\succ	Is strictly preferred to
\succeq	Is preferred to (non strictly, that is allowing for indifference)
GBM	Geometric Brownian Motion stochastic process
FSD	First-order stochastic dominance
SSD	Second-order stochastic dominance

Preface

The market for financial textbooks is crowded at both the introductory and doctoral levels, but much thinner at the intermediate level. Teaching opportunities at this level, however, have greatly increased with the advent of masters of science programs in finance (master in computational finance, in mathematical finance, and the like) and the strengthening demand for higher-level courses in MBA programs.

The Master in Banking and Finance Program at the University of Lausanne admitted its first class in the fall of 1993. One of the first such programs of its kind in Europe, its objective was to provide advanced training to finance specialists in the context of a one-year theory-based degree program. In designing the curriculum, it was felt that students should be exposed to an integrated course that would introduce the range of topics typically covered in financial economics courses at the doctoral level. Such exposure could, however, ignore the detailed proofs and arguments and concentrate on the larger set of issues and concepts to which any advanced practitioner should be exposed.

Our ambition in this text is, accordingly, first to review rigorously and concisely the main themes of financial economics (those that students should have encountered in prior courses) and, second, to introduce a number of *frontier* ideas of importance for the evolution of the discipline and of relevance from a practitioner's perspective. We want our readers not only to be at ease with the main concepts of standard finance (MPT, CAPM, etc.) but also to be aware of the principal new ideas that have marked the recent evolution of the discipline. Contrary to introductory texts, we aim at depth and rigor; contrary to higher-level texts, we do not emphasize generality. Whenever an idea can be conveyed through an example, this is the approach we chose. We have, similarly, ignored proofs and detailed technical matters unless a reasonable understanding of the related concept mandated their inclusion.

Intermediate Financial Theory is intended primarily for master level students with a professional orientation, a good quantitative background, and a preliminary education in business and finance. As such, the book is targeted for masters students in finance, but it is also appropriate for an advanced MBA class in financial economics, one with the objective of introducing students to the precise modeling of many of the concepts discussed in their capital markets and corporate finance classes. In addition, we believe the book will be a useful reference for entering doctoral candidates in finance whose lack of prior background might prevent them from drawing the full benefits of the very abstract material typically covered at that level. Finally, it is a useful refresher for well-trained practitioners.

As far as prerequisites go, we take the view that our readers will have completed at least one introductory course in Finance (or have read the corresponding text) and will not be intimidated by mathematical formalism. Although the mathematical requirements of the book are not large, some confidence in the use of calculus as well as matrix algebra is helpful.

In preparing the second edition of this text, we have emphasized the overrid-

ing concern of modern finance for the valuation of risky cash flows: Intermediate Financial Theory's main focus is thus on asset pricing. (In addition, we exclusively consider discrete time methodologies). The new Chapter 2 makes clear this emphasis while simultaneously stressing that asset pricing does not represent the totality of modern finance. This discussion then leads to a new structuring of the book into five parts, and a new ordering of the various chapters. Our goal here is to make a sharper distinction between valuation approaches that rely on equilibrium principles and those based on arbitrage considerations. We have also reorganized the treatment of Arrow-Debreu pricing to make clear how it accommodates both perspectives. Finally, a new chapter entitled "Portfolio Management in the Long Run" is included that covers recent developments that we view as especially relevant for the contemporary portfolio manager. The two appendices providing brief overviews of option pricing and continuous time valuation methods are now assigned to the text website.

Over the years, we have benefited from numerous discussions with colleagues over issues related to the material included in this book. We are especially grateful to Paolo Siconolfi, Columbia University, Rajnish Mehra, U. of California at Santa Barbara, and Erwan Morellec, University of Lausanne, the latter for his contribution to the corporate finance review of Chapter 2. We are also indebted to several generations of teaching assistants – François Christen, Philippe Gilliard, Tomas Hricko, Aydin Akgun, Paul Ehling, Oleksandra Hubal and Lukas Schmid – and of MBF students at the University of Lausanne who have participated in the shaping-up of this material. Their questions, corrections and comments have lead to a continuous questioning of the approach we have adopted and have dramatically increased the usefulness of this text. Finally we reiterate our thanks to the Fondation du 450ème of the University of Lausanne for providing "seed financing" for this project.

Jean-Pierre Danthine,
Lausanne, Switzerland

John B. Donaldson
New York City

*N'estime l'argent ni plus ni moins qu'il ne vaut :
c'est un bon serviteur et un mauvais maître
(Value money neither more nor less than it is worth:
It is a good servant and a bad master)*

Alexandre Dumas, fils, *La Dame aux Camélias* (Préface)

Part I

Introduction

Chapter 1 : On the Role of Financial Markets and Institutions

1.1 Finance: The Time Dimension

Why do we need financial markets and institutions? We chose to address this question as our introduction to this text on financial theory. In doing so we touch on some of the most difficult issues in finance and introduce concepts that will eventually require extensive development. Our purpose here is to phrase this question as an appropriate background for the study of the more technical issues that will occupy us at length. We also want to introduce some important elements of the necessary terminology. We ask the reader's patience as most of the sometimes-difficult material introduced here will be taken up in more detail in the following chapters.

A financial system is a set of institutions and markets permitting the exchange of contracts and the provision of services for the purpose of allowing the income and consumption streams of economic agents to be desynchronized – that is, made less similar. It can, in fact, be argued that indeed the *primary* function of the financial system is to permit such desynchronization. There are two dimensions to this function: the time dimension and the risk dimension. Let us start with time. Why is it useful to dissociate consumption and income across time? Two reasons come immediately to mind. First, and somewhat trivially, income is typically received at discrete dates, say monthly, while it is customary to wish to consume continuously (i.e., every day).

Second, and more importantly, consumption spending defines a *standard of living* and most individuals find it difficult to alter their standard of living from month to month or even from year to year. There is a general, if not universal, desire for a *smooth* consumption stream. Because it deeply affects everyone, the most important manifestation of this desire is the need to save (consumption smaller than income) for retirement so as to permit a consumption stream in excess of income (dissaving) after retirement begins. The *lifecycle* patterns of income generation and consumption spending are not identical, and the latter must be created from the former. The same considerations apply to shorter horizons. Seasonal patterns of consumption and income, for example, need not be identical. Certain individuals (car salespersons, department store salespersons) may experience variations in income arising from seasonal events (e.g., most new cars are purchased in the spring and summer), which they do not like to see transmitted to their ability to consume. There is also the problem created by temporary layoffs due to business cycle fluctuations. While temporarily laid off and without substantial income, workers do not want their family's consumption to be severely reduced.

Box 1.1

Representing Preference for Smoothness

The preference for a smooth consumption stream has a natural counterpart in the form of the utility function, $U(\cdot)$, typically used to represent the relative

benefit a consumer receives from a specific consumption bundle. Suppose the representative individual consumes a single consumption good (or a basket of goods) in each of two periods, now and tomorrow. Let c_1 denote today's consumption level and c_2 tomorrow's, and let $U(c_1) + U(c_2)$ represent the level of utility (benefit) obtained from a given consumption stream (c_1, c_2) .

Preference for consumption smoothness must mean, for instance, that the consumption stream $(c_1, c_2) = (4, 4)$ is preferred to the alternative $(c_1, c_2) = (3, 5)$, or

$$U(4) + U(4) > U(3) + U(5),$$

Dividing both sides of the inequality by 2, this implies

$$U(4) > \frac{1}{2}U(3) + \frac{1}{2}U(5).$$

As shown in Figure 1.1, when generalized to all possible alternative consumption pairs, this property implies that the function $U(\cdot)$ has the rounded shape that we associate with the term "strict concavity." \square

Insert Figure 1.1 about here

Furthermore, and this is quite crucial for the growth process, some people - entrepreneurs, in particular - are willing to accept a relatively small income (but not consumption!) for a period of time in exchange for the prospect of high returns (and presumably high income) in the future. They are operating a sort of 'arbitrage' over time. This does not disprove their desire for smooth consumption; rather they see opportunities that lead them to accept what is formally a low income level initially, against the prospect of a higher income level later (followed by a zero income level when they retire). They are investors who, typically, do not have enough liquid assets to finance their projects and, as a result, need to raise capital by borrowing or by selling shares.

Therefore, the first key element in finance is **time**. In a timeless world, there would be no assets, no financial transactions (although money would be used, it would have only a transaction function), and no financial markets or institutions. The very notion of a (financial) contract implies a time dimension.

Asset holding permits the desynchronization of consumption and income streams. The peasant putting aside seeds, the miser burying his gold, or the grandmother putting a few hundred dollar bills under her mattress are all desynchronizing their consumption and income, and in doing so, presumably seeking a higher level of well-being for themselves. A fully developed financial system should also have the property of fulfilling this same function **efficiently**. By that we mean that the financial system should provide versatile and diverse instruments to accommodate the widely differing needs of savers and borrowers in so far as size (many small lenders, a few big borrowers), timing and maturity of loans (how to finance long-term projects with short-term money), and

the liquidity characteristics of instruments (precautionary saving cannot be tied up permanently). In other words, the elements composing the financial system should aim at *matching* as perfectly as possible the diverse financing needs of different economic agents.

1.2 Desynchronization: The Risk Dimension

We argued above that time is of the essence in finance. When we talk of the importance of time in economic decisions, we think in particular of the relevance of choices involving the present versus the future. But the future is, by essence, uncertain: Financial decisions with implications (payouts) in the future are necessarily risky. Time and risk are inseparable. This is why **risk** is the second key word in finance.

For the moment let us compress the time dimension into the setting of a “Now and Then” (present vs. future) economy. The typical individual is motivated by the desire to smooth consumption between “Now” and “Then.” This implies a desire to identify consumption opportunities that are as smooth as possible among the different possibilities that may arise “Then.” In other words, *ceteris paribus* – most individuals would like to guarantee their family the same standard of living whatever events transpire tomorrow: whether they are sick or healthy; unemployed or working; confronted with bright or poor investment opportunities; fortunate or hit by unfavorable accidental events.¹ This characteristic of preferences is generally described as “aversion to risk.”

A productive way to start thinking about this issue is to introduce the notion of *states of nature*. A state of nature is a complete description of a possible scenario for the future across all the dimensions relevant for the problem at hand. In a “Now and Then” economy, all possible future events can be represented by an exhaustive list of states of nature or *states of the world*. We can thus extend our former argument for smoothing consumption across time by noting that the typical individual would also like to experience similar consumption levels across all future states of nature, whether good or bad.

An efficient financial system offers ways for savers to reduce or eliminate, at a fair price, the risks they are not willing to bear (risk shifting). Fire insurance contracts eliminate the financial risk of fire, while put contracts can prevent the loss in wealth associated with a stock’s price declining below a predetermined level, to mention two examples. The financial system also makes it possible to obtain relatively safe aggregate returns from a large number of small, relatively risky investments. This is the process of diversification. By permitting economic agents to *diversify*, to *insure*, and to *hedge* their risks, an efficient financial system fulfills the function of redistributing purchasing power not only over time, but also across states of nature.

Box 1.2 Representing Risk Aversion

¹“*Ceteris paribus*” is the latin expression for “everything else maintained equal”. It is part of the common language in economics.

Let us reinterpret the two-date consumption stream (c_1, c_2) of Box 1.1 as the consumption levels attained “Then” or “Tomorrow” in two alternative, equally likely, states of the world. The desire for a smooth consumption stream across the two states, which we associate with risk aversion, is obviously represented by the same inequality

$$U(4) > \frac{1}{2}U(3) + \frac{1}{2}U(5),$$

and it implies the same general shape for the utility function. In other words, assuming plausibly that decision makers are **risk averse**, an assumption in conformity with most of financial theory, implies that the utility functions used to represent agents’ preferences are **strictly concave**. \square

1.3 The Screening and Monitoring Functions of the Financial System

The business of desynchronizing consumption from income streams across time and states of nature is often more complex than our initial description may suggest. If time implies uncertainty, uncertainty may imply not only risk, but often *asymmetric information* as well. By this term, we mean situations where the individuals involved have different information, with some being potentially better informed than others. How can a saver be assured that he will be able to find a borrower with a good ability to repay - the borrower himself knows more about this, but he may not wish to reveal all he knows -, or an investor with a good project, yielding the most attractive return for him and hopefully for society as well? Again, the investor is likely to have a better understanding of the project’s prospects and of his own motivation to carry it through. What do “good” and “most attractive” mean in these circumstances? Do these terms refer to the highest potential return? What about risk? What if the anticipated return is itself affected by the actions of the investors themselves (a phenomenon labeled “moral hazard”)? How does one share the risks of a project in such a way that both investors and savers are willing to proceed, taking actions acceptable to both? An efficient financial system not only assists in these information and monitoring tasks, but also provides a range of instruments (contractual arrangements) suitable for the largest number of savers and borrowers, thereby contributing to the channeling of savings toward the most efficient projects.

In the terms of the preeminent economist, Joseph Schumpeter (1961) “*Bankers are the gatekeepers of capitalist economic development. Their strategic function is to screen potential innovators and advance the necessary purchasing power to the most promising.*” For highly risky projects, such as the creation of a new firm exploiting a new technology, venture capitalists provide a similar function today.

1.4 The Financial System and Economic Growth

The performance of the financial system matters at several levels. We shall argue that it matters for growth, that it impacts the characteristics of the business cycle, and most importantly, that it is a significant determinant of economic welfare. We tackle growth first. Channeling funds from savers to investors efficiently is obviously important. Whenever more efficient ways are found to perform this task, society can achieve a greater increase in tomorrow's consumption for a given sacrifice in current consumption.

Intuitively, more savings should lead to greater investment and thus greater future wealth. Figure 1.2 indeed suggests that, for 90 developing countries over the period 1971 to 1992, there was a strong positive association between saving rates and growth rates. When looked at more carefully, however, the evidence is usually not as strong.² One important reason may be that the hypothesized link is, of course, dependent on a *ceteris paribus* clause: It applies only to the extent savings are invested in appropriate ways. The economic performance of the former Union of Soviet Socialist Republics reminds us that it is not enough only to save; it is also important to invest judiciously. Historically, the investment/GDP (Gross Domestic Product) ratio in the Soviet Union was very high in international comparisons, suggesting the potential for very high growth rates. After 1989, however, experts realized that the value of the existing stock of capital was not consistent with the former levels of investment. A great deal of the investment must have been effectively wasted, in other words, allocated to poor or even worthless projects. Equal savings rates can thus lead to investments of widely differing degrees of usefulness from the viewpoint of future growth. However, in line with the earlier quote from Schumpeter, there are reasons to believe that the financial system has some role to play here as well.

Insert Figure 1.2 about here

The following quote from *Economic Focus* (UBS Economic Research, 1993) is part of a discussion motivated by the observation that, even for high-saving countries of Southeast Asia, the correlation between savings and growth has not been uniform.

"The paradox of raising saving without commensurate growth performance may be closely linked to the inadequate development of the financial system in a number of Asian economies. Holding back financial development ('financial repression') was a deliberate policy of many governments in Asia and elsewhere

²In a straightforward regression in which the dependent variable is the growth rate in real per capita GNP, the coefficient on the average fraction of real GNP represented by investment (I/Y) over the prior five years is positive but insignificant. Together with other results, this is interpreted as suggesting a reverse causation from real per capita GNP growth to investment spending. See Barro and Sala-i-Martin (1995), Chapter 12, for a full discussion. There is also a theoretically important distinction between the effects of increasing investment (savings) (as a proportion of national income) on an economy's *level* of wealth and its *growth rate*. Countries that save more will *ceteris paribus* be wealthier, but they need not grow more rapidly. The classic growth model of Solow (1956) illustrates this distinction.

who wished to maintain control over the flow of savings. (...) Typical measures of financial repression still include interest rate regulation, selective credit allocation, capital controls, and restricted entry into and competition within the banking sector”.

These comments take on special significance in light of the recent Asian crisis, which provides another, dramatic, illustration of the growth-finance nexus. Economists do not fully agree on what causes financial crises. There is, however, a consensus that in the case of several East-Asian countries, the weaknesses of the financial and banking sectors, such as those described as “financial repression,” must take part of the blame for the collapse and the ensuing economic regression that have marked the end of the 1990s in Southern Asia.

Let us try to go further than these general statements in the analysis of the savings and growth nexus and of the role of the financial system. Following Barro and Sala-i-Martin (1995), one can view the process of transferring funds from savers to investors in the following way.³ The least efficient system would be one in which all investments are made by the savers themselves. This is certainly inefficient because it requires a sort of “double coincidence” of intentions: Good investment ideas occurring in the mind of someone lacking past savings will not be realized. Funds that a non-entrepreneur saves would not be put to productive use. Yet, this unfortunate situation is a clear possibility if the necessary confidence in the financial system is lacking with the consequence that savers do not entrust the system with their savings. One can thus think of circumstances where savings never enter the financial system, or where only a small fraction do. When it does, it will typically enter via some sort of depository institution. In an international setting, a similar problem arises if national savings are primarily invested abroad, a situation that may reach alarming proportions in the case of underdeveloped countries.⁴ Let FS/S represent, then, the fraction of aggregate savings (S) being entrusted to the financial system (FS).

At a second level, the functioning of the financial system may be more or less costly. While funds transferred from a saver to a borrower via a direct loan are immediately and fully made available to the end user, the different functions of the financial system discussed above are often best fulfilled, or sometimes can only be fulfilled, through some form of intermediation, which typically involves some cost. Let us think of these costs as administrative costs, on the one hand, and costs linked to the reserve requirements of banks, on the other. Different systems will have different operating costs in this large sense,

³For a broader perspective and a more systematic connection with the relevant literature on this topic, see Levine (1997).

⁴The problem is slightly different here, however. Although capital flight is a problem from the viewpoint of building up a country’s home capital stock, the acquisition of foreign assets may be a perfectly efficient way of building a national capital stock. The effect on growth may be negative when measured in terms of GDP (Gross Domestic Product), but not necessarily so in terms of national income or GNP (Gross National product). Switzerland is an example of a rich country investing heavily abroad and deriving a substantial income flow from it. It can be argued that the growth rate of the Swiss Gross National Product (but probably not GDP) has been enhanced rather than decreased by this fact.

and, as a consequence, the amount of resources transferred to investors will also vary. Let us think of BOR/FS as the ratio of funds transferred from the financial system to borrowers and entrepreneurs.

Borrowers themselves may make diverse use of the funds borrowed. Some, for example, may have pure liquidity needs (analogous to the reserve needs of depository institutions), and if the borrower is the government, it may well be borrowing for consumption! For the savings and growth nexus, the issue is how much of the borrowed funds actually result in productive investments. Let I/BOR represent the fraction of borrowed funds actually invested. Note that BOR stands for borrowed funds whether private or public. In the latter case a key issue is what fraction of the borrowed funds are used to finance public investment as opposed to public consumption.

Finally let EFF denote the efficiency of the investment projects undertaken in society at a given time, with EFF normalized at unity; in other words, the average investment project has $EFF = 1$, the below-average project has $EFF < 1$, and conversely for the above average project (a project consisting of building a bridge leading nowhere would have an $EFF = 0$); K is the aggregate capital stock and Ω the depreciation rate. We may then write

$$\dot{K} = EFF \cdot I - \Omega K \quad (1.1)$$

or, multiplying and dividing I with each of the newly defined variables

$$\dot{K} = EFF \cdot (I/BOR) \cdot (BOR/FS) \cdot (FS/S) \cdot (S/Y) \cdot Y - \Omega K \quad (1.2)$$

where our notation is meant to emphasize that the growth of the capital stock at a given savings rate is likely to be influenced by the levels of the various ratios introduced above.⁵ Let us now review how this might be the case.

One can see that a financial system performing its matching function efficiently will positively affect the savings rate (S/Y) and the fraction of savings entrusted to financial institutions (FS/S). This reflects the fact that savers can find the right savings instruments for their needs. In terms of overall services net of inconvenience, this acts like an increase in the return to the fraction of savings finding its way into the financial system. The matching function is also relevant for the I/BOR ratio. With the appropriate instruments (like flexible overnight loan facilities) a firm's cash needs are reduced and a larger fraction of borrowed money can actually be used for investment.

By offering a large and diverse set of possibilities for spreading risks (insurance and hedging), an efficient financial system will also positively influence the savings ratio (S/Y) and the FS/S ratio. Essentially this works through improved return/risk opportunities, corresponding to an improved trade-off between future and present consumption (for savings intermediated through the financial system). Furthermore, in permitting entrepreneurs with risky projects to eliminate unnecessary risks by using appropriate instruments, an efficient financial system provides, somewhat paradoxically, a better platform for undertaking riskier projects. If, on average, riskier projects are also the ones with the

⁵ $\dot{K} = dK/dt$, that is, the change in K as a function of time.

highest returns, as most of financial theory reviewed later in this book leads us to believe, one would expect that the more efficiently this function is performed, the higher (*ceteris paribus*), the value of *EFF*; in other words, the higher, on average, the efficiency of the investment undertaken with the funds made available by savers.

Finally, a more efficient system may be expected to screen alternative investment projects more effectively and to better and more cost efficiently monitor the conduct of the investments (efforts of investors). The direct impact is to increase *EFF*. Indirectly this also means that, on average, the return/risk characteristics of the various instruments offered savers will be improved and one may expect, as a result, an increase in both S/Y and FS/S ratios.

The previous discussion thus tends to support the idea that the financial system plays an important role in permitting and promoting the growth of economies. Yet growth is not an objective in itself. There is such a thing as excessive capital accumulation. Jappelli and Pagano (1994) suggest that borrowing constraints,⁶ in general a source of inefficiency and the mark of a less than perfect financial system, may have led to more savings (in part unwanted) and higher growth. While their work is tentative, it underscores the necessity of adopting a broader and more satisfactory viewpoint and of more generally studying the impact of the financial system on social welfare. This is best done in the context of the theory of general equilibrium, a subject to which we shall turn in Section 1.6.

1.5 Financial Intermediation and the Business Cycle

Business cycles are the mark of all developed economies. According to much of current research, they are in part the result of external shocks with which these economies are repeatedly confronted. The depth and amplitude of these fluctuations, however, may well be affected by some characteristics of the financial system. This is at least the import of the recent literature on the financial accelerator. The mechanisms at work here are numerous, and we limit ourselves to giving the reader a flavor of the discussion.

The financial accelerator is manifest most straightforwardly in the context of monetary policy implementation. Suppose the monetary authority wishes to reduce the level of economic activity (inflation is feared) by raising real interest rates. The primary effect of such a move will be to increase firms' cost of capital and, as a result, to induce a decrease in investment spending as marginal projects are eliminated from consideration.

According to the financial accelerator theory, however, there may be further, substantial, secondary effects. In particular, the interest rate rise will reduce the value of firms' collateralizable assets. For some firms, this reduction may significantly diminish their access to credit, making them credit constrained. As a result, the fall in investment may exceed the direct impact of the higher

⁶By 'borrowing constraints' we mean the limitations that the average individual or firm may experience in his or her ability to borrow, at current market rates, from financial institutions.

cost of capital; tighter financial constraints may also affect input purchases or the financing of an adequate level of finished goods inventories. For all these reasons, the output and investment of credit-constrained firms will be more strongly affected by the action of the monetary authorities and the economic downturn may be made correspondingly more severe. By this same mechanism, any economy-wide reduction in asset values may have the effect of reducing economic activity under the financial accelerator.

Which firms are most likely to be credit constrained? We would expect that small firms, those for which lenders have relatively little information about the long-term prospects, would be principally affected. These are the firms from which lenders demand high levels of collateral. Bernanke et al. (1996) provide empirical support for this assertion using U.S. data from small manufacturing firms.

The financial accelerator has the power to make an economic downturn, of whatever origin, more severe. If the screening and monitoring functions of the financial system can be tailored more closely to individual firm needs, lenders will need to rely to a lesser extent on collateralized loan contracts. This would diminish the adverse consequences of the financial accelerator and perhaps the severity of business cycle downturns.

1.6 Financial Markets and Social Welfare

Let us now consider the role of financial markets in the allocation of resources and, consequently, their effects on social welfare. The perspective provided here places the process of financial innovation in the context of the theory of general economic equilibrium whose central concepts are closely associated with the *Ecole de Lausanne* and the names of Léon Walras, and Vilfredo Pareto.

Our starting point is the first theorem of welfare economics which defines the conditions under which the allocation of resources implied by the general equilibrium of a decentralized competitive economy is efficient or optimal in the Pareto sense.

First, let us define the terms involved. Assume a timeless economy where a large number of economic agents interact. There is an arbitrary number of goods and services, n . Consumers possess a certain quantity (possibly zero) of each of these n goods (in particular, they have the ability to work a certain number of hours per period). They can sell some of these goods and buy others at prices quoted in markets. There are a large number of firms, each represented by a production function – that is, a given ability (constrained by what is technologically feasible) to transform some of the available goods or services (inputs) into others (outputs); for instance, combining labor and capital to produce consumption goods. Agents in this economy act selfishly: Individuals maximize their well-being (utility) and firms maximize their profits.

General equilibrium theory tells us that, thanks to the action of the price system, order will emerge out of this uncoordinated chaos, provided certain conditions are satisfied. In the main, these hypotheses (conditions) are as follows:

H1: *Complete markets*. There exists a market on which a price is established for each of the n goods valued by consumers.

H2: *Perfect competition*. The number of consumers and firms (i.e., demanders and suppliers of each of the n goods in each of the n markets), is large enough so that no agent is in a position to influence (manipulate) market prices; that is, all agents take prices as given.

H3: Consumers' preferences are convex.

H4: Firms' production sets are convex as well.

H3 and H4 are technical conditions with economic implications. Somewhat paradoxically, the convexity hypothesis for consumers' preferences approximately translates into strictly concave utility functions. In particular, H3 is satisfied (in substance) if consumers display risk aversion, an assumption crucial for understanding financial markets, and one that will be made throughout this text. As already noted (Box 1.2), risk aversion translates into strictly concave utility functions (See Chapter 4 for details). H4 imposes requirements on the production technology. It specifically rules out increasing returns to scale in production. While important, this assumption is not at the heart of things in financial economics since for the most part we will abstract from the production side of the economy.

A **general competitive equilibrium** is a price vector p^* and an allocation of resources, resulting from the independent decisions of consumers and producers to buy or sell each of the n goods in each of the n markets, such that, at the equilibrium price vector p^* , supply equals demand in all markets simultaneously and the action of each agent is the most favorable to him or her among all those he/she could afford (technologically or in terms of his/her budget computed at equilibrium prices).

A **Pareto optimum** is an allocation of resources, however determined, where it is impossible to redistribute resources (i.e., to go ahead with further exchanges), without reducing the welfare of at least one agent. In a Pareto efficient (or Pareto optimal - we will use the two terminologies interchangeably) allocation of resources, it is thus not possible to make someone better off without making someone else worse off. Such a situation may not be just or fair, but it is certainly efficient in the sense of avoiding waste.

Omitting some purely technical conditions, the main results of general equilibrium theory can be summarized as follows:

1. *The existence of a competitive equilibrium*: Under H1 through H4, a competitive equilibrium is guaranteed to exist. This means that there indeed exists a price vector and an allocation of resources satisfying the definition of a competitive equilibrium as stated above.

2. *First welfare theorem*: Under H1 and H2, a competitive equilibrium, if it exists, is a Pareto optimum.

3. *Second welfare theorem*: Under H1 through H4, any Pareto-efficient allocation can be decentralized as a competitive equilibrium.

The Second welfare theorem asserts that, for any arbitrary Pareto-efficient allocation there is a price vector and a set of initial endowments such that this allocation can be achieved as a result of the free interaction of maximizing consumers and producers interacting in competitive markets. To achieve a specific Pareto-optimal allocation, some redistribution mechanism will be needed to reshuffle initial resources. The availability of such a mechanism, functioning without distortion (and thus waste) is, however, very much in question. Hence the dilemma between equity and efficiency that faces all societies and their governments.

The necessity of H1 and H2 for the optimality of a competitive equilibrium provides a rationale for government intervention when these hypotheses are not naturally satisfied. The case for antitrust and other “pro-competition” policies is implicit in H2; the case for intervention in the presence of externalities or in the provision of public goods follows from H1, because these two situations are instances of missing markets.⁷

Note that so far there does not seem to be any role for financial markets in promoting an efficient allocation of resources. To restore that role, we must abandon the fiction of a timeless world, underscoring, once again, the fact that time is of the essence in finance! Introducing the time dimension does not diminish the usefulness of the general equilibrium apparatus presented above, provided the definition of a good is properly adjusted to take into account not only its intrinsic characteristics, but also the time period in which it is available. A cup of coffee available at date t is different from a cup of coffee available at date $t + 1$ and, accordingly, it is traded on a different market and it commands a different price. Thus, if there are two dates, the number of goods in the economy goes from n to $2n$.

It is easy to show, however, that not all commodities need be traded for future as well as current delivery. The existence of a spot and forward market *for one good only* (taken as the numeraire) is sufficient to implement all the desirable allocations, and, in particular, restore, under H1 and H2, the optimality of the competitive equilibrium. This result is contained in Arrow (1964). It provides a powerful economic rationale for the existence of credit markets, markets where money is traded for future delivery.

Now let us go one step further and introduce uncertainty, which we will represent conceptually as a partition of all the relevant future scenarios into separate *states of nature*. To review, a state of nature is an exhaustive description of one possible relevant configuration of future events. Using this concept,

⁷Our model of equilibrium presumes that agents affect one another only through prices. If this is not the case, an economic externality is said to be present. These may involve either production or consumption. For example, there have been substantial negative externalities for fishermen associated with the construction of dams in the western United States: The catch of salmon has declined dramatically as these dams have reduced the ability of the fish to return to their spawning habitats. If the externality affects all consumers simultaneously, it is said to be a public good. The classic example is national defense. If any citizen is to consume a given level of national security, all citizens must be equally secure (and thus consume this public good at the same level). Both are instances of missing markets. Neither is there a market for national defense, nor for rights to disturb salmon habitats.

the applicability of the welfare theorems can be extended in a fashion similar to that used with time above, by defining goods not only according to the date but also to the state of nature at which they are (might be) available. This is the notion of contingent commodities. Under this construct, we imagine the market for ice cream decomposed into a series of markets: for ice cream today, ice cream tomorrow if it rains and the Dow Jones is at 10,000; if it rains and . . ., etc. Formally, this is a straightforward extension of the basic context: there are more goods, but this in itself is not restrictive⁸ [Arrow (1964) and Debreu (1959)].

The hypothesis that there exists a market for each and every good valued by consumers becomes, however, much more questionable with this extended definition of a typical good, as the example above suggests. On the one hand, the number of states of nature is, in principle, arbitrarily large and, on the other, one simply does not observe markets where commodities contingent on the realization of individual states of nature can routinely be traded. One can thus state that *if* markets are complete in the above sense, a competitive equilibrium is efficient, but the issue of completeness (H1) then takes center stage. Can Pareto optimality be obtained in a less formidable setup than one where there are complete contingent commodity markets? What does it mean to make markets “more complete?”

It was Arrow (1964), again, who took the first step toward answering these questions. Arrow generalized the result alluded to earlier and showed that it would be enough, in order to effect all desirable allocations, to have the opportunity to trade one good only across all states of nature. Such a good would again serve as the numeraire. The primitive security could thus be a claim promising \$1.00 (i.e., one unit of the numeraire) at a future date, contingent on the realization of a particular state, and zero under all other circumstances. We shall have a lot to say about such *Arrow-Debreu securities* (A-D securities from now on), which are also called *contingent claims*. Arrow asserted that if there is one such contingent claim corresponding to each and every one of the relevant future date/state configurations, hypothesis H1 could be considered satisfied, markets could be considered complete, and the welfare theorems would apply. Arrow’s result implies a substantial decrease in the number of required markets.⁹ However, for a complete contingent claim structure to be fully equivalent to a setup where agents could trade a complete set of contingent commodities, it must be the case that agents are assumed to know all future spot prices, contingent on the realization of all individual states of the world. Indeed, it is at these prices that they will be able to exchange the proceeds from their A-D securities for consumption goods. This hypothesis is akin to the hypothesis of rational expectations.¹⁰

A-D securities are a powerful conceptual tool and are studied in depth in

⁸Since n can be as large as one needs without restriction.

⁹Example: 2 dates, 3 basic goods, 4 states of nature: complete commodity markets require 12 contingent commodity markets plus 3 spot markets versus 4 contingent claims and 2 x 3 spot markets in the Arrow setup.

¹⁰For an elaboration on this topic, see Drèze (1971).

Chapters 8 and 10. They are not, however, the instruments we observe being traded in actual markets. Why is this the case, and in what sense is what we do observe an adequate substitute? To answer these questions, we first allude to a result (derived later on) which states that there is no single way to make markets complete. In fact there is potentially a large number of alternative financial structures achieving the same goal, and the complete A-D securities structure is only one of them. For instance, we shall describe, in Chapter 10, a context in which one might think of achieving an essentially complete market structure with options or derivative securities. We shall make use of this fact for pricing alternative instruments using arbitrage techniques. Thus, the failure to observe anything close to A-D securities being traded is not evidence against the possibility that markets are indeed complete.

In an attempt to match this discussion on the role played by financial markets with the type of markets we see in the real world, one can identify the different needs met by trading A-D securities in a complete markets world. In so doing, we shall conclude that, in reality, different types of needs are met through trading alternative specialized financial instruments (which, as we shall later prove, will all appear as portfolios of A-D securities).

As we have already observed, the time dimension is crucial for finance and, correspondingly, the need to exchange purchasing power across time is essential. It is met in reality through a variety of specific *non contingent* instruments, which are promised future payments independent of specific states of nature, except those in which the issuer is unable to meet his obligations (bankruptcies). Personal loans, bank loans, money market and capital market instruments, social security and pension claims are all assets fulfilling this basic need for redistributing purchasing power in the time dimension. In a complete market setup implemented through A-D securities, the needs met by these instruments would be satisfied by a certain configuration of positions in A-D securities. In reality, the specialized instruments mentioned above fulfill the demand for exchanging income through time.

One reason for the formidable nature of the complete markets requirement is that a state of nature, which is a complete description of the relevant future for a particular agent, includes some purely personal aspects of almost unlimited complexity. Certainly the future is different for you, in a relevant way, if you lose your job, or if your house burns, without these contingencies playing a very significant role for the population at large. In a pure A-D world, the description of the states of nature should take account of these *individual contingencies* viewed from the perspective of each and every market participant! In the real world, insurance contracts are the specific instruments that deal with the need for exchanging income across purely individual events or states. The markets for these contracts are part and parcel of the notion of complete financial markets. While such a specialization makes sense, it is recognized as unlikely that the need to trade across individual contingencies will be fully met through insurance markets because of specific difficulties linked with the hidden quality of these contingencies, (i.e., the inherent asymmetry in the information possessed by suppliers and demanders participating in these markets). The presence of

these asymmetries strengthens our perception of the impracticality of relying exclusively on pure A-D securities to deal with personal contingencies.

Beyond time issues and personal contingencies, most other financial instruments not only imply the exchange of purchasing power through time, but are also more specifically contingent on the realization of particular events. The relevant events here, however, are defined on a collective basis rather than being based on individual contingencies; they are contingent on the realization of events affecting groups of individuals and observable by everyone. An example of this is the situation where a certain level of profits for a firm implies the payment of a certain dividend against the ownership of that firm's equity. Another is the payment of a certain sum of money associated with the ownership of an option or a financial futures. In the later cases, the contingencies (sets of states of nature) are dependent on the value of the underlying asset itself.

1.7 Conclusion

To conclude this introductory chapter, we advance a vision of the financial system progressively evolving toward the complete markets paradigm, starting with the most obviously missing markets and slowly, as technological innovation decreases transaction costs and allows the design of more sophisticated contracts, completing the market structure. Have we arrived at a complete market structure? Have we come significantly closer? There are opposing views on this issue. While a more optimistic perspective is proposed by Merton (1990) and Allen and Gale (1994), we choose to close this chapter on two healthily skeptical notes. Tobin (1984, p.10), for one, provides an unambiguous answer to the above question:

“New financial markets and instruments have proliferated over the last decade, and it might be thought that the enlarged menu now spans more states of nature and moves us closer to the Arrow-Debreu ideal. Not much closer, I am afraid. The new options and futures contracts do not stretch very far into the future. They serve mainly to allow greater leverage to short-term speculators and arbitrageurs, and to limit losses in one direction or the other. Collectively they contain considerable redundancy. Every financial market absorbs private resources to operate, and government resources to police. The country cannot afford all the markets the enthusiasts may dream up. In deciding whether to approve proposed contracts for trading, the authorities should consider whether they really fill gaps in the menu and enlarge the opportunities for Arrow-Debreu insurance, not just opportunities for speculation and financial arbitrage.”

Shiller (1993, pp. 2–3) is even more specific with respect to missing markets:

“It is odd that there appear to have been no practical proposals for establishing a set of markets to hedge the biggest risks to standards of living. Individuals and organizations could hedge or insure themselves against risks to their standards of living if an array of risk markets – let us call them macro markets – could be established. These would be large international markets, securities, futures, options, swaps or analogous markets, for claims on major components of incomes (including service flows) shared by many people or organizations.

The settlements in these markets could be based on income aggregates, such as national income or components thereof, such as occupational incomes, or prices that value income flows, such as real estate prices, which are prices of claims on real estate service flows."

References

- Allen, F., Gale, D. (1994), *Financial Innovation and Risk Sharing*, MIT Press, Cambridge, Massachusetts.
- Arrow, K. J., (1964), "The Role of Securities in the Allocation of Risk," *Review of Economic Studies*, 31, 91–96.
- Barro, R. J., Sala-i-Martin, X. (1995), *Economic Growth*, McGraw-Hill, New York.
- Bernanke, B., Gertler, M., Gilchrist, S. (1996), "The Financial Accelerator and the Flight to Quality," *The Review of Economics and Statistics* 78, 1–15.
- Bernstein, P. L. (1992), *Capital Ideas. The Improbable Origins of Modern Wall Street*, The Free Press, New York.
- Debreu, G. (1959), *Theory of Value: An Axiomatic Analysis of Economic Equilibrium*, Wiley, New York.
- Drèze, J. H. (1971), "Market Allocation Under Uncertainty," *European Economic Review* 2, 133–165.
- Jappelli, T., Pagano, M. (1994), "Savings, Growth, and Liquidity Constraints," *Quarterly Journal of Economics* 109, 83–109.
- Levine, R. (1997), "Financial Development and Economic Growth: Views and Agenda," *Journal of Economic Literature* 35, 688–726.
- Merton, R.C. (1990), "The Financial System and Economic Performance," *Journal of Financial Services*, 4, 263–300
- Mishkin, F. (1992), *The Economics of Money, Banking and Financial Markets*, 3rd edition. Harper Collins, New York, Chapter 8.
- Schumpeter, J. (1934), *The Theory of Economic Development*, Duncker & Humblot, Leipzig. Trans. Opie, R. (1934), Harvard University Press, Cambridge, Massachusetts. Reprinted, Oxford University Press, New York (1964).
- Shiller, R. J. (1993), *Macro Markets – Creating Institutions for Managing Society's Largest Economic Risks*, Clarendon Press, Oxford.
- Solow, R. M. (1956), "A Contribution to the Theory of Economic Growth," *Quarterly Journal of Economics* 32, 65–94.

Tobin, J. (1984), "On the Efficiency of the Financial System," *Lloyds Bank Review*, 1–15.

UBS Economic Research, (1993), *Economic Focus*, Union Bank of Switzerland, no. 9.

Complementary Readings

As a complement to this introductory chapter, the reader will be interested in the historical review of financial markets and institutions found in the first chapter of Allen and Gale (1994). Bernstein (1992) provides a lively account of the birth of the major ideas making up modern financial theory including personal portraits of their authors.

Appendix: Introduction to General Equilibrium Theory

The goal of this appendix is to provide an introduction to the essentials of General Equilibrium Theory thereby permitting a complete understanding of Section 1.6 of the present chapter and facilitating the discussion of subsequent chapters (from Chapter 8 on). To make this presentation as simple as possible we'll take the case of a hypothetical exchange economy (that is, one with no production) with two goods and two agents. This permits using a very useful pedagogical tool known as the Edgeworth-Bowley box.

Insert Figure A.1.1 about here

Let us analyze the problem of allocating efficiently a given economy-wide endowment of 10 units of good 1 and 6 units of good 2 among two agents, A and B. In Figure A1.1, we measure good 2 on the vertical axis and good 1 on the horizontal axis. Consider the choice problem from the origin of the axes for Mr. A, and upside down, (that is placing the origin in the upper right corner), for Ms. B. An allocation is then represented as a point in a rectangle of size 6 x 10. Point *E* is an allocation at which Mr. A receives 4 units of good 2 and 2 units of good 1. Ms. B gets the rest, that is, 2 units of good 2 and 8 units of good 1. All other points in the box represent feasible allocations, that is, alternative ways of allocating the resources available in this economy.

Pareto Optimal Allocations

In order to discuss the notion of Pareto optimal or efficient allocations, we need to introduce agents' preferences. They are fully summarized, in the graphical context of the Edgeworth-Bowley box, by indifference curves (IC) or utility level curves. Thus, starting from the allocation *E* represented in Figure A1.1, we can record all feasible allocations that provide the same utility to Mr. A. The precise shape of such a level curve is person specific, but we can at least be confident that it slopes downward. If we take away some units of good 1, we have to compensate him with some extra units of good 2 if we are to leave his utility level unchanged. It is easy to see as well that the ICs of a consistent

person do not cross, a property associated with the notion of transitivity (and with rationality) in Chapter 3. And we have seen in Boxes 1.1 and 1.2 that the preference for smoothness translates into a strictly concave utility function, or, equivalently, convex-to-the-origin level curves as drawn in Figure A1.1. The same properties apply to the IC of Ms. B, of course viewed upside down with the upper right corner as the origin.

Insert Figure A.1.2 about here

With this simple apparatus we are in a position to discuss further the concept of Pareto optimality. Arbitrarily tracing the level curves of Mr.A and Ms.B as they pass through allocation E , (but in conformity with the properties derived in the previous paragraph), only two possibilities may arise: they cross each other at E or they are tangent to one another at point E . The first possibility is illustrated in Figure A1.1, the second in Figure A1.2. In the first case, allocation E cannot be a Pareto optimal allocation. As the picture illustrates clearly, by the very definition of level curves, if the ICs of our two agents cross at point E there is a set of allocations (corresponding to the shaded area in Figure A1.1) that are simultaneously preferred to E by both Mr. A and Ms. B. These allocations are Pareto superior to E , and, in that situation, it would indeed be socially inefficient or wasteful to distribute the available resources as indicated by E . Allocation D , for instance, is feasible and preferred to E by both individuals.

If the ICs are tangent to one another at point E' as in Figure A1.2, no redistribution of the given resources exists that would be approved by both agents. Inevitably, moving away from E' decreases the utility level of one of the two agents if it favors the other. In this case, E' is a Pareto optimal allocation. Figure A1.2 illustrates that it is not generally unique, however. If we connect all the points where the various ICs of our two agents are tangent to each other, we draw the line, labeled the contract curve, representing the infinity of Pareto optimal allocations in this simple economy.

An indifference curve for Mr.A is defined as the set of allocations that provide the same utility to Mr.A as some specific allocation; for example, allocation E : $\{(c_1^A, c_2^A) : U(c_1^A, c_2^A) = U(E)\}$. This definition implies that the slope of the IC can be derived by taking the total differential of $U(c_1^A, c_2^A)$ and equating it to zero (no change in utility along the IC), which gives:

$$\frac{\partial U(c_1^A, c_2^A)}{\partial c_1^A} dc_1^A + \frac{\partial U(c_1^A, c_2^A)}{\partial c_2^A} dc_2^A = 0, \quad (1.3)$$

and thus

$$-\frac{dc_2^A}{dc_1^A} = \frac{\frac{\partial U(c_1^A, c_2^A)}{\partial c_1^A}}{\frac{\partial U(c_1^A, c_2^A)}{\partial c_2^A}} \equiv MRS_{1,2}^A. \quad (1.4)$$

That is, the negative (or the absolute value) of the slope of the IC is the ratio of the marginal utility of good 1 to the marginal utility of good 2 specific

to Mr.A and to the allocation (c_1^A, c_2^A) at which the derivatives are taken. It defines Mr.A's Marginal Rate of Substitution (MRS) between the two goods.

Equation (1.4) permits a formal characterization of a Pareto optimal allocation. Our former discussion has equated Pareto optimality with the tangency of the ICs of Mr.A and Ms.B. Tangency, in turn, means that the slopes of the respective ICs are identical. Allocation E , associated with the consumption vector $(c_1^A, c_2^A)^E$ for Mr.A and $(c_1^B, c_2^B)^E$ for Ms.B, is thus Pareto optimal if and only if

$$MRS_{1,2}^A = \frac{\frac{\partial U(c_1^A, c_2^A)^E}{\partial c_1^A}}{\frac{\partial U(c_1^A, c_2^A)^E}{\partial c_2^A}} = \frac{\frac{\partial U(c_1^B, c_2^B)^E}{\partial c_1^B}}{\frac{\partial U(c_1^B, c_2^B)^E}{\partial c_2^B}} = MRS_{1,2}^B. \quad (1.5)$$

Equation (1.5) provides a complete characterization of a Pareto optimal allocation in an exchange economy except in the case of a corner allocation, that is, an allocation at the frontier of the box where one of the agents receives the entire endowment of one good and the other agent receives none. In that situation it may well be that the equality could not be satisfied except, hypothetically, by moving to the outside of the box, that is to allocations that are not feasible since they require giving a negative amount of one good to one of the two agents.

So far we have not touched on the issue of how the discussed allocations may be determined. This is the viewpoint of Pareto optimality which analysis is exclusively concerned with deriving efficiency properties of given allocations, irrespective of how they were achieved. Let us now turn to the concept of competitive equilibrium.

Competitive equilibrium

Associated with the notion of competitive equilibrium is the notion of markets and prices. One price vector (one price for each of our two goods), or simply a relative price taking good 1 as the numeraire, and setting $p_1 = 1$, is represented in the Edgeworth-Bowley box by a downward sloping line. From the viewpoint of either agent, such a line has all the properties of the budget line. It also represents the frontier of their opportunity set. Let us assume that the initial allocation, before any trade, is represented by point I in Figure A1.3. Any line sloping downward from I does represent the set of allocations that Mr.A, endowed with I , can obtain by going to the market and exchanging (competitively, taking prices as given) good 1 for 2 or vice versa. He will maximize his utility subject to this budget constraint by attempting to climb to the highest IC making contact with his budget set. This will lead him to select the allocation corresponding to the tangency point between one of his ICs and the price line. Because the same prices are valid for both agents, an identical procedure, viewed upside down from the upper right-hand corner of the box, will lead Ms.B to a tangency point between one of her ICs and the price line. At this stage, only two possibilities may arise: Mr.A and Ms.B have converged to the same allocation, (the two markets, for good 1 and 2, clear – supply and demand for the two goods are equal and we are at a competitive equilibrium);

or the two agents' separate optimizing procedures have lead them to select two different allocations. Total demand does not equal total supply and an equilibrium is not achieved. The two situations are described, respectively, in Figures A1.3 and A1.4.

Insert Figures A.1.3 and A.1.4 about here

In the disequilibrium case of Figure A1.4, prices will have to adjust until an equilibrium is found. Specifically, with Mr.A at point A and Ms.B at point B, there is an excess demand of good 2 but insufficient demand for good 1. One would expect the price of 2 to increase relative to the price of good 1 with the likely result that both agents will decrease their net demand for 2 and increase their net demand for 1. Graphically, this is depicted by the price curve tilting with point I as the axis and looking less steep (indicating, for instance, that if both agents wanted to buy good 1 only, they could now afford more of it). With regular ICs, the respective points of tangencies will converge until an equilibrium similar to the one described in Figure A1.3 is reached.

We will not say anything here about the conditions guaranteeing that such a process will converge. Let us rather insist on one crucial necessary precondition: that an equilibrium exists. In the text we have mentioned that assumptions H1 to H4 are needed to guarantee the existence of an equilibrium. Of course H4 does not apply here. H1 states the necessity of the existence of a price for each good, which is akin to specifying the existence of a price line. H2 defines one of the characteristics of a competitive equilibrium: that prices are taken as given by the various agents and the price line describes their perceived opportunity sets. Our discussion here can enlighten the need for H3. Indeed, in order for an equilibrium to have a chance to exist, the geometry of Figure A1.3 makes clear that the shapes of the two agents' ICs is relevant. The price line must be able to separate the "better than" areas of the two agents' ICs passing through a same point – the candidate equilibrium allocation. The better than area is simply the area above a given IC. It represents all the allocations providing higher utility than those on the level curve. This separation by a price line is not generally possible if the ICs are not convex, in which case an equilibrium cannot be guaranteed to exist. The problem is illustrated in Figure A1.5.

Insert Figure A.1.5 about here

Once a competitive equilibrium is observed to exist, which logically could be the case even if the conditions that guarantee existence are not met, the Pareto optimality of the resulting allocation is ensured by H1 and H2 only. In substance this is because once the common price line at which markets clear exists, the very fact that agents optimize taking prices as given, leads them to a point of tangency between their highest IC and the common price line. At the resulting allocation, both MRS are equal to the same price line and, consequently, are identical. The conditions for Pareto optimality are thus fulfilled.

Chapter 2 : The Challenges of Asset Pricing: A Roadmap

2.1 The main question of financial theory

Valuing risky cash flows or, equivalently, pricing risky assets is at the heart of financial theory.

Our discussion thus far has been conducted from the perspective of society as a whole, and it argues that a progressively more complete set of financial markets will generally enhance societal welfare by making it easier for economic agents to transfer income across future dates and states via the sale or purchase of individually tailored portfolios of securities. The desire of agents to construct such portfolios will be as much dependent on the market prices of the relevant securities as on their strict availability, and this leads us to the main topic of the text.

Indeed, the major practical question in finance is “how do we value a risky cash flow?”, and the main objective of this text is to provide a complete and up-to-date treatment of how it can be answered. For the most part, this textbook is thus a text on asset pricing. Indeed, an asset is nothing else than the right to future cash flows, whether these future cash flows are the result of interest payments, dividend payments, insurance payments, or the resale value of the asset. Conversely, when we compute the risk-adjusted present value (PV), we are, in effect, asking the question: If this project’s cash flow were traded as though it were a security, at what price would it sell given that it should pay the prevailing rate for securities of that same systematic risk level? We compare its fair market value, estimated in this way, with its cost, P_0 . Evaluating a project is thus a special case of evaluating a complex security.

Viewed in this way and abstracting from risk for the moment, the key object of our attention, be it an asset or an investment project, can be summarized as in Table 2.1.

Table 2.1 : Valuing a Risk-Free Cash Flow

$t = 0$	$t = 1$	$t = 2$	\dots	$t = T$
$P_0?$	\tilde{CF}_1	\tilde{CF}_2	\tilde{CF}_τ	\tilde{CF}_T
	$\frac{CF_1}{(1+r_1^f)}$	$\frac{CF_2}{(1+r_2^f)^2}$	$\frac{CF_\tau}{(1+r_\tau^f)^\tau}$	$\frac{CF_T}{(1+r_T^f)^T}$

In Table 2.1, $t = 0, 1, 2, \dots, \tau, \dots, T$ represents future dates. The duration of each period, the length of time between $\tau - 1$ and τ is arbitrary and can be viewed as one day, one month, one quarter or one year. The expression CF_τ stands for the possibly uncertain cash flows in period τ (whenever useful, we will identify random variables with a tilde), r_τ^f is the risk free, per-period interest rate prevailing between date $\tau - 1$ and τ , and P_0 denotes the to-be-determined

current price or valuation of the future cash flow. If the future cash flows will be available for sure, valuing the flow of future payments is easy. It requires adding the future cash flows after discounting them by the risk-free rate of interest, that is, adding the cells in the last line of the Table. The discounting procedure is indeed at the heart of our problem: it clearly serves to translate future payments into current dollars (those that are to be used to purchase the right to these future cash flows or in terms of which the current value of the future cash flow is to be expressed); in other words, the discounting procedure is what makes it possible to compare future dollars (i.e., dollars that will be available in the future) with current dollars.

If, however, the future cash flows will not be available for certain but are subject to random events - the interest payments depend on the debtor remaining solvent, the dividend payments depend on the financial strength of the equity issuer, the returns to the investment project depend on its commercial success -, then the valuation question becomes trickier, so much so that there does not exist a universally agreed way of proceeding that dominates all others.

In the same way that one dollar for sure tomorrow does not generally have the same value as one current dollar, one dollar tomorrow under a set of more or less narrowly defined circumstances, that is, in a subset of all possible states of nature, is also not worth one current dollar, not even one current dollar discounted at the risk free rate. Assume the risk free rate of return is 5% per year, then discounting at the risk free rate one dollar available in one year yields $\frac{1\$}{1.05} \cong \$.95$. This equality is an equality: it states that \$1 tomorrow will have a market price of \$.95 today when one year risk free securities earn 5%. It is a market assessment to the extent that the 5% risk free rate is an equilibrium market price. Now if \$1 for sure tomorrow is worth \$.95, it seems likely that \$1 tomorrow “maybe”, that is, under a restrictive subset of the states of nature, should certainly be worth less than \$.95. One can speculate for instance that if the probability of \$1 in a year is about $\frac{1}{2}$, then one should not be willing to pay more than $\frac{1}{2} \times \$.95$ for that future cash flow. But we have to be more precise than this. To that end, several lines of attack will be pursued. Let us outline them.

2.2 Discounting risky cash flows: various lines of attack

First, as in the certainty case, it is plausible to argue (and it can be formally demonstrated) that the valuation process is additive: the value of a sum of future cash flows will take the form of the sum of the values of each of these future cash flows. Second, as already anticipated, we will work with probabilities, so that the random cash flow occurring at a future date τ will be represented by a random variable: $\tilde{C}F_\tau$, for which a natural reference value is its expectation $E\tilde{C}F_\tau$. Another would be the value of this expected future cash flow discounted at the risk free rate: $\frac{E\tilde{C}F_\tau}{(1+r_f^\tau)^\tau}$. Now the latter expression cannot generally be the solution to our problem, although it is intuitively understandable that it will be when the risk issue does not matter; that is, when market participants

can be assumed as risk neutral. In the general case where risk needs to be taken into account, which typically means that risk bearing behavior needs to be remunerated, alterations to that reference formula are necessary. These alterations may take the following form:

1. The most common strategy consists of discounting at a rate that is higher than the risk free rate, that is, to discount at a rate which is the risk free rate increased by a certain amount π as in

$$\frac{E\tilde{C}F_\tau}{(1 + r_\tau^f + \pi)^\tau};$$

The underlying logic is straightforward: To price an asset equal to the present value of its expected future cash flows discounted at a particular rate is to price the asset in a manner such that, at its present value price, it is expected to earn that discount rate. The appropriate rate, in turn, must be the analyst's estimate of the rate of return on other financial assets that represent title to cash flows similar in risk and timing to that of the asset in question. This strategy has the consequence of pricing the asset to pay the prevailing competitive rate for its risk class. When we follow this approach, the key issue is to compute the appropriate risk premium.

2. Another approach in the same spirit consists of correcting the expected cash flow itself in such a way that one can continue discounting at the risk free rate. The standard way of doing this is to decrease the expected future cash flow by a factor Π that once again will reflect some form of risk or insurance premium as in

$$\frac{E\tilde{C}F_\tau - \Pi_\tau}{(1 + r_\tau^f)^\tau}.$$

3. The same idea can take the form, it turns out quite fruitfully, of distorting the probability distribution over which the expectations operator is applied so that taking the expected cash flow with this modified probability distribution justifies once again discounting at the risk free rate:

$$\frac{\hat{E}\tilde{C}F_\tau}{(1 + r_\tau^f)^\tau};$$

Here \hat{E} denotes the expectation taken with respect to the modified probability distribution.

4. Finally, one can think of decomposing the future cash flow $\tilde{C}F_\tau$ into its state by state elements. Denote $(CF(\theta_\tau))$ the actual payment that will occur in the specific possible state of nature θ_τ . If one is able to find the price today of \$1 tomorrow conditional on that particular state θ_τ being

realized, say $q(\theta_\tau)$, then surely the appropriate current valuation of \tilde{CF}_τ is

$$\sum_{\theta_\tau \in \Theta_\tau} q(\theta_\tau) CF(\theta_\tau),$$

where the summation takes place over all the possible future states θ_τ .

The procedures described above are really alternative ways of attacking the difficult valuation problem we have outlined, but they can only be given content in conjunction with theories on how to compute the risk premia (cases 1 or 2), to identify the distorted probability distribution (case 3) or to price future dollars state by state (case 4). For strategies 1 and 2, this can be done using the CAPM, the CCAPM or the APT; strategy 3 is characteristic of the Martingale approach; strategy 4 describes the perspective of Arrow-Debreu pricing.

2.3 Two main perspectives: Equilibrium vs. Arbitrage

There is another, even more fundamental way of classifying alternative valuation theories. All the known valuation theories borrow one of two main methodologies: the equilibrium approach or the arbitrage approach.

The traditional equilibrium approach consists of an analysis of the factors determining the supply and demand for the cash flow (asset) in question. The arbitrage approach attempts to value a cash flow on the basis of observations made on the various elements making up that cash flow.

Let us illustrate this distinction with an analogy. You are interested to price a bicycle. There are two ways to approach the question. If you follow the equilibrium approach you will want to study the determinants of supply and demand. Who are the producers? How many bicycles are they able to produce? What are the substitutes, including probably the existing stock of old bicycles potentially appearing on the second-hand market? After dealing with supply, turn to demand: Who are the buyers? What are the forecasts on the demand for bicycles? Etc.. Finally you will turn to the market structure. Is the market for bicycles competitive? If so, we know how the equilibrium price will emerge as a result of the matching process between demanders and suppliers. The equilibrium perspective is a sophisticated, complex approach, with a long tradition in economics, one that has also been applied in finance, at least since the fifties. We will follow it in the first part of this book, adopting standard assumptions that simplify, without undue cost, the supply and demand analysis for financial objects: the supply of financial assets at any point in time is assumed to be fixed and financial markets are viewed as competitive. Our analysis can thus focus on the determinants of the demand for financial assets. This requires that we first spend some time discussing the preferences and attitudes toward risk of investors, those who demand the assets (Chapters 3 and 4), before modeling the investment process, that is, how the demand for financial assets is determined (Chapters 5 and 6). Armed with these tools we will review the three main equilibrium theories, the CAPM in Chapter 7, AD pricing in Chapter 8 and the CCAPM in Chapter 9.

The other approach to valuing bicycles starts from observing that a bicycle is not (much) more than the sum of its parts. With a little knowledge, in almost infinite supply, and some time (which is not and this suggests that the arbitrage approach holds only as an approximation that may be rather imprecise in circumstances where the time and intellectual power required to “assemble the bicycle” from the necessary spare parts are non-trivial; that is, when the remuneration of the necessary “engineers” matters), it is possible to re-engineer or replicate any bicycle with the right components. It then follows that if you know the price of all the necessary elements - frame, handle, wheel, tire, saddle, break and gearshift - you can determine relatively easily the market value of the bicycle.

The arbitrage approach is, in a sense, much more straightforward than the equilibrium approach. It is also more robust: if the no arbitrage relation between the price of the bicycle and the price of its parts did not hold, anyone with a little time could become a bicycle manufacturer and make good money. If too many people exploit that idea, however, the prices of parts and the prices of bicycles will start adjusting and be forced into line. This very idea is powerful for the object at hand, financial assets, because if markets are complete in the sense discussed in section 1.6, then it can easily be shown that all the component prices necessary to value any arbitrary cash flow are available. Furthermore, little time and few resources (relative to the global scale of these product markets) are needed to exploit arbitrage opportunities in financial markets.

There is, however, an obvious limitation to the arbitrage approach. Where do we get the price of the parts if not through an equilibrium approach? That is, the arbitrage approach is much less ambitious and more partial than the equilibrium approach. Even though it may be more practically useful in the domains where the price of the parts are readily available, it does not make up for a general theory of valuation and, in that sense, has to be viewed as a complement to the equilibrium approach. In addition, the equilibrium approach, by forcing us to rationalize investors’ demand for financial assets, provides useful lessons for the practice of asset management. The foundations of this inquiry will be put in place in Chapters 3 to 6 - which together make up Part II of the book - while Chapter 14 will extend the treatment of this topic beyond the case of the traditional one-period static portfolio analysis and focus on the specificities of long run portfolio management. Finally, the arbitrage and equilibrium approaches can be combined. In particular, one fundamental insight that we will develop in Chapter 10 is that any cash flow can be viewed as a portfolio of, that is, can be replicated with Arrow-Debreu securities. This makes it very useful to start using the arbitrage approach with A-D securities as the main building blocks for pricing assets or valuing cash flows. Conversely the same chapter will show that options can be very useful in completing the markets and thus in obtaining a full set of prices for “the parts” that will then be available to price the bicycles. In other words, the Arrow-Debreu equilibrium pricing theory is a good platform for arbitrage valuation. The link between the two approaches is indeed so tight that we will use our acquired knowledge of equilibrium models to understand one of the major arbitrage approaches,

the Martingale pricing theory (Chapters 11 and 12). We will then propose an overview of the Arbitrage Pricing Theory (APT) in Chapter 13. Chapters 10 to 13 together make up Part IV of this book.

Part V will focus on three extensions. As already mentioned, Chapter 14 deals with long run asset management. Chapter 15 focuses on some implications of incomplete markets whose consequences we will illustrate from the twin viewpoints of the equilibrium and arbitrage approaches. We will use it as a pretext to review the Modigliani-Miller theorem and, in particular, to understand why it depends on the hypothesis of complete markets. Finally, in Chapter 16 we will open up, just a little, the Pandora's box of heterogeneous beliefs. Our goal is to understand a number of issues that are largely swept under the rug in standard asset management and pricing theories and, in the process, restate the Efficient Market hypothesis.

Table 2.2: The Roadmap

	Equilibrium	Arbitrage
Preliminaries	Utility theory - Ch.3-4 Investment demand - Ch.5-6	
Computing risk premia	CAPM - Ch.7 CCAPM - Ch.9	APT - Ch. 13
Identifying distorted probabilities		Martingale measure - Ch. 11 & 12
Pricing future dollars state by state	A-D pricing I - Ch. 8	A-D pricing II - Ch. 10

2.4 This is not all of finance!

2.4.1 Corporate Finance

Intermediate Financial Theory focuses on the valuation of risky cash flows. Pricing a future (risky) dollar is a dominant ingredient in most financial problems. But it is not all of finance! Our capital markets perspective in particular side-steps many of the issues surrounding how the firm generates and protects the cash flow streams to be priced. It is this concern that is at the core of corporate financial theory or simply “corporate finance.”

In a broad sense, corporate finance is concerned with decision making at the firm level whenever it has a financial dimension, has implications for the financial situation of the firm, or is influenced by financial considerations. In particular it is a field concerned, first and foremost, with the investment decision (what

projects should be accepted), the financing decision (what mix of securities should be issued and sold to finance the chosen investment projects), the payout decision (how should investors in the firm, and in particular the equity investors, be compensated), and risk management (how corporate resources should be protected against adverse outcomes). Corporate finance also explores issues related to the size and the scope of the firm, e.g, mergers and acquisitions and the pricing of conglomerates, the internal organization of the firm, the principles of corporate governance and the forms of remuneration of the various stakeholders.¹

All of these decisions individually and collectively do influence the firm's free cash flow stream and, as such, have asset pricing implications. The decision to increase the proportion of debt in the firm's capital structure, for example, increases the riskiness of its equity cash flow stream, the standard deviation of the equilibrium return on equity, etc.

Of course when we think of the investment decision itself, the solution to the valuation problem is of the essence, and indeed many of the issues typically grouped under the heading of capital budgeting are intimately related to the preoccupations of the present text. We will be silent, however, on most of the other issues listed above which are better viewed as arising in the context of bilateral (rather than market) relations and, as we will see, in situations where asymmetries of information play a dominant role.

The goal of this section is to illustrate the difference in perspectives by reviewing, selectively, the corporate finance literature particularly as regards the capital structure of the firm, and contrasting it with the capital markets perspective that we will be adopting throughout this text. In so doing we also attempt to give the flavor of an important research area while detailing many of the topics this text elects not to address.

2.4.2 Capital structure

We focus on the capital structure issue in Chapter 15 where we explore the assumption underlying the famous Modigliani-Miller irrelevance result: in the absence of taxes, subsidies and contracting costs, the value of a firm is independent of its capital structure if the firm's investment policy is fixed and financial markets are complete. Our emphasis will concern how this result fundamentally rests on the complete markets assumption.

The corporate finance literature has not ignored the completeness issue but rather has chosen to explore its underlying causes, most specifically information asymmetries between the various agents concerned, managers, shareholders etc.²

¹The recent scandals (Enron, WorldCom) in the U.S. and in Europe (Parmalat, ABB) place in stark light the responsibilities of boards of directors for ultimate firm oversight as well as their frequent failure in doing so. The large question here is what sort of board structure is consistent with superior long run firm performance? Executive compensation also remains a large issue: In particular, to what extent are the incentive effects of stock or stock options compensation influenced by the manager's outside wealth?

²Tax issues have tended to dominate the corporate finance capital structure debate until recently and we will review this arena shortly. The relevance of taxes is not a distinguishing

While we touch on the issue of heterogeneity of information in a market context, we do so only in our last Chapter (16), emphasizing there that heterogeneity raises a number of tough modeling difficulties. These difficulties justify the fact that most of capital market theory either is silent on the issue of heterogeneity (in particular, when it adopts the arbitrage approach) or explicitly assumes homogeneous information on the part of capital market participants.

In contrast, the bulk of corporate finance builds on asymmetries of information and explores the various problems they raise. These are typically classified as leading to situations of ‘moral hazard’ or ‘adverse selection’. An instance of the former is when managers are tempted to take advantage of their superior information to implement investment plans that may serve their own interests at the expense of those of shareholders or debt holders. An important branch of the literature concerns the design of contracts which take moral hazard into account. The choice of capital structure, in particular, will be seen potentially to assist in their management (see, e.g., Zwiebel (1996)).

A typical situation of ‘adverse selection’ occurs when information asymmetries between firms and investors make firms with ‘good’ investment projects indistinguishable to outside investors from firms with poor projects. This suggests a tendency for all firms to receive the same financing terms (a so-called “pooling equilibrium” where firms with less favorable prospects may receive better than deserved financing arrangements). Firms with good projects must somehow indirectly distinguish themselves in order to receive the more favorable financing terms they merit. For instance they may want to attach more collateral to their debt securities, an action that firms with poor projects may find too costly to replicate (see, e.g., Stein (1992)). Again, the capital structure decision may sometimes help in providing a resolution of the “adverse selection” problem. Below we review the principal capital structure perspectives.

2.4.3 Taxes and capital structure

Understanding the determinants of a firm’s capital structure (the proportion of debt and equity securities it has outstanding in value terms) is the ‘classical’ problem in corporate finance. Its intellectual foundations lie in the seminal work of Modigliani and Miller (1958) who argue for capital structure irrelevance in a world without taxes and with complete markets (an hypothesis that excludes information asymmetries).

The corporate finance literature has also emphasized the fact that when one security type receives favored tax treatment (typically this is debt via the tax deductibility of interest), then the firm’s securities become more valuable in the aggregate if more of that security is issued, since to do so is to reduce the firm’s overall tax bill and thus enhance the free cash flow to the security holders. Since the bondholders receive the same interest and principal payments, irrespective of the tax status of these payments from the firm’s perspective, any tax based

feature of the corporate finance perspective alone. Taxes also matter when we think of valuing risky cash flows although we will have very little to say about it except that all the cash flows we consider are to be thought of as after tax cash flows.

cash flow enhancement is captured by equity holders. Under a number of further specialized assumptions (including the hypothesis that the firm's debt is risk-free), these considerations lead to the classical relationship

$$V_L = V_U + \tau D;$$

the value of a firm's securities under partial debt financing (V_L , where 'L' denotes leverage in the capital structure) equals its value under all equity financing (V_U , where 'U' denotes unlevered or an all equity capital structure) plus the present value of the interest tax subsidies. This latter quantity takes the form of the corporate tax rate (τ) times the value of debt outstanding (D) when debt is assumed to be perpetual (unchanging capital structure).

In return terms this value relationship can be transformed into a relationship between levered and unlevered equity returns:

$$r_L^e = r_U^e + (1 - \tau)(D/E)(r_U^e - r_f);$$

i.e., the return on levered equity, r_L^e , is equal to the return on unlevered equity, r_U^e , plus a risk premium due to the inherently riskier equity cash flow that the presence of the fixed payments to debt creates. This premium, as indicated, is related to the tax rate, the firm's debt/equity ratio (D/E), a measure of the degree of leverage, and the difference between the unlevered equity rate and the risk free rate, r_f . Immediately we observe that capital structure considerations influence not only expected equilibrium equity returns via

$$Er_L^e = Er_U^e + (1 - \tau)D/E(Er_U^e - r_f),$$

where E denotes the expectations operator, but also the variance of returns since

$$\sigma_{r_L^e}^2 = (1 + (1 - \tau)D/E)^2 \sigma_{r_U^e}^2 > \sigma_{r_U^e}^2$$

under the mild assumption that r_f is constant in the very short run. These relationships illustrate but one instance of corporate financial considerations affecting the patterns of equilibrium returns as observed in the capital markets.

The principal drawback to this tax based theory of capital structure is the natural implication that if one security type receives favorable tax treatment (usually debt), then if the equity share price is to be maximized the firm's capital structure should be composed exclusively of that security type – i.e., all debt, which is not observed. More recent research in corporate finance has sought to avoid these extreme tax based conclusions by balancing the tax benefits of debt with various costs of debt, including bankruptcy and agency costs. Our discussion broadly follows Harris and Raviv (1991).

2.4.4 Capital structure and agency costs

An important segment of the literature seeks to explain financial decisions by examining the conflicts of interests among claimholders within the firm. While agency conflicts can take a variety of forms, most of the literature has focused

on shareholders' incentives to increase investment risk – the asset substitution problem – or to reject positive NPV projects – the underinvestment problem. Both of these conflicts increase the cost of debt and thus reduce the firm's value maximizing debt ratio.

Another commonly discussed determinant of capital structure arises from manager-stockholder conflicts. Managers and shareholders have different objectives. In particular, managers tend to value investment more than shareholders do. Although there are a number of potentially powerful internal mechanisms to control managers, the control technology normally does not permit the costless resolution of this conflict between managers and investors. Nonetheless, the cash-flow identity implies that constraining financing, hedging and payout policy places indirect restrictions on investment policy. Hence, even though investment policy is not contractible, by restricting the firm in other dimensions, it is possible to limit the manager's choice of an investment policy. For instance, Jensen (1986) argues that debt financing can increase firm value by reducing the free cash flow. This idea is formalized in more recent papers by Stulz (1990) and Zwiebel (1996). Also, by reducing the likelihood of both high and low cash flows, risk management not only can control shareholders' underinvestment incentives but managers' ability to overinvest as well.

More recently, the corporate finance literature has put some emphasis on the cost that arises from conflicts of interests between controlling and minority shareholders. In most countries, publicly traded companies are not widely held, but rather have controlling shareholders. Moreover, these controlling shareholders have the power to pursue private benefits at the expense of minority shareholders, within the limits imposed by investor protection. The recent "law and finance" literature following Shleifer and Vishny (1997) and La Porta et al. (1998) argues that the expropriation of minority shareholders by the controlling shareholder is at the core of agency conflicts in most countries. While these conflicts have been widely discussed in qualitative terms, the literature has largely been silent on the magnitude of their effects.

2.4.5 The pecking order theory of investment financing

The seminal reference here is Myers and Majluf (1984) who again base their work on the assumption that investors are generally less well informed (asymmetric information) than insider-managers vis-à-vis the firm's investment opportunities. As a result, new equity issues to finance new investments may be so underpriced (reflecting average project quality) that NPV positive projects from a societal perspective may have a negative NPV from the perspective of existing shareholders and thus not be financed. Myers and Majluf (1984) argue that this underpricing can be avoided if firms finance projects with securities which have more assured payout patterns and thus are less susceptible to undervaluation: internal funds and, to a slightly lesser extent, debt securities especially risk free debt. It is thus in the interests of shareholders to finance projects first with retained earnings, then with debt, and lastly with equity. An implication of this qualitative theory is that the announcement of a new equity issuance is likely

to be accompanied by a fall in the issuing firm's stock price since it indicates that the firm's prospects are too poor for the preferred financing alternatives to be accessible.

The pecking order theory has led to a large literature on the importance of security design. For example, Stein (1992) argues that companies may use convertible bonds to get equity into their capital structures "through the back-door" in situations where informational asymmetries make conventional equity issues unattractive. In other words, convertible bonds represent an indirect mechanism for implementing equity financing that mitigates the adverse selection costs associated with direct equity sales. This explanation for the use of convertibles emphasizes the role of the call feature – that will allow good firms to convert the bond into common equity – and costs of financial distress – that will prevent bad firms from mimicking good ones. Thus, the announcement of a convertible bond issue should be greeted with a less negative – and perhaps even positive – stock price response than an equity issue of the same size by the same company.

2.5 Conclusions

We have presented four general approaches and two main perspectives to the valuation of risky cash flows. This discussion was meant as providing an organizing principle and a roadmap for the extended treatment of a large variety of topics on which we are now embarking. Our brief excursion into corporate finance was intended to suggest some of the agency issues that are part and parcel of a firm's cash flow determination. That we have elected to focus on pricing issues surrounding those cash flow streams does not diminish the importance of the issues surrounding their creation.

References

- Harris, M., Raviv, A. (1991), "The Theory of Capital Structure," *Journal of Finance*, 46, 297-355.
- Jensen, M. (1986), "Agency Costs of Free Cash Flow, Corporate Finance and Takeovers," *American Economic Review*, 76, 323-329.
- Jensen, M., Meckling, W. (1976), "Theory of the Firm: Managerial Behavior, Agency Costs, and Capital Structure," *Journal of Financial Economics*, 3.
- La Porta, R., F. Lopes de Silanes, A. Shleifer, and R. Vishny (1998), "Law and Finance", *Journal of Political Economy*, 106, 1113-1155.
- Myers, S., Majluf, N. (1984), "Corporate Financing and Investment Decisions when Firms Have Information that Investors Do Not Have," *Journal of Financial Economics*, 13, 187-221.

- Modigliani, F., Miller, M. (1958), "The Cost of Capital, Corporate Finance, and the Theory of Investment," *American Economic Review* 48, 261-297.
- Shleifer A. and R. Vishny, (1997), "A Survey of Corporate Governance", *Journal of Finance*, 52, 737-783.
- Stein, J. (1992), "Convertible Bonds as Backdoor Equity Financing", *Journal of Financial Economics* 32, 3-23.
- Stulz, R. (1990), "Managerial Discretion and Optimal Financial Policies", *Journal of Financial Economics* 26, 3-27.
- Zwiebel, J. (1996), "Dynamic Capital Structure under Managerial Entrenchment", *American Economic Review* 86, 1197-1215.

Part II

The Demand For Financial Assets

Chapter 3: Making Choices in Risky Situations

3.1 Introduction

The first stage of the equilibrium perspective on asset pricing consists in developing an understanding of the determinants of the **demand** for securities of various risk classes. Individuals demand securities (in exchange for current purchasing power) in their attempt to redistribute income across time and states of nature. This is a reflection of the consumption-smoothing and risk-reallocation function central to financial markets.

Our endeavor requires an understanding of three building blocks:

1. how financial risk is defined and measured;
2. how an investor's attitude toward or tolerance for risk is to be conceptualized and then measured;
3. how investors' risk attitudes interact with the subjective uncertainties associated with the available assets to determine an investor's desired portfolio holdings (demands).

In this and the next chapter we give a detailed overview of points 1 and 2; point 3 is treated in succeeding chapters.

3.2 Choosing Among Risky Prospects: Preliminaries

When we think of the "risk" of an investment, we are typically thinking of uncertainty in the future cash flow stream to which the investment represents title. Depending on the state of nature that may occur in the future, we may receive different payments and, in particular, much lower payments in some states than others. That is, we model an asset's associated cash flow in any future time period as a **random variable**.

Consider, for example, the investments listed in Table 3.1, each of which pays off next period in either of two equally likely possible states. We index these states by $\theta = 1, 2$ with their respective probabilities labelled π_1 and π_2 .

Table 3.1: Asset Payoffs (\$)

	Cost at $t = 0$	Value at $t = 1$	
		$\pi_1 = \pi_2 = 1/2$	
		$\theta = 1$	$\theta = 2$
Investment 1	-1,000	1,050	1,200
Investment 2	-1,000	500	1,600
Investment 3	-1,000	1,050	1,600

First, this comparison serves to introduce the important notion of **dominance**. Investment 3 clearly dominates both investments 1 and 2 in the sense that it pays as much in all states of nature, and strictly more in at least one state. The **state-by-state dominance** illustrated here is the strongest possible form of dominance. Without any qualification, we will assume that all rational

individuals would prefer investment 3 to the other two. Basically this means that we are assuming the typical individual to be non-satiated in consumption: she desires more rather than less of the consumption goods these payoffs allow her to buy.

In the case of dominance the choice problem is trivial and, in some sense, the issue of defining risk is irrelevant. The ranking defined by the concept of dominance is, however, very incomplete. If we compare investments 1 and 2, one sees that neither dominates the other. Although it performs better in state 2, investment 2 performs much worse in state 1. There is no ranking possible on the basis of the dominance criterion. The different prospects must be characterized from a different angle. The concept of risk enters necessarily.

On this score, we would probably all agree that investments 2 and 3 are comparatively riskier than investment 1. Of course for investment 3, the dominance property means that the only risk is an upside risk. Yet, in line with the preference for smooth consumption discussed in Chapter 1, the large variation in date 2 payoffs associated with investment 3 is to be viewed as undesirable in itself. When comparing investments 1 and 2, the qualifier “riskier” undoubtedly applies to the latter. In the worst state, the payoff associated with 2 is worse; in the best state it is better.

These comparisons can alternatively, and often more conveniently, be represented if we describe investments in terms of their performance on a per dollar basis. We do this by computing the state contingent rates of return (ROR) that we will typically associate with the symbol r . In the case of the above investments, we obtain the results in Table 3.2:

Table 3.2: State Contingent ROR (r)

	$\theta = 1$	$\theta = 2$
Investment 1	5%	20%
Investment 2	-50%	60%
Investment 3	5%	60%

One sees clearly that all rational individuals should prefer investment 3 to the other two and that this same dominance cannot be expressed when comparing 1 and 2.

The fact that investment 2 is riskier, however, does not mean that all rational risk-averse individuals would necessarily prefer 1. Risk is not the only consideration and the ranking between the two projects is, in principle, preference dependent. This is more often the case than not; dominance usually provides a very incomplete way of ranking prospects. This is why we have to turn to a description of preferences, the main object of this chapter.

The most well-known approach at this point consists of summarizing such investment return distributions (that is, the random variables representing returns) by their mean (Er_i) and variance (σ_i^2), $i = 1, 2, 3$. The variance (or its square root, the standard deviation) of the rate of return is then naturally used

as the measure of “risk” of the project (or the asset). For the three investments just listed, we have:

$$\begin{aligned} Er_1 &= 12.5\% ; \sigma_1^2 = \frac{1}{2}(5 - 12.5)^2 + \frac{1}{2}(20 - 12.5)^2 = (7.5)^2, \text{ or } \sigma_1 = 7.5\% \\ Er_2 &= 5\% ; \sigma_2 = 55\% \text{ (similar calculation)} \\ Er_3 &= 32.5\% ; \sigma_3 = 27.5\% \end{aligned}$$

If we decided to summarize these return distributions by their means and variances only, investment 1 would clearly appear more attractive than investment 2: It has both a higher mean return and a lower variance. In terms of the mean-variance criterion, investment 1 dominates investment 2; 1 is said to *mean-variance dominate* 2. Our previous discussion makes it clear that **mean-variance dominance** is neither as strong, nor as general a concept as state-by-state dominance. Investment 3 mean-variance dominates 2 but not 1, although it dominates them both on a state-by-state basis! This is surprising and should lead us to be cautious when using any mean variance return criterion. We will, later on, detail circumstances where it is fully reliable. At this point let us anticipate that it will not be generally so, and that restrictions will have to be imposed to legitimize its use.

The notion of mean-variance dominance can be expressed in the form of a criterion for selecting investments of equal magnitude, which plays a prominent role in modern portfolio theory:

1. For investments of the same Er , choose the one with the lowest σ .
2. For investments of the same σ , choose the one with the greatest Er .

In the framework of modern portfolio theory, one could not understand a rational agent choosing 2 rather than investment 1.

We cannot limit our inquiry to the concept of dominance, however. Mean-variance dominance provides only an incomplete ranking among uncertain prospects, as Table 3.3 illustrates:

Table 3.3: State-Contingent Rates of Return

	$\theta = 1$	$\theta = 2$
Investment 4	3%	5%
Investment 5	2%	8%
	$\pi_1 = \pi_2 = 1/2$ $ER_4 = 4\%; \sigma_4 = 1\%$ $ER_5 = 5\%; \sigma_5 = 3\%$	

Comparing these two investments, it is not clear which is best; there is no dominance in either state-by-state or mean-variance terms. Investment 5 is expected to pay 1.25 times the expected return of investment 4, but, in terms of standard deviation, it is also three times riskier. The choice between 4 and 5, when restricted to mean-variance characterizations, would require specifying the terms at which the decision maker is willing to *substitute* expected return for a given risk reduction. In other words, what decrease in expected return is

he willing to accept for a 1% decrease in the standard deviation of returns? Or conversely, does the 1 percentage point additional expected return associated with investment 5 adequately compensate for the (3 times) larger risk? Responses to such questions are preference dependent (i.e., vary from individual to individual).

Suppose, for a particular individual, the terms of the trade-off are well represented by the index E/σ (referred to as the “Sharpe” ratio). Since $(E/\sigma)_4 = 4$ while $(E/\sigma)_5 = 5/3$, investment 4 is better than investment 5 for that individual. Of course another investor may be less risk averse; that is, he may be willing to accept more extra risk for the same expected return. For example, his preferences may be adequately represented by $(E - 1/3\sigma)$ in which case he would rank investment 5 (with an index value of 4) above investment 4 (with a value of $3\frac{2}{3}$).¹

All these considerations strongly suggest that we have to adopt a more general viewpoint for comparing potential return distributions. This viewpoint is part of utility theory, to which we turn after describing some of the problems associated with the empirical characterization of return distributions in Box 3.1.

Box 3.1 Computing Means and Variances in Practice

Useful as it may be conceptually, calculations of distribution moments such as the mean and the standard deviation are difficult to implement in practice. This is because we rarely know what the future states of nature are, let alone their probabilities. We also do not know the returns in each state. A frequently used proxy for a future return distribution is its historical distribution. This amounts to selecting a historical time period and a periodicity, say monthly prices for the past 60 months, and computing the historical returns as follows:

$$r_{s,j} = \text{return to stock } s \text{ in month } j = ((p_{s,j} + d_{s,j})/p_{s,j-1}) - 1$$

where $p_{s,j}$ is the price of stock s in month j , and $d_{s,j}$ its dividend, if any, that month. We then summarize the past distribution of stock returns by the average historical return and the variance of the historical returns. By doing so we, in effect, assign a probability of $\frac{1}{60}$ to each past observation or event.

In principle this is an acceptable way to estimate a return distribution for the future if we think the “mechanism” generating these returns is “stationary”: that the future will in some sense closely resemble the past. In practice, this hypothesis is rarely fully verified and, at the minimum, it requires careful checking. Also necessary for such a straightforward, although customary, application is that the return realizations are independent of each other, so that

¹Observe that the Sharpe ratio criterion is not immune to the criticism discussed above. With the Sharpe ratio criterion, investment 3 ($E/\sigma = 1.182$) is inferior to investment 1 ($E/\sigma = 1.667$). Yet we know that 3 dominates 1 since it pays a higher return in every state. This problem is pervasive with the mean-variance investment criterion. For any mean-variance choice criterion, whatever the terms of the trade-off between mean and variance or standard deviation, one can produce a paradox such as the one illustrated above. This confirms this criterion is not generally applicable without additional restrictions. The name Sharpe ratio refers to Nobel Prize winner William Sharpe, who first proposed the ratio for this sort of comparison.

today's realization does not reveal anything materially new about the probabilities of tomorrow's returns (formally, that the conditional and unconditional distributions are identical). \square

3.3 A Prerequisite: Choice Theory Under Certainty

A good deal of financial economics is concerned with how people make choices. The objective is to understand the systematic part of individual behavior and to be able to predict (at least in a loose way) how an individual will react to a given situation. Economic theory describes individual behavior as the result of a process of optimization under constraints, the objective to be reached being determined by individual preferences, and the constraints being a function of the person's income or wealth level and of market prices. This approach, which defines the *homo economicus* and the notion of **economic rationality**, is justified by the fact that individuals' behavior is predictable only to the extent that it is systematic, which must mean that there is an attempt at achieving a set objective. It is not to be taken literally or normatively.²

To develop this sense of rationality systematically, we begin by summarizing the objectives of investors in the most basic way: we postulate the existence of a preference relation, represented by the symbol \succeq , describing investors' ability to compare various bundles of goods, services, and money. For two bundles a and b , the expression

$$a \succeq b$$

is to be read as follows: For the investor in question, bundle a is strictly preferred to bundle b , or he is indifferent between them. Pure indifference is denoted by $a \sim b$, strict preference by $a \succ b$.

The notion of economic rationality can then be summarized by the following assumptions:

A.1 Every investor possesses such a preference relation and it is *complete*, meaning that he is able to decide whether he prefers a to b , b to a , or both, in which case he is indifferent with respect to the two bundles. That is, for any two bundles a and b , either $a \succeq b$ or $b \succeq a$ or both. If both hold, we say that the investor is indifferent with respect to the bundles and write $a \sim b$.

A.2 This preference relation satisfies the fundamental property of transitivity: For any bundles a , b , and c , if $a \succeq b$ and $b \succeq c$, then $a \succeq c$.

A further requirement is also necessary for technical reasons:

A.3 The preference relation \succeq is continuous in the following sense: Let $\{x_n\}$ and $\{y_n\}$ be two sequences of consumption bundles such that $x_n \mapsto x$ and $y_n \mapsto y$.³ If $x_n \succeq y_n$ for all n , then the same relationship is preserved in the

²By this we mean that economic science does not *prescribe* that individuals maximize, optimize, or simply behave as if they were doing so. It just finds it productive to summarize the systematic behavior of economic agents with such tools.

³We use the standard sense of (normed) convergence on R^N .

limit: $x \succeq y$.

A key result can now be expressed by the following proposition.

Theorem 3.1:

Assumptions A.1 through A.3 are sufficient to guarantee the existence of a continuous, time-invariant, real-valued utility function⁴ u , such that for any two objects of choice (consumption bundles of goods and services; amounts of money, etc.) a and b ,

$$\begin{aligned} a &\succeq b \text{ if and only if} \\ u(a) &\geq u(b). \end{aligned}$$

Proof:

See, for example, Mas-Colell et. al. (1995), Proposition 3.c.1. \square

This result asserts that the assumption that decision makers are endowed with a utility function (which they are assumed to maximize) is, in reality, no different than assuming their preferences among objects of choice define a relation possessing the (weak) properties summarized in A1 through A3.

Notice that Theorem 3.1 implies that if $u(\cdot)$ is a valid representation of an individual's preferences, any increasing transformation of $u(\cdot)$ will do as well since such a transformation by definition will preserve the ordering induced by $u(\cdot)$. Notice also that the notion of a consumption bundle is, formally, very general. Different elements of a bundle may represent the consumption of the same good or service in different time periods. One element might represent a vacation trip in the Bahamas this year; another may represent exactly the same vacation next year. We can further expand our notion of different goods to include the same good consumed in mutually exclusive states of the world. Our preference for hot soup, for example, may be very different if the day turns out to be warm rather than cold. These thoughts suggest Theorem 3.1 is really quite general, and can, formally at least, be extended to accommodate uncertainty. Under uncertainty, however, ranking bundles of goods (or vectors of monetary payoffs, see below) involves more than pure elements of taste or preferences. In the hot soup example, it is natural to suppose that our preferences for hot soup are affected by the probability we attribute to the day being hot or cold. Disentangling pure preferences from probability assessments is the subject to which we now turn.

3.4 Choice Theory Under Uncertainty: An Introduction

Under certainty, the choice is among consumption baskets with known characteristics. Under uncertainty, however, our emphasis changes. The objects of

⁴In other words, $u: R^n \rightarrow R^+$

choice are typically no longer consumption bundles but vectors of state contingent money payoffs (we'll reintroduce consumption in Chapter 5). Such vectors are formally what we mean by an *asset* that we may purchase or an *investment*. When we purchase a share of a stock, for example, we know that its sale price in one year will differ depending on what events transpire within the firm and in the world economy. Under financial uncertainty, therefore, the choice is among alternative investments leading to different possible income levels and, hence, ultimately different consumption possibilities. As before, we observe that people do make investment choices, and if we are to make sense of these choices, there must be a stable underlying order of preference defined over different alternative investments. The spirit of Theorem 3.1 will still apply. With appropriate restrictions, these preferences can be represented by a utility index defined on investment possibilities, but obviously something deeper is at work. It is natural to assume that individuals have no intrinsic taste for the assets themselves (IBM stock as opposed Royal Dutch stock, for example); rather, they are interested in what payoffs these assets will yield and with what likelihood (see Box 3.2, however).

Box 3.2 Investing Close to Home

Although the assumption that investors only care for the final payoff of their investment without any trace of romanticism is standard in financial economics, there is some evidence to the contrary and, in particular, for the assertion that many investors, at the margin at least, prefer to purchase the claims of firms whose products or services are familiar to them. In a recent paper, Huberman (2001) examines the stock ownership records of the seven regional Bell operating companies (RBOCs). He discovered that, with the exception of residents of Montana, Americans are more likely to invest in their local regional Bell operating company than in any other. When they do, their holdings average \$14,400. For those who venture farther from home and hold stocks of the RBOC of a region other than their own, the average holding is only \$8,246. Considering that every local RBOC cannot be a better investment choice than all of the other six, Huberman interprets his findings as suggesting investors' psychological need to feel comfortable with where they put their money. □

One may further hypothesize that investor preferences are indeed very simple after uncertainty is resolved: They prefer a higher payoff to a lower one or, equivalently, to earn a higher return rather than a lower one. Of course they do not know *ex ante* (that is, before the state of nature is revealed) which asset will yield the higher payoff. They have to choose among prospects, or probability distributions representing these payoffs. And, as we saw in Section 3.2, typically, no one investment prospect will strictly dominate the others. Investors will be able to imagine different possible scenarios, some of which will result in a higher return for one asset, with other scenarios favoring other assets. For instance, let us go back to our favorite situation where there are only two states of nature; in other words, two conceivable scenarios and two assets, as seen in Table 3.4.

Table 3.4: Forecasted Price per Share in One Period

	State 1	State 2
IBM	\$100	\$150
Royal Dutch	\$90	\$160
	Current Price of both assets is \$100	

There are two key ingredients in the choice between these two alternatives. The first is the probability of the two states. All other things being the same, the more likely is state 1, the more attractive IBM stock will appear to prospective investors. The second is the *ex post* (once the state of nature is known) level of utility provided by the investment. In Table 3.4 above, IBM yields \$100 in state 1 and is thus preferred to Royal Dutch, which yields \$90 if this scenario is realized; Royal Dutch, however, provides \$160 rather than \$150 in state 2. Obviously, with unchanged state probabilities, things would look different if the difference in payoffs were increased in one state as in Table 3.5.

Table 3.5: Forecasted Price per Share in One Period

	State 1	State 2
IBM	\$100	\$150
Royal Dutch	\$90	\$200
	Current Price of both assets is \$100	

Here even if state 1 is slightly more likely, the superiority of Royal Dutch in state 2 makes it look more attractive. A more refined perspective is introduced if we go back to our first scenario but now introduce a third contender, Sony, with payoffs of \$90 and \$150, as seen in Table 3.6.

Table 3.6: Forecasted Price per Share in One Period

	State 1	State 2
IBM	\$100	\$150
Royal Dutch	\$90	\$160
Sony	\$90	\$150
	Current Price of all assets is \$100	

Sony is dominated by both IBM and Royal Dutch. But the choice between the latter two can now be described in terms of an improvement of \$10 over the Sony payoff, either in state 1 or in state 2. Which is better? The relevant feature is that IBM adds \$10 when the payoff is low (\$90) while Royal Dutch adds the same amount when the payoff is high (\$150). Most people would think IBM more desirable, and with equal state probabilities, would prefer IBM. Once again this is an illustration of the preference for smooth consumption

(smoother income allows for smoother consumption).⁵ In the present context one may equivalently speak of risk aversion or of the well-known microeconomic assumption of decreasing marginal utility (the incremental utility when adding ever more consumption or income is smaller and smaller).

The expected utility theorem provides a set of hypotheses under which an investor's preference ranking over investments with uncertain money payoffs may be represented by a utility index combining, in the most elementary way (i.e., linearly), the two ingredients just discussed — the preference ordering on the *ex post* payoffs and the respective probabilities of these payoffs.

We first illustrate this notion in the context of the two assets considered earlier. Let the respective probability distributions on the price per share of IBM and Royal Dutch (RDP) be described, respectively, by $\tilde{p}_{IBM} = p_{IBM}(\theta_i)$ and $\tilde{p}_{RDP} = p_{RDP}(\theta_i)$ together with the probability π_i that the state of nature θ_i will be realized. In this case the expected utility theorem provides sufficient conditions on an agent's preferences over uncertain asset payoffs, denoted \succeq , such that

$$\tilde{p}_{IBM} \succeq \tilde{p}_{RDP}$$

if and only if there exists a real valued function U for which

$$\begin{aligned} EU(\tilde{p}_{IBM}) &= \pi_1 U(p_{IBM}(\theta_1)) + \pi_2 U(p_{IBM}(\theta_2)) \\ &> \pi_1 U(p_{RDP}(\theta_1)) + \pi_2 U(p_{RDP}(\theta_2)) = EU(\tilde{p}_{RDP}) \end{aligned}$$

More generally, the utility of any asset A with payoffs $p_A(\theta_1), p_A(\theta_2), \dots, p_A(\theta_N)$ in the N possible states of nature with probabilities $\pi_1, \pi_2, \dots, \pi_N$ can be represented by

$$\mathbb{U}(A) = EU(p_A(\theta_i)) = \sum_{i=1}^N \pi_i U(p_A(\theta_i))$$

in other words, by the weighted mean of *ex post* utilities with the state probabilities as weights. $\mathbb{U}(A)$ is a real number. Its precise numerical value, however, has no more meaning than if you are told that the temperature is 40 degrees when you do not know if the scale being used is Celsius or Fahrenheit. It is useful, however, for comparison purposes. By analogy, if it is 40° today, but it will be 45° tomorrow, you at least know it will be warmer tomorrow than it is today. Similarly, the expected utility number is useful because it permits attaching a number to a probability distribution and this number is, under appropriate hypotheses, a good representation of the relative ranking of a particular member of a family of probability distributions (assets under consideration).

⁵Of course, for the sake of our reasoning, one must assume that nothing else important is going on simultaneously in the background, and that other things, such as income from other sources, if any, and the prices of the consumption goods to be purchased with the assets' payoffs, are not tied to what the payoffs actually are.

3.5 The Expected Utility Theorem

Let us discuss this theorem in the simple context where objects of choice take the form of simple lotteries. The generic lottery is denoted (x, y, π) ; it offers payoff (consequence) x with probability π and payoff (consequence) y with probability $1 - \pi$. This notion of a lottery is actually very general and encompasses a huge variety of possible payoff structures. For example, x and y may represent specific monetary payoffs as in Figure 3.1.a, or x may be a payment while y is a lottery as in Figure 3.1.b, or even x and y may both be lotteries as in Figure 3.1.c. Extending these possibilities, some or all of the x_i 's and y_i 's may be lotteries, etc. We also extend our choice domain to include individual payments, lotteries where one of the possible monetary payoffs is certain; for instance,

$$(x, y, \pi) = x \text{ if (and only if) } \pi = 1 \text{ (see axiom C.1).}$$

Moreover, the theorem holds as well for assets paying a continuum of possible payoffs, but our restriction makes the necessary assumptions and justifying arguments easily accessible. Our objective is a conceptual transparency rather than absolute generality. All the results extend to much more general settings.

Insert Figure 3.1 about here

Under these representations, we will adopt the following **axioms and conventions**:

- C.1**
- a. $(x, y, 1) = x$
 - b. $(x, y, \pi) = (y, x, 1 - \pi)$
 - c. $(x, z, \pi) = (x, y, \pi + (1 - \pi)\tau)$ if $z = (x, y, \tau)$

C.1c informs us that agents are concerned with the net cumulative probability of each outcome. Indirectly, it further accommodates lotteries with multiple outcomes; see Figure 3.2, for an example where $p = (x, y, \pi')$, and $q = (z, w, \hat{\pi})$, and $\pi = \pi_1 + \pi_2 = \frac{\pi_1}{\pi_1 + \pi_2}$, etc.

Insert Figure 3.2 about here

C.2 There exists a preference relation \succeq , defined on lotteries, which is complete and transitive.

C.3 The preference relation is continuous in the sense of A.3 in the earlier section.

By C.2 and C.3 alone we know (Theorem 3.1) that there exists a utility function, which we will denote by $\mathbb{U}(\cdot)$, defined both on lotteries and on specific payments since, by assumption C.1a, a payment may be viewed as a (degenerate) lottery. For any payment x , we have

$$U(x) = \mathbb{U}((x, y, 1))$$

Our remaining assumptions are thus necessary only to guarantee that this function assumes the expected utility form.

C.4 Independence of irrelevant alternatives. Let (x, y, π) and (x, z, π) be any two lotteries; then, $y \succeq z$ if and only if $(x, y, \pi) \succeq (x, z, \pi)$.

C.5 For simplicity, we also assume that there exists a best (i.e., most preferred lottery), b , as well as a worst, least desirable, lottery w .

In our argument to follow (which is constructive, i.e., we explicitly exhibit the expected utility function), it is convenient to use relationships that follow directly from these latter two assumptions. In particular, we'll use C.6 and C.7:

C.6 Let x, k, z be consequences or payoffs for which $x > k > z$. Then there exists a π such that $(x, z, \pi) \sim k$.

C.7 Let $x \succ y$. Then $(x, y, \pi_1) \succeq (x, y, \pi_2)$ if and only if $\pi_1 > \pi_2$. This follows directly from C.4.

Theorem 3.2:

If axioms C.1 to C.7 are satisfied, then there exists a utility function U defined on the lottery space so that:

$$\mathbb{U}((x, y, \pi)) = \pi U(x) + (1 - \pi)U(y)$$

Proof:

We outline the proof in a number of steps:

1. Without loss of generality, we may normalize $\mathbb{U}(\cdot)$ so that $\mathbb{U}(b) = 1$, $\mathbb{U}(w) = 0$.

2. For all other lotteries z , define $\mathbb{U}(z) = \pi_z$ where π_z satisfies $(b, w, \pi_z) \sim z$

Constructed in this way $\mathbb{U}(z)$ is well defined since,

a. by C.6, $\mathbb{U}(z) = \pi_z$ exists, and

b. by C.7, $\mathbb{U}(z)$ is unique. To see this latter implication, assume, to the contrary, that $\mathbb{U}(z) = \pi_z$ and also $\mathbb{U}(z) = \pi'_z$ where $\pi_z > \pi'_z$. By assumption C.4,

$$z \sim (b, w, \pi_z) \succ (b, w, \pi'_z) \sim z; \text{ a contradiction.}$$

3. It follows also from C.7 that if $m \succ n$, $\mathbb{U}(m) = \pi_m > \pi_n = \mathbb{U}(n)$. Thus, $\mathbb{U}(\cdot)$ has the property of a utility function.

4. Lastly, we want to show that $\mathbb{U}(\cdot)$ has the required property. Let x, y be monetary payments, π a probability. By C.1a, $U(x), U(y)$ are well-defined real numbers. By C.6,

$$\begin{aligned}(x, y, \pi) &\sim ((b, w, \pi_x), (b, w, \pi_y)), \pi \\ &\sim (b, w, \pi\pi_x + (1 - \pi)\pi_y), \text{ by C.1c.}\end{aligned}$$

Thus, by definition of $\mathbb{U}(\cdot)$,

$$\mathbb{U}((x, y, \pi)) = \pi\pi_x + (1 - \pi)\pi_y = \pi U(x) + (1 - \pi)U(y).$$

Although we have chosen x, y as monetary payments, the same conclusion holds if they are lotteries. \square

Before going on to a more careful examination of the assumptions underlying the expected utility theorem, a number of clarifying thoughts are in order. First, the overall Von-Neumann Morgenstern (VNM) utility function $\mathbb{U}(\cdot)$ defined over lotteries, is so named after the originators of the theory, the justly celebrated mathematicians John von Neumann and Oskar Morgenstern. In the construction of a VNM utility function, it is customary first to specify its restriction to certainty monetary payments, the so-called utility of money function or simply the *utility function*. Note that the VNM utility function and its associated utility of money function are not the same. The former is defined over uncertain asset payoff structures while the latter is defined over individual monetary payments.

Given the objective specification of probabilities (thus far assumed), it is the utility function that uniquely characterizes an investor. As we will see shortly, different additional assumptions on $U(\cdot)$ will identify an investor's tolerance for risk. We do, however, impose the maintained requirement that $U(\cdot)$ be increasing for all candidate utility functions (more money is preferred to less). Second, note also that the expected utility theorem confirms that investors are concerned only with an asset's final payoffs and the cumulative probabilities of achieving them. For expected utility investors the structure of uncertainty resolution is irrelevant (Axiom C.1a).⁶

Third, although the introduction to this chapter concentrates on comparing rates of return distributions, our expected utility theorem in fact gives us a tool for comparing different asset payoff distributions. Without further analysis, it does not make sense to think of the utility function as being defined over a rate of return. This is true for a number of reasons. First, returns are expressed on a per unit (per dollar, Swiss Francs (SF) etc.) basis, and do not identify the magnitude of the initial investment to which these rates are to be applied. We thus have no way to assess the implications of a return distribution for an investor's wealth position. It could, in principle, be anything. Second, the notion of a rate of return implicitly suggests a time interval: The payout is received after the asset is purchased. So far we have only considered the atemporal

⁶See Section 3.7 for a generalization on this score.

evaluation of uncertain investment payoffs. In Chapter 4, we generalize the VNM representation to preferences defined over rates of returns.

Finally, as in the case of a general order of preferences over bundles of commodities, the VNM representation is preserved under a certain class of linear transformations. If $\mathbb{U}(\cdot)$ is a Von-Neuman-Morgenstern utility function, then $\mathbb{V}(\cdot) = a\mathbb{U}(\cdot) + b$ where $a > 0$, is also such a function. Let (x, y, π) be some uncertain payoff and let $U(\cdot)$ be the utility of money function associated with \mathbb{U} .

$$\begin{aligned}\mathbb{V}((x, y, \pi)) &= a\mathbb{U}((x, y, \pi)) + b = a[\pi U(x) + (1 - \pi)U(y)] + b \\ &= \pi[a\mathbb{U}(x) + b] + (1 - \pi)[a\mathbb{U}(y) + b] \equiv \pi\mathbb{V}(x) + (1 - \pi)\mathbb{V}(y)\end{aligned}$$

Every linear transformation of an expected utility function is thus also an expected utility function. The utility of money function associated with \mathbb{V} is $[aU(\cdot) + b]$; $\mathbb{V}(\cdot)$ represents the same preference ordering over uncertain payoffs as $\mathbb{U}(\cdot)$. On the other hand, a nonlinear transformation doesn't always respect the preference ordering. It is in that sense that utility is said to be **cardinal** (see Exercise 3.1).

3.6 How Restrictive Is Expected Utility Theory? The Allais Paradox

Although apparently innocuous, the above set of axioms has been hotly contested as representative of rationality. In particular, it is not difficult to find situations in which investor preferences violate the independence axiom. Consider the following four possible asset payoffs (lotteries):

$$\begin{aligned}L^1 &= (10,000, 0, 0.1) \quad L^2 = (15,000, 0, 0.09) \\ L^3 &= (10,000, 0, 1) \quad L^4 = (15,000, 0, 0.9)\end{aligned}$$

When investors are asked to rank these payoffs, the following ranking is frequently observed:

$$L^2 \succ L^1,$$

(presumably because L^2 's positive payoff in the favorable state is much greater than L^1 's while the likelihood of receiving it is only slightly smaller) and,

$$L^3 \succ L^4,$$

(here it appears that the certain prospect of receiving 10,000 is worth more than the potential of an additional 5,000 at the risk of receiving nothing).

By the structure of compound lotteries, however, it is easy to see that:

$$\begin{aligned}L^1 &= (L^3, L^0, 0.1) \\ L^2 &= (L^4, L^0, 0.1) \text{ where } L^0 = (0, 0, 1)\end{aligned}$$

By the independence axiom, the ranking between L^1 and L^2 on the one hand, and L^3 and L^4 on the other, should thus be identical!

This is the Allais Paradox.⁷ There are a number of possible reactions to it.

1. Yes, my choices were inconsistent; let me think again and revise them.
2. No, I'll stick to my choices. The following kinds of things are missing from the theory of choice expressed solely in terms of asset payoffs:
 - the pleasure of gambling, and/or
 - the notion of regret.

The idea of regret is especially relevant to the Allais paradox, and its application in the prior example would go something like this. L^3 is preferred to L^4 because of the regret involved in receiving nothing if L^4 were chosen and the bad state ensued. We would, at that point, regret not having chosen L^3 , the certain payment. The expected regret is high because of the nontrivial probability (.10) of receiving nothing under L^4 . On the other hand, the expected regret of choosing L^2 over L^1 is much smaller (the probability of the bad state is only .01 greater under L^2 and in either case the probability of success is small), and insufficient to offset the greater expected payoff. Thus L^2 is preferred to L^1 .

Box 3.3 On the Rationality of Collective Decision Making

Although the discussion in the text pertains to the rationality of individual choices, it is a fact that many important decisions are the result of collective decision making. The limitations to the rationality of such a process are important and, in fact, better understood than those arising at the individual level. It is easy to imagine situations in which transitivity is violated once choices result from some sort of aggregation over more basic preferences.

Consider three portfolio managers who decide which stocks to add to the portfolios they manage by majority voting. The stocks currently under consideration are General Electric (GE), Daimler-Chrysler (DC), and Sony (S). Based on his fundamental research and assumptions, each manager has rational (i.e., transitive) preferences over the three possibilities:

Manager 1: $GE \succ_1 DC \succ_1 S$
Manager 2: $S \succ_2 GE \succ_2 DC$
Manager 3: $DC \succ_3 S \succ_3 GE$

If they were to vote all at once, they know each stock would receive one vote (each stock has its advocate). So they decide to vote on pair-wise choices: (GE vs. DB), (DB vs. S), and (S vs. GE). The results of this voting (GE dominates DB, DB dominates S, and S dominates GE) suggest an intransitivity in the aggregate ordering. Although this example illustrates an intransitivity, it is an intransitivity that arises from the operation of a collective choice mechanism (voting) rather than being present in the individual orders of preference of the participating agents. There is a large literature on this subject that is closely

⁷Named after the Nobel prize winner Maurice Allais who was the first to uncover the phenomenon.

identified with Arrow's "Impossibility Theorem," See Arrow (1963) for a more exhaustive discussion. \square

The Allais paradox is but the first of many phenomena that appear to be inconsistent with standard preference theory. Another prominent example is the general pervasiveness of *preference reversals*, events that may approximately be described as follows. Individuals, participating in controlled experiments were asked to choose between two lotteries,

$(4, 0, .9)$ and $(40, 0, .1)$. More than 70 percent typically chose $(4, 0, .9)$. When asked at what price they would be willing to sell the lotteries if they were to own them, however, a similar percentage demanded the higher price for $(40, 0, .1)$. At first appearances, these choices would seem to violate transitivity. Let x, y be, respectively, the sale prices of $(4, 0, .9)$ and $(40, 0, .10)$. Then this phenomenon implies

$$x \sim (4, 0, .9) \succ (40, 0, .1) \sim y, \text{ yet } y > x.$$

Alternatively, it may reflect a violation of the assumed principle of procedure invariance, which is the idea that investors' preference for different objects should be indifferent to the manner by which their preference is elicited. Surprisingly, more narrowly focused experiments, which were designed to force a subject with expected utility preferences to behave consistently, gave rise to the same reversals. The preference reversal phenomenon could thus, in principle, be due either to preference intransitivity, or to a violation of the independence axiom, or of procedure invariance.

Various researchers who, through a series of carefully constructed experiments, have attempted to assign the blame for preference reversals lay the responsibility largely at the feet of procedure invariance violations. But this is a particularly alarming conclusion as Thaler (1992) notes. It suggests that *"the context and procedures involved in making choices or judgements influence the preferences that are implied by the elicited responses. In practical terms this implies that (economic) behavior is likely to vary across situations which economists (would otherwise) consider identical."* This is tantamount to the assertion that the notion of a preference ordering is not well defined. While investors may be able to express a consistent (and thus mathematically representable) preference ordering across television sets with different features (e.g., size of the screen, quality of the sound, etc.), this may not be possible with lotteries or consumption baskets containing widely diverse goods.

Grether and Plott (1979) summarize this conflict in the starkest possible terms:

"Taken at face value, the data demonstrating preference reversals are simply inconsistent with preference theory and have broad implications about research priorities within economics. The inconsistency is deeper than the mere lack of transitivity or even stochastic transitivity. It suggests that no optimization principles of any sort lie behind the simplest of human choices and that the uniformities in human choice behavior which lie behind market behavior result from principles which are of a completely different sort from those generally accepted."

At this point it is useful to remember, however, that the goal of economics and finance is not to describe individual, but rather market, behavior. There is a real possibility that occurrences of individual irrationality essentially “wash out” when aggregated at the market level. On this score, the proof of the pudding is in the eating and we have little alternative but to see the extent to which the basic theory of choice we are using is able to illuminate financial phenomena of interest. All the while, the discussion above should make us alert to the possibility that unusual phenomena might be the outcome of deviations from the generally accepted preference theory articulated above. While there is, to date, no preference ordering that accommodates preference reversals – and it is not clear there will ever be one – more general constructs than expected utility have been formulated to admit other, seemingly contradictory, phenomena.

3.7 Generalizing the VNM Expected Utility Representation

Objections to the assumptions underlying the VNM expected utility representation have stimulated the development of a number of alternatives, which we will somewhat crudely aggregate under the title *non-expected utility theory*. Elements of this theory differ with regard to which fundamental postulate of expected utility is relaxed. We consider four and refer the reader to Machina (1987) for a more systematic survey.

3.7.1 Preference for the Timing of Uncertainty Resolution

To grasp the idea here we must go beyond our current one period setting. Under the VNM expected utility representation, investors are assumed to be concerned only with actual payoffs and the cumulative probabilities of attaining them. In particular, they are assumed to be indifferent to the timing of uncertainty resolution. To get a better idea of what this means, consider the two investment payoff trees depicted in Figure 3.3. These investments are to be evaluated from the viewpoint of date 0 (today).

Insert Figure 3.3 about here

Under the expected utility postulates, these two payoff structures would be valued (in utility terms) identically as

$$EU(\tilde{P}) = U(100) + [\pi U(150) + (1 - \pi)U(25)]$$

This means that a VNM investor would not care if the uncertainty were resolved in period 0 (immediately) or one period later. Yet, people are, in fact, very different in this regard. Some want to know the outcome of an uncertain event as soon as possible; others prefer to postpone it as long as possible.

Kreps and Porteus (1978) were the first to develop a theory that allowed for these distinctions. They showed that if investor preferences over uncertain sequential payoffs were of the form

$$U_0(P_1, P_2(\tilde{\theta})) = W(P_1, E(U_1(P_1, P_2(\tilde{\theta}))),$$

then investors would prefer early (late) resolution of uncertainty according to whether $W(P_1, \cdot)$ is convex (concave) (loosely, whether $W_{22} > 0$ or $W_{22} < 0$). In the above representation P_i is the payoff in period $i = 1, 2$. If $W(P_1, \cdot)$ were concave, for example, the expected utility of investment 1 would be lower than investment 2.

The idea can be easily illustrated in the context of the example above. We assume functional forms similar to those used in an illustration of Kreps and Porteus (1978); in particular, assume $W(P_1, EU) = EU^{1.5}$, and $U_1(P_1, P_2(\tilde{\theta})) = (P_1 + P_2(\tilde{\theta}))^{1/2}$. Let $\pi = .5$, and note that the overall composite function $U_0(\cdot)$ is concave in all of its arguments. In computing utilities at the decision nodes [0], [1a], [1b], and [1c] (the latter decisions are trivial ones), we must be especially scrupulous to observe exactly the dates at which the uncertainty is resolved under the two alternatives:

$$\begin{aligned} [1a] : EU_1^{1a}(P_1, P_2(\theta)) &= (100 + 150)^{1/2} = 15.811 \\ [1b] : EU_1^{1b}(P_1, P_2(\theta)) &= (100 + 25)^{1/2} = 11.18 \\ [1c] : EU_1^{1c}(P_1, P_2(\theta)) &= .5(100 + 150)^{1/2} + .5(100 + 25)^{1/2} = 13.4955 \end{aligned}$$

At $t = 0$, the expected utility on the upper branch is

$$\begin{aligned} EU_0^{1a,1b}(P_1, P_2(\tilde{\theta})) &= EW^{1a,1b}(P_1, P_2(\tilde{\theta})) \\ &= .5W(100, 15.811) + .5W(100, 11.18) \\ &= .5(15.811)^{1.5} + .5(11.18)^{1.5} \\ &= 50.13, \end{aligned}$$

while on the lower branch

$$EU_0^{1c}(P_1, P_2(\tilde{\theta})) = W(100, 13.4955) = (13.4955)^{1.5} = 49.57.$$

This investor clearly prefers early resolution of uncertainty which is consistent with the convexity of the $W(\cdot)$ function. Note that the result of the example is simply an application of Jensen's inequality.⁸ If $W(\cdot)$ were concave, the ordering would be reversed.

There have been numerous specializations of this idea, some of which we consider in Chapter 4 (See Weil (1990) and Epstein and Zin (1989)). At the moment it is sufficient to point out that such representations are not consistent with the VNM axioms.

⁸Let $a = (100 + 150)^{1/2}, b = (100 + 25)^{1/2}, g(x) = x^{1.5}(\text{convex})$, $EU_0^{1a,1b}(P, \tilde{P}_2(\theta)) = Eg(x) > g(E\tilde{x}) = EU_0^{1c}(P, \tilde{P}_2(\theta))$ where $\tilde{x} = a$ with prob = .5 and $\tilde{x} = b$ with prob = .5.

3.7.2 Preferences That Guarantee Time-Consistent Planning

Our setting is once again intertemporal, where uncertainty is resolved in each future time period. Suppose that at each date $t \in \{0, 1, 2, \dots, T\}$, an agent has a preference ordering \succeq_t defined over all future (state-contingent) consumption bundles, where \succeq_t will typically depend on her past consumption history. The notion of *time-consistent planning* is this: if, at each date, the agent could plan against any future contingency, what is the required relationship among the family of orderings $\{\succeq_t : t = 0, 1, 2, \dots, T\}$ that will cause plans which were optimal with respect to preferences \succeq_0 to remain optimal in all future time periods given all that has happened in the interim; (i.e., intermediate consumption experiences and the specific way uncertainty has evolved)? In particular, what utility function representation will guarantee this property?

When considering decision problems over time, such as portfolio investments over a multiperiod horizon, time consistency seems to be a natural property to require. In its absence, one would observe portfolio rebalancing not motivated by any outside event or information flow, but simply resulting from the inconsistency of the date t preference ordering of the investor compared with the preferences on which her original portfolio positions were taken. Asset trades would then be fully motivated by endogenous and unobservable preference issues and would thus be basically unexplainable.

To see what it takes for a utility function to be time consistent, let us consider two periods where at date 1 any one of $s \in S$ possible states of nature may be realized. Let c_0 denote a possible consumption level at date 0, and let $c_1(s)$ denote a possible consumption level in period 1 if state “ s ” occurs. Johnsen and Donaldson (1985) demonstrate that if initial preferences \succeq_0 , with utility representation $U(\cdot)$, are to guarantee time-consistent planning, there must exist continuous and monotone increasing functions $f(\cdot)$ and $\{U_s(\cdot, \cdot) : s \in S\}$ such that

$$U(c_0, c_1(s) : s \in S) = f(c_0, U_s(c_0, c_1(s) : s \in S)), \quad (3.1)$$

where $U_s(\cdot, \cdot)$ is the state s contingent utility function.

This result means the utility function must be of a form such that the utility representations in future states can be recursively nested as individual arguments of the overall utility function. This condition is satisfied by the VNM expected utility form,

$$U(c_0, c_1(s) : s \in S) = U_0(c_0) + \sum_s \pi_s U(c_1(s)),$$

which clearly is of a form satisfying Equation (3.1). The VNM utility representation is thus time consistent, but the latter property can also accommodate more general utility functions. To see this, consider the following special case of Equation (3.1), where there are three possible states at $t = 1$:

$$\begin{aligned}
& U(c_0, c_1(1), c_1(2), c_1(3)) \\
&= \left\{ c_0 + \pi_1 U_1(c_0, c_1(1)) + [\pi_2 U_2(c_0, c_1(2))]^{1/3} \pi_3 U_3(c_0, c_1(3)) \right\}^{1/2}
\end{aligned} \tag{3.2}$$

where

$$\begin{aligned}
U_1(c_0, c_1(1)) &= \log(c_0 + c_1(1)), \\
U_2(c_0, c_1(2)) &= c_0^{1/2} (c_1(2))^{1/2}, \text{ and} \\
U_3(c_0, c_1(3)) &= c_0 c_1(3).
\end{aligned}$$

In this example, preferences are not linear in the probabilities and thus not of the VNM expected utility type. Nevertheless, Equation (3.2) is of the form of Equation (3.1). It also has the feature that preferences in any future state are independent of irrelevant alternatives, where the irrelevant alternatives are those consumption plans for states that do not occur. As such, agents with these preferences will never experience regret and the Allais Paradox will not be operational.

Consistency of choices seems to make sense and turns out to be important for much financial modeling, but is it borne out empirically? Unfortunately, the answer is: frequently not. A simple illustration of this is a typical pure-time preference experiment from the psychology literature (uncertainty in future states is not even needed). Participants are asked to choose among the following monetary prizes:⁹

Question 1: Would you prefer \$100 today or \$200 in 2 years?

Question 2: Would you prefer \$100 in 6 years or \$200 in 8 years?

Respondents often prefer the \$100 in question 1 and the \$200 in question 2, not realizing that question 2 involves the same choice as question 1 but with a 6-year delay. If these people are true to their answers, they will be time inconsistent. In the case of question 2, although they state their preference now for the \$200 prize in 8 years, when year 6 arrives they will take the \$100 and run!

3.7.3 Preferences Defined over Outcomes Other Than Fundamental Payoffs

Under the VNM expected utility theory, the utility function is defined over actual payoff outcomes. Tversky and Kahneman (1992) and Kahneman and Tversky (1979) propose formulations whereby preferences are defined, not over actual payoffs, but rather over gains and losses relative to some benchmark, so that losses are given the greater utility weight. The benchmark can be thought of as either a minimally acceptable payment or, under the proper transformations, a cutoff rate of return. It can be changing through time reflecting prior experience. Their development is called *prospect theory*.

⁹See Ainslie and Haslan (1992) for details.

Insert Figure 3.4 about here

A simple illustration of this sort of representation is as follows: Let \bar{Y} denote the benchmark payoff, and define the investor's utility function $U(Y)$ by

$$U(Y) = \begin{cases} \frac{(|Y - \bar{Y}|)^{1-\gamma_1}}{1-\gamma_1}, & \text{if } Y \geq \bar{Y} \\ -\frac{\lambda(|Y - \bar{Y}|)^{1-\gamma_2}}{1-\gamma_2}, & \text{if } Y \leq \bar{Y} \end{cases}$$

where $\lambda > 1$ captures the extent of the investor's aversion to "losses" relative to the benchmark, and γ_1 and γ_2 need not coincide. In other words, the curvature of the function may differ for deviations above or below the benchmark. Clearly both features could have a large impact on the relative ranking of uncertain investment payoff. See Figure 3.4 for an illustration. Not all economic transactions (e.g., the normal purchase or sale of commodities) are affected by loss aversion since, in normal circumstances, one does not suffer a loss in trading a good. An investor's willingness to hold stocks, however, may be significantly affected if he has experienced losses in prior periods.

3.7.4 Nonlinear Probability Weights

Under the VNM representation, the utility outcomes are linear weighted by their respective probability of outcome. Under prospect theory and its close relatives, this need not be the case: outcomes can be weighted using nonlinear functions of the probabilities and may be asymmetric. More general theories of investor psychology replace the objective mathematical expectation operator entirely with a model of subjective expectations. See Barberis et. al. (1996) for an illustration.

3.8 Conclusions

The expected utility theory is the workhorse of choice theory under uncertainty. It will be put to use almost systematically in this book as it is in most of financial theory. We have argued in this chapter that the expected utility construct provides a straightforward, intuitive mechanism for comparing uncertain asset payoff structures. As such, it offers a well-defined procedure for ranking the assets themselves.

Two ingredients are necessary for this process:

1. An estimate of the probability distribution governing the asset's uncertain payments. While it is not trivial to estimate this quantity, it must also be estimated for the much simpler and less flexible mean/variance criterion.
2. An estimate of the agents' utility of money function; it is the latter that fully characterizes his preference ordering. How this can be identified is one of the topics of the next chapter.

References

- Ainslie, G., Haslam, N. (1992), "Hyperbolic Discounting," in *Choice over Time*, eds. G. Lowenstein and J. Elster, New York: Russell Sage Foundation.
- Allais, M. (1964), "Le comportement de l'homme rationnel devant le risque: Critique des postulats de l'école Américaine," *Econometrica*, 21, 503–546.
- Arrow, K.J. (1963), *Social Choice and Individual Values*, Yale University Press, New Haven, CT.
- Barberis, N., Schleifer, A., Vishney, R. (1998), "A Model of Investor Sentiment" *Journal of Financial Economics*, 49, 307–343.
- Epstein, L., Zin, S. (1989), "Substitution, Risk Aversion, and the Temporal Behavior of Consumption and Asset Returns: A Theoretical Framework," *Econometrica* 57, 937–969.
- Grether, D., Plott, C. (1979), "Economic Theory of Choice and the Preference Reversal Phenomenon," *American Economic Review* 75, 623–638.
- Huberman, G. (2001), "Familiarity Breeds Investment," *Review of Financial Studies*, 14, 659–680.
- Johnsen, T., Donaldson, J. B. (1985), "The Structure of Intertemporal Preferences Under Uncertainty and Time Consistent Plans," *Econometrica* 53, 1451–1458.
- Kahneman, D., Tversky, A. (1979), "Prospect Theory: An Analysis of Decision Under Risk," *Econometrica* 47, 263–291.
- Kreps, D., Porteus, E. (1978), "Temporal Resolution of Uncertainty and Dynamic Choice Theory," *Econometrica* 46, 185–200.
- Machina, M. (1987), "Choice Under Uncertainty: Problems Solved and Unsolved," *Journal of Economic Perspectives* 1, 121–154.
- Mas-Colell, A., Whinston, M. D., Green, J. R. (1995), *Microeconomic Theory*, Oxford University Press, Oxford.
- Thaler, R.H. (1992), *The Winner's Curse*, Princeton University Press, Princeton, NJ.
- Tversky, A., Kahneman, D. (1992), "Advances in Prospect Theory: Cumulative Representation of Uncertainty," *Journal of Risk and Uncertainty* 5, 297–323.
- Weil, P. (1990), "Unexpected Utility in Macroeconomics," *Quarterly Journal of Economics* 105, 29–42.

Chapter 4 : Measuring Risk and Risk Aversion

4.1 Introduction

We argued in Chapter 1 that the desire of investors to avoid risk, that is, to smooth their consumption across states of nature and for that reason avoid variations in the value of their portfolio holdings, is one of the primary motivations for financial contracting. But we have not thus far imposed restrictions on the VNM expected utility representation of investor preferences, which necessarily guarantee such behavior. For that to be the case, our representation must be further specialized.

Since the probabilities of the various state payoffs are objectively given, independently of agent preferences, further restrictions must be placed on the utility-of-money function $U(\cdot)$ if the VNM (von Neumann-Morgenstern - expected utility) representation is to capture this notion of risk aversion. We will now define risk aversion and discuss its implications for $U(\cdot)$.

4.2 Measuring Risk Aversion

What does the term risk aversion imply about an agent's utility function? Consider a financial contract where the potential investor either receives an amount h with probability $\frac{1}{2}$, or must pay an amount h with probability $\frac{1}{2}$. Our most basic sense of risk aversion must imply that for any level of personal wealth Y , a risk-averse investor would not wish to own such a security. In utility terms this must mean

$$U(Y) > \left(\frac{1}{2}\right)U(Y+h) + \left(\frac{1}{2}\right)U(Y-h) = EU,$$

where the expression on the right-hand side of the inequality sign is the VNM expected utility associated with the random wealth levels:

$$\begin{array}{l} y+h, \text{ probability} = \frac{1}{2} \\ y-h, \text{ probability} = \frac{1}{2}. \end{array}$$

This inequality can only be satisfied for all wealth levels Y if the agent's utility function has the form suggested in Figure 4.1. When this is the case we say the utility function is strictly concave.

The important characteristics implied by this and similarly shaped utility functions is that the slope of the graph of the function decreases as the agent becomes wealthier (as Y increases); that is, the marginal utility (MU), represented by the derivative $\frac{d(U(Y))}{d(Y)} \equiv U'(Y)$, decreases with greater Y . Equivalently, for twice differentiable utility functions, $\frac{d^2(U(Y))}{d(Y)^2} \equiv U''(Y) < 0$. For this class of functions, the latter is indeed a necessary and sufficient condition for risk aversion.

Insert Figure 4.1 about here

As the discussion indicates, both consumption smoothing and risk aversion are directly related to the notion of decreasing MU . Whether they are envisaged across time or states, decreasing MU basically implies that income (or consumption) deviations from a fixed average level diminish rather than increase utility. Essentially, the positive deviations do not help as much as the negative ones hurt.

Risk aversion can also be represented in terms of indifference curves. Figure 4.2 illustrates the case of a simple situation with two states of nature. If consuming c_1 in state 1 and c_2 in state 2 represents a certain level of expected utility EU , then the convex-to-the-origin indifference curve that is the appropriate translation of a strictly concave utility function indeed implies that the utility level generated by the average consumption $\frac{c_1+c_2}{2}$ in both states (in this case a certain consumption level) is larger than EU .

Insert Figure 4.2 about here

We would like to be able to measure the degree of an investor's aversion to risk. This will allow us to compare whether one investor is more risk averse than another and to understand how an investor's risk aversion affects his investment behavior (for example, the composition of his portfolio).

As a first attempt toward this goal, and since $U''(\cdot) < 0$ implies risk aversion, why not simply say that investor A is more risk averse than investor B , if and only if $|U''_A(Y)| \geq |U''_B(Y)|$, for all income levels Y ? Unfortunately, this approach leads to the following inconsistency. Recall that the preference ordering described by a utility function is invariant to linear transformations. In other words, suppose $U_A(\cdot)$ and $\bar{U}_A(\cdot)$ are such that $\bar{U}_A(\cdot) = a + bU_A(\cdot)$ with $b > 0$. These utility functions describe the identical ordering, and thus must display identical risk aversion. Yet, if we use the above measure we have

$$|U''_B(Y)| > |U''_A(Y)|, \text{ if, say, } b > 1.$$

This implies that investor A is more risk averse than he is himself, which must be a contradiction.

We therefore need a measure of risk aversion that is invariant to linear transformations. Two widely used measures of this sort have been proposed by, respectively, Pratt (1964) and Arrow (1971):

$$(i) \text{ absolute risk aversion} = -\frac{U''(Y)}{U'(Y)} \equiv R_A(Y)$$

$$(ii) \text{ relative risk aversion} = -\frac{YU''(Y)}{U'(Y)} \equiv R_R(Y).$$

Both of these measures have simple behavioral interpretations. Note that instead of speaking of risk aversion, we could use the inverse of the measures proposed above and speak of risk tolerance. This terminology may be preferable on various occasions.

4.3 Interpreting the Measures of Risk Aversion

4.3.1 Absolute Risk Aversion and the Odds of a Bet

Consider an investor with wealth level Y who is offered — at no charge — an investment involving winning or losing an *amount* h , with probabilities π and $1-\pi$, respectively. Note that any investor will accept such a bet if π is high enough (especially if $\pi = 1$) and reject it if π is small enough (surely if $\pi = 0$). Presumably, the willingness to accept this opportunity will also be related to his level of current wealth, Y . Let $\pi = \pi(Y, h)$ be that probability at which the agent is indifferent between accepting or rejecting the investment. It is shown that

$$\pi(Y, h) \cong 1/2 + (1/4)hR_A(Y), \quad (4.1)$$

where \cong denotes “is approximately equal to.”

The higher his measure of absolute risk aversion, the more favorable odds he will demand in order to be willing to accept the investment. If $R_A^1(Y) \geq R_A^2(Y)$, for agents 1 and 2 respectively, then investor 1 will always demand more favorable odds than investor 2, and in this sense investor 1 is more risk averse.

It is useful to examine the magnitude of this probability. Consider, for example, the family of VNM utility-of-money functions with the form:

$$U(Y) = -\frac{1}{\nu}e^{-\nu Y},$$

where ν is a parameter.

For this case,

$$\pi(Y, h) \cong 1/2 + (1/4)h\nu,$$

in other words, the odds requested are independent of the level of initial wealth (Y); on the other hand, the more wealth at risk (h), the greater the odds of a favorable outcome demanded. This expression advances the parameter ν as the appropriate measure of the degree of *absolute* risk aversion for these preferences.

Let us now derive Equation (4.1). By definition, $\pi(Y, h)$ must satisfy

$$\underbrace{U(Y)}_{\text{utility if he foregoes the bet}} = \underbrace{\pi(Y, h)U(Y + h) + [1 - \pi(Y, h)]U(Y - h)}_{\text{expected utility if the investment is accepted}} \quad (4.2)$$

By an approximation (Taylor’s Theorem) we know that:

$$\begin{aligned} U(Y + h) &= U(Y) + hU'(Y) + \frac{h^2}{2}U''(Y) + H_1 \\ U(Y - h) &= U(Y) - hU'(Y) + \frac{h^2}{2}U''(Y) + H_2, \end{aligned}$$

where H_1, H_2 are remainder terms of order higher than h^2 . Substituting these quantities into Equation (4.2) gives

$$U(Y) = \pi(Y, h)[U(Y) + hU'(Y) + \frac{h^2}{2}U''(Y) + H_1] + (1 - \pi(Y, h))[U(Y) - hU'(Y) + \frac{h^2}{2}U''(Y) + H_2] \quad (4.3)$$

Collecting terms gives

$$U(Y) = U(Y) + (2\pi(Y, h) - 1) \left[hU'(Y) + \frac{h^2}{2}U''(Y) \right] + \underbrace{\pi(Y, h)H_1 + (1 - \pi(Y, h))H_2}_{=\text{def. } H(\text{small})}$$

Solving for $\pi(Y, h)$ yields

$$\pi(Y, h) = \frac{1}{2} + \frac{h}{4} \left[\frac{-U''(Y)}{U'(Y)} \right] - \frac{H}{2hU'(Y)}, \quad (4.4)$$

which is the promised expression, since the last remainder term is small - it is a weighted average of terms of order higher than h^2 and is, thus, itself of order higher than h^2 - and it can be ignored in the approximation.

4.3.2 Relative Risk Aversion in Relation to the Odds of a Bet

Consider now an investment opportunity similar to the one just discussed except that the amount at risk is a *proportion* of the investor's wealth, in other words, $h = \theta Y$, where θ is the fraction of wealth at risk. By a derivation almost identical to the one above, it can be shown that

$$\pi(Y, \theta) \cong \frac{1}{2} + \frac{1}{4}\theta R_R(Y). \quad (4.5)$$

If $R_R^1(Y) \geq R_R^2(Y)$, for investors 1 and 2, then investor 1 will always demand more favorable odds, for any level of wealth, when the fraction θ of his wealth is at risk.

It is also useful to illustrate this measure by an example. A popular family of VNM utility-of-money functions (for reasons to be detailed in the next chapter) has the form:

$$\begin{aligned} U(Y) &= \frac{Y^{1-\gamma}}{1-\gamma}, \text{ for } 0 > \gamma \neq 1 \\ U(Y) &= \ln Y, \text{ if } \gamma = 1. \end{aligned}$$

In the latter case, the probability expression becomes

$$\pi(Y, \theta) \cong \frac{1}{2} + \frac{1}{4}\theta.$$

In this case, the requested odds of winning are not a function of initial wealth (Y) but depend upon θ , the fraction of wealth that is at risk: The lower the fraction θ , the more investors are willing to consider entering into bet that is close to being fair (a risky opportunity where the probabilities of success or

failure are both $\frac{1}{2}$). In the former, more general, case the analogous expression is

$$\pi(Y, \theta) \cong \frac{1}{2} + \frac{1}{4}\theta\gamma.$$

Since $\gamma > 0$, these investors demand a higher probability of success. Furthermore, if $\gamma_2 > \gamma_1$, the investor characterized by $\gamma = \gamma_2$ will always demand a higher probability of success than will an agent with $\gamma = \gamma_1$, for the same fraction of wealth at risk. In this sense a higher γ denotes a greater degree of *relative* risk aversion for this investor class.

4.3.3 Risk Neutral Investors

One class of investors deserves special mention at this point. They are significant, as we shall later see, for the influence they have on the financial equilibria in which they participate. This is the class of investors who are risk neutral and who are identified with utility functions of a linear form

$$U(Y) = cY + d,$$

where c and d are constants and $c > 0$.

Both of our measures of the degree of risk aversion, when applied to this utility function give the same result:

$$R_A(Y) \equiv 0 \text{ and } R_R(Y) \equiv 0.$$

Whether measured as a proportion of wealth or as an absolute amount of money at risk, such investors do not demand better than even odds when considering risky investments of the type under discussion. They are indifferent to risk, and are concerned only with an asset's expected payoff.

4.4 Risk Premium and Certainty Equivalence

The context of our discussion thus far has been somewhat artificial because we were seeking especially convenient probabilistic interpretations for our measures of risk aversion. More generally, a risk-averse agent ($U''(\cdot) < 0$) will always value an investment at something less than the expected value of its payoffs. Consider an investor, with current wealth Y , evaluating an uncertain risky payoff \tilde{Z} . For any distribution function F_z ,

$$U(Y + E\tilde{Z}) \geq E[U(Y + \tilde{Z})]$$

provided that $U''(\cdot) < 0$. This is a direct consequence of a standard mathematical result known as Jensen's inequality.

Theorem 4.1 (Jensen's Inequality):

Let $g(\cdot)$ be a concave function on the interval (a, b) , and \tilde{x} be a random variable such that $\text{Prob} \{\tilde{x} \in (a, b)\} = 1$. Suppose the expectations $E(\tilde{x})$ and $Eg(\tilde{x})$ exist; then

$$E[g(\tilde{x})] \leq g[E(\tilde{x})].$$

Furthermore, if $g(\cdot)$ is strictly concave and $\text{Prob} \{\tilde{x} = E(\tilde{x})\} \neq 1$, then the inequality is strict.

This theorem applies whether the interval (a, b) on which $g(\cdot)$ is defined is finite or infinite and, if a and b are finite, the interval can be open or closed at either endpoint. If $g(\cdot)$ is convex, the inequality is reversed. See De Groot (1970).

To put it differently, if an uncertain payoff is available for sale, a risk-averse agent will only be willing to buy it at a price less than its expected payoff. This statement leads to a pair of useful definitions. The (maximal) certain sum of money a person is willing to pay to acquire an uncertain opportunity defines his **certainty equivalent** (CE) for that risky prospect; the difference between the CE and the expected value of the prospect is a measure of the uncertain payoff's **risk premium**. It represents the maximum amount the agent would be willing to pay to avoid the investment or gamble.

Let us make this notion more precise. The context of the discussion is as follows. Consider an agent with current wealth Y and utility function $U(\cdot)$ who has the opportunity to acquire an uncertain investment \tilde{Z} with expected value $E\tilde{Z}$. The certainty equivalent (to the risky investment \tilde{Z} , $CE(Y, \tilde{Z})$), and the corresponding risk or insurance premium, $\Pi(Y, \tilde{Z})$, are the solutions to the following equations:

$$EU(Y + \tilde{Z}) = U(Y + CE(Y, \tilde{Z})) \quad (4.6a)$$

$$= U(Y + E\tilde{Z} - \Pi(Y, \tilde{Z})) \quad (4.6b)$$

which, implies

$$CE(\tilde{Z}, Y) = E\tilde{Z} - \Pi(Y, \tilde{Z}) \text{ or } \Pi(Y, \tilde{Z}) = E\tilde{Z} - CE(\tilde{Z}, Y)$$

These concepts are illustrated in Figure 4.3.

Insert Figure 4.3 about here

It is intuitively clear that there is a direct relationship between the size of the risk premium and the degree of risk aversion of a particular individual. The link can be made quite easily in the spirit of the derivations of the previous section. For simplicity, the derivation that follows applies to the case of an actuarially fair prospect \tilde{Z} , one for which $E\tilde{Z} = 0$. Using Taylor series approximations we can develop the left-hand side (LHS) and right-hand side (RHS) of the definitional Equations (4.6a) and (4.6b).

LHS:

$$\begin{aligned} EU(Y + \tilde{Z}) &= EU(Y) + E[\tilde{Z}U'(Y)] + E\left[\frac{1}{2}\tilde{Z}^2U''(Y)\right] + EH(\tilde{Z}^3) \\ &= U(Y) + \frac{1}{2}\sigma_z^2U''(Y) + EH(\tilde{Z}^3) \end{aligned}$$

RHS:

$$U(Y - \Pi(Y, \tilde{Z})) = U(Y) - \Pi(Y, \tilde{Z})U'(Y) + H(\Pi^2)$$

or, ignoring the terms of order Z^3 or Π^2 or higher ($EH(\tilde{Z}^3)$ and $H(\Pi^2)$),

$$\Pi(Y, \tilde{Z}) \cong \frac{1}{2}\sigma_z^2 \left(\frac{-U''(Y)}{U'(Y)} \right) = \frac{1}{2}\sigma_z^2 R_A(Y).$$

To illustrate, consider our earlier example in which $U(Y) = \frac{Y^{1-\gamma}}{1-\gamma}$, and suppose $\gamma = 3$, $Y = \$500,000$, and

$$\tilde{Z} = \begin{cases} \$100,000 & \text{with probability} = \frac{1}{2} \\ -\$100,000 & \text{with probability} = \frac{1}{2} \end{cases}$$

For this case the approximation specializes to

$$\Pi(Y, \tilde{Z}) = \frac{1}{2}\sigma_z^2 \frac{\gamma}{Y} = \frac{1}{2}(100,000)^2 \left(\frac{3}{500,000} \right) = \$30,000.$$

To confirm that this approximation is a good one, we must show that:

$$U(Y - \Pi(Y, \tilde{Z})) = U(500,000 - 30,000) = \frac{1}{2}U(600,000) + \frac{1}{2}U(400,000) = EU(Y + \tilde{Z}),$$

or

$$(4.7)^{-2} = \frac{1}{2}(6)^{-2} + \frac{1}{2}(4)^{-2},$$

or

$$.0452694 \cong .04513; \text{ confirmed.}$$

Note also that for this preference class, the insurance premium is directly proportional to the parameter γ .

Can we convert these ideas into statements about rates of return? Let the equivalent risk-free return be defined by

$$U(Y(1 + r_f)) = U(Y + CE(\tilde{Z}, Y)).$$

The random payoff \tilde{Z} can also be converted into a rate of return distribution via $\tilde{Z} = \tilde{r}Y$, or, $\tilde{r} = \tilde{Z}/Y$. Therefore, r_f is defined by the equation

$$U(Y(1 + r_f)) \equiv EU(Y(1 + \tilde{r})).$$

By risk aversion, $E\tilde{r} > r_f$. We thus define the rate of return risk premium Π^r as $\Pi^r = E\tilde{r} - r_f$, or $E\tilde{r} = r_f + \Pi^r$, where Π^r depends on the degree of risk aversion of the agent in question. Let us conclude this section by computing the rate of return premium in a particular case. Suppose $U(Y) = \ln Y$, and that the random payoff \tilde{Z} satisfies

$$\tilde{Z} = \begin{cases} \$100,000 & \text{with probability} = \frac{1}{2} \\ -\$50,000 & \text{with probability} = \frac{1}{2} \end{cases}$$

from a base of $Y = \$500,000$. The risky rate of return implied by these numbers is clearly

$$\tilde{r} = \begin{cases} 20\% & \text{with probability} = \frac{1}{2} \\ -10\% & \text{with probability} = \frac{1}{2} \end{cases}$$

with an expected return of 5%. The certainty equivalent $CE(Y, \tilde{Z})$ must satisfy

$$\ln(500,000 + CE(Y, \tilde{Z})) = \frac{1}{2} \ln(600,000) + \frac{1}{2} \ln(450,000), \text{ or}$$

$$CE(Y, \tilde{Z}) = e^{\frac{1}{2} \ln(600,000) + \frac{1}{2} \ln(450,000)} - 500,000$$

$$CE(Y, \tilde{Z}) = 19,618, \text{ so that}$$

$$(1 + r_f) = \frac{519,618}{500,000} = 1.0392.$$

The rate of return risk premium is thus $5\% - 3.92\% = 1.08\%$. Let us be clear: This rate of return risk premium does not represent a market or equilibrium premium. Rather it reflects personal preference characteristics and corresponds to the premium over the risk-free rate necessary to compensate, utility-wise, a specific individual, with the postulated preferences and initial wealth, for engaging in the risky investment.

4.5 Assessing the Level of Relative Risk Aversion

Suppose that agents' utility functions are of the form $U(Y) = \frac{Y^{1-\gamma}}{1-\gamma}$ class. As noted earlier, a quick calculation informs us that $R_R(Y) \equiv \gamma$, and we say that $U(\cdot)$ is of the constant relative risk aversion class. To get a feeling as to what this measure means, consider the following uncertain payoff:

$$\begin{cases} \$50,000 & \text{with probability } \pi = .5 \\ \$100,000 & \text{with probability } \pi = .5 \end{cases}$$

Assuming your utility function is of the type just noted, what would you be willing to pay for such an opportunity (i.e., what is the certainty equivalent for this uncertain prospect) if your current wealth were Y ? The interest in asking such a question resides in the fact that, given the amount you are willing to pay,

it is possible to infer your coefficient of relative risk aversion $R_R(Y) = \gamma$, provided your preferences are adequately represented by the postulated functional form. This is achieved with the following calculation.

The CE , the maximum amount you are willing to pay for this prospect, is defined by the equation

$$\frac{(Y + CE)^{1-\gamma}}{1-\gamma} = \frac{\frac{1}{2}(Y + 50,000)^{1-\gamma}}{1-\gamma} + \frac{\frac{1}{2}(Y + 100,000)^{1-\gamma}}{1-\gamma}$$

Assuming zero initial wealth ($Y = 0$), we obtain the following sample results (clearly, $CE > 50,000$):

$\gamma = 0$	$CE = 75,000$ (risk neutrality)
$\gamma = 1$	$CE = 70,711$
$\gamma = 2$	$CE = 66,667$
$\gamma = 5$	$CE = 58,566$
$\gamma = 10$	$CE = 53,991$
$\gamma = 20$	$CE = 51,858$
$\gamma = 30$	$CE = 51,209$

Alternatively, if we suppose a current wealth of $Y = \$100,000$ and a degree of risk aversion of $\gamma = 5$, the equation results in a $CE = \$66,532$.

4.6 The Concept of Stochastic Dominance

In response to dissatisfaction with the standard ranking of risky prospects based on mean and variance, a theory of choice under uncertainty with general applicability has been developed. In this section we show that the postulates of expected utility lead to a definition of two weaker alternative concepts of dominance with wider applicability than the concept of state-by-state dominance. These are of interest because they circumscribe the situations in which rankings among risky prospects are preference free, or, can be defined independently of the specific trade-offs (among return, risk, and other characteristics of probability distributions) represented by an agent's utility function.

We start with an illustration. Consider two investment alternatives, Z_1 and Z_2 , with the characteristics outlined in Table 4.1:

Table 4.1: Sample Investment Alternatives

Payoffs	10	100	2000
Prob Z_1	.4	.6	0
Prob Z_2	.4	.4	.2
$EZ_1 = 64, \sigma_{z_1} = 44$			
$EZ_2 = 444, \sigma_{z_2} = 779$			

First observe that under standard mean-variance analysis, these two investments cannot be ranked: Although investment Z_2 has the greater mean, it also has the greater variance. Yet, all of us would clearly prefer to own investment 2. It at least matches investment 1 and has a positive probability of exceeding it.

Insert Figure 4.4 about here

To formalize this intuition, let us examine the cumulative probability distributions associated with each investment, $F_1(\bar{Z})$ and $F_2(\bar{Z})$ where $F_i(\bar{Z}) = \text{Prob}(Z_i \leq \bar{Z})$.

In Figure 4.4 we see that $F_1(\cdot)$ always lies above $F_2(\cdot)$. This observation leads to Definition 4.1.

Definition 4.1:

Let $F_A(\tilde{x})$ and $F_B(\tilde{x})$, respectively, represent the cumulative distribution functions of two random variables (cash payoffs) that, without loss of generality assume values in the interval $[a, b]$. We say that $F_A(\tilde{x})$ first order stochastically dominates (*FSD*) $F_B(\tilde{x})$ if and only if $F_A(x) \leq F_B(x)$ for all $x \in [a, b]$.

Distribution A in effect assigns more probability to higher values of x ; in other words, higher payoffs are more likely. That is, the distribution functions of A and B generally conform to the following pattern: if F_A *FSD* F_B , then F_A is everywhere below and to the right of F_B as represented in Figure 4.5. By this criterion, investment 2 in Figure 4.5 stochastically dominates investment 1. It should, intuitively, be preferred. Theorem 4.2 summarizes our intuition in this latter regard.

Insert Figure 4.5 about here

Theorem 4.2:

Let $F_A(\tilde{x})$, $F_B(\tilde{x})$, be two cumulative probability distributions for random payoffs $\tilde{x} \in [a, b]$. Then $F_A(\tilde{x})$ *FSD* $F_B(\tilde{x})$ if and only if $E_A U(\tilde{x}) \geq E_B U(\tilde{x})$ for all non-decreasing utility functions $U(\cdot)$.

Proof:

See Appendix.

Although it is not equivalent to state-by-state dominance (see Exercise 4.8), *FSD* is an extremely strong condition. As is the case with the former, it is so strong a concept that it induces only a very incomplete ranking among uncertain prospects. Can we find a broader measure of comparison, for instance, which would make use of the hypothesis of risk aversion as well?

Consider the two independent investments in Table 4.2.¹

¹In this example, contrary to the previous one, the two investments considered are statistically independent.

Table 4.2: Two Independent Investments

Investment 3		Investment 4	
Payoff	Prob.	Payoff	Prob.
4	0.25	1	0.33
5	0.50	6	0.33
9	0.25	8	0.33

Which of these investments is better? Clearly, neither investment (first order) stochastically dominates the other as Figure 4.6 confirms. The probability distribution function corresponding to investment 3 is not everywhere below the distribution function of investment 4. Yet, we would probably prefer investment 3. Can we formalize this intuition (without resorting to the mean/variance criterion, which in this case accords with intuition: $ER_4 = 5$, $ER_3 = 5.75$; $\sigma_4 = 2.9$, and $\sigma_3 = 1.9$)? This question leads to a weaker notion of stochastic dominance that explicitly compares distribution functions.

Definition 4.2: Second Order Stochastic Dominance (SSD).

Let $F_A(\tilde{x})$, $F_B(\tilde{x})$, be two cumulative probability distributions for random payoffs in $[a, b]$. We say that $F_A(\tilde{x})$ second order stochastically dominates (SSD) $F_B(\tilde{x})$ if and only if for any x :

$$\int_{-\infty}^x [F_B(t) - F_A(t)] dt \geq 0.$$

(with strict inequality for some meaningful interval of values of t).

The calculations in Table 4.3 reveal that, in fact, investment 3 second order stochastically dominates investment 4 (let $f_i(x)$, $i = 3, 4$, denote the density functions corresponding to the cumulative distribution function $F_i(x)$). In geometric terms (Figure 4.6), this would be the case as long as area B is smaller than area A .

Insert Figure 4.6 about here

As Theorem 4.3 shows, this notion makes sense, especially for risk-averse agents:

Theorem 4.3:

Let $F_A(\tilde{x})$, $F_B(\tilde{x})$, be two cumulative probability distributions for random payoffs \tilde{x} defined on $[a, b]$. Then, $F_A(\tilde{x})$ SSD $F_B(\tilde{x})$ if and only if $E_A U(\tilde{x}) \geq E_B U(\tilde{x})$ for all nondecreasing and concave U .

Proof:

See Laffont (1989), Chapter 2, Section 2.5

Table 4.3: Investment 3 Second Order Stochastically Dominates Investment 4

Values of x	$\int_0^x f_3(t)dt$	$\int_0^x F_3(t)dt$	$\int_0^x f_4(t)dt$	$\int_0^x F_4(t)dt$	$\int_0^x [F_4(t) - F_3(t)]dt$
0	0	0	0	0	0
1	0	0	1/3	1/3	1/3
2	0	0	1/3	2/3	2/3
3	0	0	1/3	1	1
4	.25	.25	1/3	4/3	13/12
5	.75	1	1/3	5/3	2/3
6	.75	1.75	2/3	7/3	7/12
7	.75	2.5	2/3	3	1/2
8	.75	3.25	1	4	3/4
9	.75	4	1	5	1
10	.75	4.75	1	6	5/4
11	.75	5.5	1	7	3/2
12	1	6.5	1	8	3/2
13	1	7.5	1	9	3/2

That is, all risk-averse agents will prefer the second-order stochastically dominant asset. Of course, *FSD* implies *SSD*: If for two investments Z_1 and Z_2 , Z_1 *FSD* Z_2 , then it is also true that Z_1 *SSD* Z_2 . But the converse is not true.

4.7 Mean Preserving Spreads

Theorems 4.2 and 4.3 attempt to characterize the notion of “better/worse” relevant for probability distributions or random variables (representing investments). But there are two aspects to such a comparison: the notion of “more or less risky” and the trade-off between risk and return. Let us now attempt to isolate the former effect by comparing only those probability distributions with identical means. We will then review Theorem 4.3 in the context of this latter requirement.

The concept of more or less risky is captured by the notion of a mean preserving spread. In our context, this notion can be informally stated as follows: Let $f_A(x)$ and $f_B(x)$ describe, respectively, the probability density functions on payoffs to assets A and B . If $f_B(x)$ can be obtained from $f_A(x)$ by removing some of the probability weight from the center of $f_A(x)$ and distributing it to the tails in such a way as to leave the mean unchanged, we say that $f_B(x)$ is related to $f_A(x)$ via a **mean preserving spread**. Figure 4.7 suggests what this notion would mean in the case of normal-type distributions with identical mean, yet different variances.

Insert Figure 4.7 about here

How can this notion be made both more intuitive and more precise? Consider a set of possible payoffs \tilde{x}_A that are distributed according to $F_A(\cdot)$. We further randomize these payoffs to obtain a new random variable \tilde{x}_B according to

$$\tilde{x}_B = \tilde{x}_A + \tilde{z} \quad (4.7)$$

where, for any x_A value, $E(\tilde{z}) = \int z dH_{x_A}(\tilde{z}) = 0$; in other words, we add some pure randomness to \tilde{x}_A . Let $F_B(\cdot)$ be the distribution function associated with \tilde{x}_B . We say that $F_B(\cdot)$ is a mean preserving spread of $F_A(\cdot)$.

A simple example of this is as follows. Let

$$\tilde{x}_A = \begin{cases} 5 & \text{with prob } 1/2 \\ 2 & \text{with prob } 1/2 \end{cases}$$

and suppose

$$\tilde{z} = \begin{cases} +1 & \text{with prob } 1/2 \\ -1 & \text{with prob } 1/2 \end{cases}$$

Then,

$$\tilde{x}_B = \begin{cases} 6 & \text{with prob } 1/4 \\ 4 & \text{with prob } 1/4 \\ 3 & \text{with prob } 1/4 \\ 1 & \text{with prob } 1/4 \end{cases}$$

Clearly, $E\tilde{x}_A = E\tilde{x}_B = 3.5$; we would also all agree that $F_B(\cdot)$ is intuitively riskier.

Our final theorem (Theorem 4.4) relates the sense of a mean preserving spread, as captured by Equation (4.7), to our earlier results.

Theorem 4.4:

Let $F_A(\cdot)$ and $F_B(\cdot)$ be two distribution functions defined on the same state space with identical means. If this is true, the following statements are equivalent:

- (i) $F_A(\tilde{x}) \text{ SSD } F_B(\tilde{x})$
- (ii) $F_B(\tilde{x})$ is a mean preserving spread of $F_A(\tilde{x})$ in the sense of Equation (4.7).

Proof:

See Rothschild and Stiglitz (1970).

But what about distributions that are not stochastically dominant under either definition and for which the mean-variance criterion does not give a relative ranking? For example, consider (independent) investments 5 and 6 in Table 4.4.

In this case we are left to compare distributions by computing their respective expected utilities. That is to say, the ranking between these two investments is preference dependent. Some risk-averse individuals will prefer investment 5 while other risk-averse individuals will prefer investment 6. This is not bad.

Table 4.4: Two Investments; No Dominance

Investment 5		Investment 6	
Payoff	Prob.	Payoff	Prob.
1	0.25	3	0.33
7	0.5	5	0.33
12	0.25	8	0.34

There remains a systematic basis of comparison. The task of the investment advisor is made more complex, however, as she will have to elicit more information on the preferences of her client if she wants to be in position to provide adequate advice.

4.8 Conclusions

The main topic of this chapter was the VNM expected utility representation specialized to admit risk aversion. Two measures of the degree of risk aversion were presented. Both are functions of an investor's current level of wealth and, as such, we would expect them to change as wealth changes. Is there any systematic relationship between $R_A(Y)$, $R_R(Y)$, and Y which it is reasonable to assume?

In order to answer that question we must move away from the somewhat artificial setting of this chapter. As we will see in Chapter 5, systematic relationships between wealth and the measures of absolute and relative risk aversion are closely related to investors' portfolio behavior.

References

- Arrow, K. J. (1971), *Essays in the Theory of Risk Bearing*, Markham, Chicago.
- De Groot, M. (1970), *Optimal Statistical Decisions*, McGraw Hill, New York.
- Laffont, J.-J. (1989), *The Economics of Uncertainty and Information*, MIT Press, Cambridge, MA.
- Pratt, J. (1964), "Risk Aversion in the Small and the Large," *Econometrica* 32, 122–136.
- Rothschild, M., Stiglitz, J.E. (1970), "Increasing Risk: A Definition," *Journal of Economic Theory* 2, 225–243.

Appendix: Proof of Theorem 4.2

\Rightarrow There is no loss in generality in assuming $U(\cdot)$ is differentiable, with $U'(\cdot) > 0$.

Suppose $F_A(x) \text{ FSD } F_B(x)$, and let $U(\cdot)$ be a utility function defined on $[a, b]$ for which $U'(\cdot) > 0$. We need to show that

$$E_A U(\tilde{x}) = \int_a^b U(\tilde{x}) dF_A(\tilde{x}) > \int_a^b U(\tilde{x}) dF_B(\tilde{x}) = E_B U(\tilde{x}).$$

This result follows from integration by parts (recall the relationship $\int_a^b u dv = uv|_a^b - \int_a^b v du$).

$$\begin{aligned} & \int_a^b U(\tilde{x}) dF_A(\tilde{x}) - \int_a^b U(\tilde{x}) dF_B(\tilde{x}) \\ &= U(b)F_A(b) - U(a)F_A(a) - \int_a^b F_A(\tilde{x})U'(\tilde{x})d\tilde{x} - \\ & \quad \left\{ U(b)F_B(b) - U(a)F_B(a) - \int_a^b F_B(\tilde{x})U'(\tilde{x})d\tilde{x} \right\} \\ &= - \int_a^b F_A(\tilde{x})U'(\tilde{x})d\tilde{x} + \int_a^b F_B(\tilde{x})U'(\tilde{x})d\tilde{x}, \\ & \quad (\text{since } F_A(b) = F_B(b) = 1, \text{ and } F_A(a) = F_B(a) = 0) \\ &= \int_a^b [F_B(\tilde{x}) - F_A(\tilde{x})] U'(\tilde{x})d\tilde{x} \geq 0. \end{aligned}$$

The desired inequality follows since, by the definition of FSD and the assumption that the marginal utility is always positive, both terms within the integral are positive. If there is some subset $(c, a) \subset [a, b]$ on which $F_A(x) > F_B(x)$, the final inequality is strict.

\Leftarrow Proof by contradiction. If $F_A(\tilde{x}) \leq F_B(\tilde{x})$ is false, then there must exist an $\bar{x} \in [a, b]$ for which $F_A(\bar{x}) > F_B(\bar{x})$. Define the following nondecreasing function $\hat{U}(x)$ by

$$\hat{U}(x) = \begin{cases} 1 & \text{for } b \geq x > \bar{x} \\ 0 & \text{for } a \leq x < \bar{x} \end{cases}.$$

We'll use integration by parts again to obtain the required contradiction.

$$\begin{aligned} & \int_a^b \hat{U}(\tilde{x}) dF_A(\tilde{x}) - \int_a^b \hat{U}(\tilde{x}) dF_B(\tilde{x}) \\ &= \int_a^b \hat{U}(\tilde{x}) [dF_A(\tilde{x}) - dF_B(\tilde{x})] \\ &= \int_{\bar{x}}^b 1 [dF_A(\tilde{x}) - dF_B(\tilde{x})] \\ &= F_A(b) - F_B(b) - [F_A(\bar{x}) - F_B(\bar{x})] - \int_{\bar{x}}^b [F_A(\tilde{x}) - F_B(\tilde{x})](0)d\tilde{x} \\ &= F_B(\bar{x}) - F_A(\bar{x}) < 0. \end{aligned}$$

Thus we have exhibited an increasing function $\hat{U}(x)$ for which $\int_a^b \hat{U}(\tilde{x}) dF_A(\tilde{x}) < \int_a^b U(\tilde{x}) dF_B(\tilde{x})$, a contradiction. \square

Chapter 5: Risk Aversion and Investment Decisions, Part 1

5.1 Introduction

Chapters 3 and 4 provided a systematic procedure for assessing an investor's relative preference for various investment payoffs: rank them according to expected utility using a VNM utility representation constructed to reflect the investor's preferences over random payments. The subsequent postulate of risk aversion further refined this idea: it is natural to hypothesize that the utility-of-money function entering the investor's VNM index is concave ($U''(\cdot) < 0$). Two widely used measures were introduced and interpreted each permitting us to assess an investor's degree of risk aversion. In the setting of a zero-cost investment paying either $(+h)$ or $(-h)$, these measures were shown to be linked with the minimum probability of success above one half necessary for a risk averse investor to take on such a prospect willingly. They differ only as to whether (h) measures an absolute amount of money or a proportion of the investors' initial wealth.

In this chapter we begin to use these ideas with a view towards understanding an investor's demand for assets of different risk classes and, in particular, his or her demand for risk-free versus risky assets. This is an essential aspect of the investor's portfolio allocation decision.

5.2 Risk Aversion and Portfolio Allocation: Risk Free vs. Risky Assets

5.2.1 The Canonical Portfolio Problem

Consider an investor with wealth level Y_0 , who is deciding what amount, a , to invest in a risky portfolio with uncertain rate of return \tilde{r} . We can think of the risky asset as being, in fact, the market portfolio under the "old" Capital Asset Pricing Model (CAPM), to be reviewed in Chapter 7. The alternative is to invest in a risk-free asset which pays a certain rate of return r_f . The time horizon is one period. The investor's wealth at the end of the period is given by

$$\tilde{Y}_1 = (1 + r_f)(Y_0 - a) + a(1 + \tilde{r}) = Y_0(1 + r_f) + a(\tilde{r} - r_f)$$

The choice problem which he must solve can be expressed as

$$\max_a EU(\tilde{Y}_1) = \max_a EU(Y_0(1 + r_f) + a(\tilde{r} - r_f)), \quad (5.1)$$

where $U(\cdot)$ is his utility-of-money function, and E the expectations operator.

This formulation of the investor's problem is fully in accord with the lessons of the prior chapter. Each choice of a leads to a different uncertain payoff distribution, and we want to find the choice that corresponds to the most preferred

such distribution. By construction of his VNM representation, this is the payoff pattern that maximizes his expected utility.

Under risk aversion ($U''(\cdot) < 0$), the necessary and sufficient first order condition for problem (5.1) is given by:

$$E[U'(Y_0(1+r_f) + a(\tilde{r} - r_f))(\tilde{r} - r_f)] = 0 \quad (5.2)$$

Analyzing Equation (5.2) allows us to describe the relationship between the investor's degree of risk aversion and his portfolio's composition as per the following theorem:

Theorem 5.1:

Assume $U'(\cdot) > 0$, and $U''(\cdot) < 0$ and let \hat{a} denote the solution to problem (5.1). Then

$$\begin{aligned} \hat{a} &> 0 \Leftrightarrow E\tilde{r} > r_f \\ \hat{a} &= 0 \Leftrightarrow E\tilde{r} = r_f \\ \hat{a} &< 0 \Leftrightarrow E\tilde{r} < r_f \end{aligned}$$

Proof: Since this is a fundamental result, its worthwhile to make clear its (straightforward) justification. We follow the argument presented in Arrow (1971), Chapter 2.

Define $W(a) = E\{U(Y_0(1+r_f) + a(\tilde{r} - r_f))\}$. The FOC (5.2) can then be written $W'(a) = E[U'(Y_0(1+r_f) + a(\tilde{r} - r_f))(\tilde{r} - r_f)] = 0$. By risk aversion ($U'' < 0$), $W''(a) = E[U''(Y_0(1+r_f) + a(\tilde{r} - r_f))(\tilde{r} - r_f)^2] < 0$, that is, $W'(a)$ is everywhere decreasing. It follows that \hat{a} will be positive if and only if $W'(0) = U'(Y_0(1+r_f))E(\tilde{r} - r_f) > 0$ (since then a will have to be increased from the value of 0 to achieve equality in the FOC). Since U' is always strictly positive, this implies $\hat{a} > 0$ if and only if $E(\tilde{r} - r_f) > 0$. \square

The other assertion follows similarly.

Theorem 5.1 asserts that a risk averse agent will invest in the risky asset or portfolio only if the expected return on the risky asset exceeds the risk free rate. On the other hand, a risk averse agent will *always* participate (possibly via an arbitrarily small stake) in a risky investment when the odds are favorable.

5.2.2 Illustration and Examples

It is worth pursuing the above result to get a sense of how large a is relative to Y_0 . Our findings will, of course, be preference dependent. Let us begin with the fairly standard and highly tractable utility function $U(Y) = \ln Y$. For added simplicity let us also assume that the risky asset is forecast to pay either of two returns (corresponding to an “up” or “down” stock market), $r_2 > r_1$, with probabilities π and $1-\pi$ respectively. It makes sense (why?) to assume $r_2 > r_f > r_1$ and $E\tilde{r} = \pi r_2 + (1-\pi)r_1 > r_f$.

Under this specification, the F.O.C (5.2) becomes

$$E \left\{ \frac{\tilde{r} - r_f}{Y_0(1 + r_f) + a(\tilde{r} - r_f)} \right\} = 0.$$

Writing out the expectation explicitly yields

$$\frac{\pi(r_2 - r_f)}{Y_0(1 + r_f) + a(r_2 - r_f)} + \frac{(1 - \pi)(r_1 - r_f)}{Y_0(1 + r_f) + a(r_1 - r_f)} = 0,$$

which, after some straightforward algebraic manipulation, gives:

$$\frac{a}{Y_0} = \frac{-(1 + r_f)[E\tilde{r} - r_f]}{(r_1 - r_f)(r_2 - r_f)} > 0. \quad (5.3)$$

This is an intuitive sort of expression: the fraction of wealth invested in risky assets increases with the return premium paid by the risky asset ($E\tilde{r} - r_f$) and decreases with an increase in the return dispersion around r_f as measured by $(r_2 - r_f)(r_f - r_1)$.¹

Suppose $r_f = .05$, $r_2 = .40$, and $r_1 = -.20$ and $\pi = \frac{1}{2}$ (the latter information guarantees $E\tilde{r} = .10$). In this case $\frac{a}{Y_0} = .6$: 60% of the investor's wealth turns out to be invested in the risky asset. Alternatively, suppose $r_2 = .30$ and $r_1 = -.10$ (same r_f , π and $E\tilde{r}$); here we find that $\frac{a}{Y_0} = 1.4$. This latter result must be interpreted to mean that an investor would prefer to invest at least his full wealth in the risky portfolio. If possible, he would even want to borrow an additional amount, equal to 40% of his initial wealth, at the risk free rate and invest this amount in the risky portfolio as well. In comparing these two examples, we see that the return dispersion is much smaller in the second case (lower risk in a mean-variance sense) with an unchanged return premium. With less risk and unchanged mean returns, it is not surprising that the proportion invested in the risky asset increases very substantially. We will see, however, that, somewhat surprisingly, this result does not generalize without further assumption on the form of the investor's preferences.

5.3 Portfolio Composition, Risk Aversion and Wealth

In this section we consider how an investor's portfolio decision is affected by his degree of risk aversion and his wealth level. A natural first exercise is to compare the portfolio composition across individuals of differing risk aversion. The answer to this first question conforms with intuition: if John is more risk averse than Amos, he optimally invests a smaller fraction of his wealth in the risky asset. This is the essence of our next two theorems.

Theorem 5.2 (Arrow, 1971):

Suppose, for all wealth levels Y , $R_A^1(Y) > R_A^2(Y)$ where $R_A^i(Y)$ is the measure of absolute risk aversion of investor i , $i = 1, 2$. Then $\hat{a}_1(Y) < \hat{a}_2(Y)$.

¹That this fraction is independent of the wealth level is not a general result, as we shall find out in Section 5.3.

That is, the more risk averse agent, as measured by his absolute risk aversion measure, will always invest less in the risky asset, given the same level of wealth. This result does not depend on measuring risk aversion via the absolute Arrow-Pratt measure. Indeed, since $R_A^1(Y) > R_A^2(Y) \Leftrightarrow R_R^1(Y) > R_R^2(Y)$, Theorem 5.2 can be restated as

Theorem 5.3:

Suppose, for all wealth levels $Y > 0$, $R_R^1(Y) > R_R^2(Y)$ where $R_R^i(Y)$ is the measure of relative risk aversion of investor $i, i = 1, 2$. Then $\hat{a}_1(Y) < \hat{a}_2(Y)$.

Continuing with the example of Section 5.2.2, suppose now that the investor's utility function has the form $U(Y) = \frac{Y^{1-\gamma}}{1-\gamma}, \gamma > 1$. This utility function displays both greater absolute and greater relative risk aversion than $U(Y) = \ln Y$ (you are invited to prove this statement). From Theorems 5.2 and 5.3, we would expect this greater risk aversion to manifest itself in a reduced willingness to invest in the risky portfolio. Let us see if this is the case.

For these preferences the expression corresponding to (5.3) is

$$\frac{a}{Y_0} = \frac{(1+r_f) \left\{ [(1-\pi)(r_f-r_1)]^{\frac{1}{\gamma}} - (\pi(r_2-r_f))^{\frac{1}{\gamma}} \right\}}{(r_1-r_f) \left\{ \pi(r_2-r_f) \right\}^{\frac{1}{\gamma}} - (r_2-r_f) \left\{ (1-\pi)(r_f-r_1) \right\}^{\frac{1}{\gamma}}} \quad (5.4)$$

In the case of our first example, but with $\gamma = 3$, we obtain, by simple direct substitution,

$$\frac{a}{Y_0} = .24;$$

indeed only 24% of the investor's assets are invested in the risky portfolio, down from 60% earlier.

The next logical question is to ask how the investment in the risky asset varies with the investor's total wealth as a function of his degree of risk aversion. Let us begin with statements appropriate to the absolute measure of risk aversion.

Theorem 5.4 (Arrow, 1971):

Let $\hat{a} = \hat{a}(Y_0)$ be the solution to problem (5.1) above; then:

- (i) $R'_A(Y) < 0 \Leftrightarrow \hat{a}'(Y_0) > 0$
- (ii) $R'_A(Y) = 0 \Leftrightarrow \hat{a}'(Y_0) = 0$
- (iii) $R'_A(Y) > 0 \Leftrightarrow \hat{a}'(Y_0) < 0$.

Case (i) is referred to as declining absolute risk aversion (DARA). Agents with this property become more willing to accept greater bets as they become wealthier. Theorem 5.4 says that such agents will also increase the amount invested in the risky asset ($\hat{a}'(Y_0) > 0$). To state matters slightly differently, an agent with the indicated declining absolute risk aversion will, if he becomes wealthier, be willing to put some of that additional wealth at risk. Utility functions of this form are quite common: those considered in the example,

$U(Y) = \ln Y$ and $U(Y) = \frac{Y^{1-\gamma}}{1-\gamma}$, $\gamma > 0$, display this property. It also makes intuitive sense.

Under constant absolute risk aversion or CARA, case(ii), the amount invested in the risky asset is unaffected by the agent's wealth. This result is somewhat counter-intuitive. One might have expected that a CARA decision maker, in particular one with little risk aversion, would invest some of his or her increase in initial wealth in the risky asset. Theorem 5.4 disproves this intuition. An example of a CARA utility function is $U(Y) = -e^{-\nu Y}$. Indeed,

$$R_A(Y) = \frac{-U''(Y)}{U'(Y)} = \frac{-(-\nu^2)e^{-\nu Y}}{\nu e^{-\nu Y}} = \nu$$

Let's verify the claim of Theorem 5.4 for this utility function. Consider

$$\max_a E \left(-e^{-\nu(Y_0(1+r_f)+a(\tilde{r}-r_f))} \right)$$

The F.O.C. is

$$E \left[\nu (\tilde{r} - r_f) e^{-\nu(Y_0(1+r_f)+a(\tilde{r}-r_f))} \right] = 0$$

Now compute $\frac{da}{dY_0}$; by differentiating the above equation, we obtain:

$$\begin{aligned} E \left[\nu (\tilde{r} - r_f) e^{-\nu(Y_0(1+r_f)+a(\tilde{r}-r_f))} \left(1 + r_f + (\tilde{r} - r_f) \frac{da}{dY_0} \right) \right] &= 0 \\ (1 + r_f) \underbrace{E \left[\nu (\tilde{r} - r_f) e^{-\nu(Y_0(1+r_f)+a(\tilde{r}-r_f))} \right]}_{=0(\text{by the FOC})} + E \left[\underbrace{\nu (\tilde{r} - r_f)^2}_{>0} \frac{da}{dY_0} \underbrace{e^{-\nu(Y_0(1+r_f)+a(\tilde{r}-r_f))}}_{>0} \right] &= 0; \end{aligned}$$

therefore, $\frac{da}{dY_0} \equiv 0$.

For the above preference ordering, and our original two state risky distribution,

$$\hat{a} = \frac{1}{\nu} \left(\frac{1}{r_1 - r_2} \right) \ln \left(\frac{(1-\pi)}{\pi} \left(\frac{r_f - r_1}{r_2 - r_f} \right) \right)$$

Note that in order for \hat{a} to be positive, it must be that

$$0 < \frac{(1-\pi)}{\pi} \left(\frac{r_f - r_1}{r_2 - r_f} \right) < 1.$$

A sufficient condition is that $\pi > 1/2$.

Case (iii) is one with increasing absolute risk aversion (IARA). It says that as an agent becomes wealthier, he reduces his investments in risky assets. This does not make much sense and we will generally ignore this possibility. Note, however, that the quadratic utility function, which is of some significance as we will see later on, possesses this property.

Let us now think in terms of the relative risk aversion measure. Since it is defined for bets expressed as a *proportion of wealth*, it is appropriate to think in terms of *elasticities*, or of how the fraction invested in the risky asset changes as wealth changes. Define $\eta(Y, \hat{a}) = \frac{d\hat{a}/\hat{a}}{dY/Y} = \frac{Y}{\hat{a}} \frac{d\hat{a}}{dY}$, i.e., the wealth elasticity of investment in the risky asset. For example, if $\eta(Y, \hat{a}) > 1$, as wealth Y increases, the **percentage** increase in the amount optimally invested in the risky portfolio exceeds the percentage increase in Y . Or as wealth increases, the **proportion** optimally invested in the risky asset increases. Analogous to Theorem 5.4 is Theorem 5.5:

Theorem 5.5 (Arrow, 1971):

If, for all wealth levels Y ,

- (i) $R'_R(Y) = 0$ (CRRA) then $\eta = 1$
- (ii) $R'_R(Y) < 0$ (DRRA) then $\eta > 1$
- (iii) $R'_R(Y) > 0$ (IRRA) then $\eta < 1$

In his article, Arrow gives support for the hypothesis that the rate of relative risk aversion should be constant and CRRA ≈ 1 . In particular it can be shown that if an investor's utility of wealth is to be bounded above as wealth tends to ∞ , then $\lim_{Y \rightarrow \infty} R'_R(Y) \geq 1$; similarly, if $U(Y)$ is to be bounded below as $Y \rightarrow 0$, then $\lim_{Y \rightarrow 0} R'_R(Y) \leq 1$. These results suggest, that if we wish to assume CRRA, then CRRA = 1 is the appropriate value². Utility functions of the CRRA class include: $U(Y) = \frac{Y^{1-\gamma}}{1-\gamma}$, where $R_R(Y) = \gamma$, $R_A(Y) = \frac{\gamma}{Y}$.

5.4 Special Case of Risk Neutral Investors

As noted in Chapter 4, a risk neutral investor is one who does not care about risk; he ranks investments solely on the basis of their expected returns. The utility function of such an agent is necessarily of the form $U(Y) = c + dY$, c, d constants, $d > 0$ (Check that $U'' = 0$ in this case).

What proportion of his wealth will such an agent invest in the risky asset? The answer is: provided $E\tilde{r} > r_f$ (as we have assumed), **all** of his wealth will be invested in the risky asset. This is clearly seen from the following. Consider the agent's portfolio problem:

$$\begin{aligned} & \max_a E(c + d(Y_0(1 + r_f) + a(\tilde{r} - r_f))) \\ &= \max_a [c + d(Y_0(1 + r_f)) + da(E\tilde{r} - r_f)] \end{aligned}$$

²Note that the above comments also suggest the appropriateness of weakly increasing relative risk aversion as an alternative working assumption.

With $E\tilde{r} > r_f$ and, consequently, $d(E\tilde{r} - r_f) > 0$, this expression is increasing in a . This means that if the risk neutral investor is unconstrained, he will attempt to borrow as much as possible at r_f and reinvest the proceeds in the risky portfolio. He is willing, without bound, to exchange certain payments for uncertain claims of greater expected value. As such he stands willing to absorb all of the economy's financial risk. If we specify that the investor is prevented from borrowing then the maximum will occur at $a = Y_0$

5.5 Risk Aversion and Risky Portfolio Composition

So far we have considered the question of how an investor should allocate his wealth between a risk free asset and a risky asset or portfolio. We now go one step further and ask the following question: when is the composition of the portfolio (i.e., the percentage of the portfolio's value invested in each of the J risky assets that compose it) independent of the agent's wealth level? This question is particularly relevant in light of current investment practices whereby portfolio decisions are usually taken in steps. Step 1, often associated with the label "asset allocation", is the choice of instruments: stocks, bonds and riskless assets (possibly alternative investments as well, such as hedge funds, private equity and real estate); Step 2 is the country or sector allocation decision: here the issue is to optimize not across asset class but across geographical regions or industrial sectors. Step 3 consists of the individual stock picking decisions made on the basis of information provided by financial analysts. The issuing of asset and country/sector allocation "grids" by all major financial institutions, tailored to the risk profile of the different clients, but independent of their wealth levels (and of changes in their wealth), is predicated on the hypothesis that differences in wealth (across clients) and changes in their wealths do not require adjustments in portfolio composition provided risk tolerance is either unchanged or controlled for.

Let us illustrate the issue in more concrete terms; take the example of an investor with invested wealth equal to \$12,000 and optimal portfolio proportions of $a_1 = \frac{1}{2}$, $a_2 = \frac{1}{3}$, and $a_3 = \frac{1}{6}$ (only 3 assets are considered). In other words, this individual's portfolio holdings are \$6,000 in asset 1, \$4,000 in asset 2 and \$2,000 in asset 3. The implicit assumption behind the most common asset management practice is that, were the investor's wealth to double to \$24,000, the new optimal portfolio would naturally be :

$$\begin{aligned} \text{Asset 1: } \frac{1}{2} (\$24,000) &= \$12,000 \\ \text{Asset 2: } \frac{1}{3} (\$24,000) &= \$8,000 \\ \text{Asset 3: } \frac{1}{6} (\$24,000) &= \$4,000. \end{aligned}$$

The question we pose in the present section is: Is this hypothesis supported by theory? The answer is generally no, in the sense that it is only for very specific preferences (utility functions) that the asset allocation is optimally left

unchanged in the face of changes in wealth levels. Fortunately, these specific preferences include some of the major utility representations. The principal result in this regard is as follows:

Theorem 5.6 (Cass and Stiglitz, 1970):

Let the vector $\begin{pmatrix} \hat{a}_1(Y_0) \\ \vdots \\ \hat{a}_J(Y_0) \end{pmatrix}$ denote the amount optimally invested in the J risky assets if the wealth level is Y_0 .

Then $\begin{pmatrix} \hat{a}_1(Y_0) \\ \vdots \\ \hat{a}_J(Y_0) \end{pmatrix} = \begin{pmatrix} a_1 \\ \vdots \\ a_J \end{pmatrix} f(Y_0)$

(for some arbitrary function $f(\cdot)$) if and only if either

- (i) $U'(Y_0) = (\theta Y_0 + \kappa)^\Delta$ or
- (ii) $U'(Y_0) = \xi e^{-\nu Y_0}$

There are, of course, implicit restrictions on the choice of $\theta, \kappa, \Delta, \xi$ and ν to insure, in particular, that $U''(Y_0) < 0$.³

Integrating (i) and (ii), respectively, in order to recover the utility functions corresponding to these marginal utilities, one finds, significantly, that the first includes the CRRA class of functions:

$U(Y_0) = \frac{1}{1-\gamma} Y_0^{(1-\gamma)}$ $\gamma \neq 1$, and $U(Y_0) = \ln(Y_0)$, while the second corresponds to the CARA class:

$$U(Y_0) = \frac{\xi}{-\nu} e^{-\nu Y_0}.$$

In essence, Theorem 5.6 states that it is only in the case of utility functions satisfying constant absolute or constant relative risk aversion preferences (and some generalization of these functions of minor interest) that the relative composition of the risky portion of an investor's optimal portfolio is invariant to changes in his wealth⁴. Only in these cases, should the investor's portfolio composition be left unchanged as invested wealth increases or decreases. It is only with such utility specifications that the standard "grid" approach to portfolio investing is formally justified⁵.

³For (i), we must have **either** $\theta > 0, \Delta < 0$, and Y_0 such that $\theta Y_0 + \kappa \geq 0$ **or** $\theta < 0, \kappa < 0, \Delta > 0$, and $Y_0 \leq -\frac{\kappa}{\theta}$. For (ii), $\xi > 0, -\nu < 0$ and $Y_0 \geq 0$.

⁴As noted earlier, the constant absolute risk aversion class of preferences has the property that the total amount invested in risky assets is invariant to the level of wealth. It is not surprising therefore that the proportionate allocation among the available risky assets is similarly invariant as this theorem asserts.

⁵Theorem 5.6 does not mean, however, that the fraction of initial wealth invested in the risk free asset vs. the risky 'mutual fund' is invariant to changes in Y_0 . The CARA class of preferences discussed in the previous footnote is a case in point.

5.6 Risk Aversion and Savings Behavior

5.6.1 Savings and the Riskiness of Returns

We have thus far considered the relationship between an agent's degree of risk aversion and the composition of his portfolio. A related, though significantly different, question is to ask how an agent's **savings rate** is affected by an increase in the degree of risk facing him. It is to be expected that the answer to this question will be influenced, in a substantial way, by the agent's degree of risk aversion.

Consider first an agent solving the following two period consumption-savings problem:

$$\begin{aligned} \max_s & E\{U(Y_0 - s) + \delta U(s\tilde{R})\}, \\ \text{s.t.} \quad & Y_0 \geq s \geq 0 \end{aligned} \quad (5.5)$$

where Y_0 is initial (period zero) wealth, s is the amount saved and entirely invested in a risky portfolio with uncertain gross risky return, $\tilde{R} = 1 + \tilde{r}$, $U(\cdot)$ is the agent's period utility-of-consumption function, and δ is his subjective discount factor⁶. Note that this is the first occasion where we have explicitly introduced a time dimension into the analysis (i.e., where the important trade-off involves the present vs. the future): the discount rate $\delta < 1$ captures the extent to which the investor values consumption in the future less than current consumption.

The first order condition for this problem (assuming an interior solution) is given by:

$$U'(Y_0 - s) = \delta E\{U'(s\tilde{R})\tilde{R}\} \quad (5.6)$$

It is clear from the above equation that the properties of the return distribution \tilde{R} will influence the optimal level of s . One is particularly interested to know how optimal savings is affected by the riskiness of returns.

To be concrete, let us think of two return distributions \tilde{R}_A , \tilde{R}_B such that \tilde{R}_B is riskier than \tilde{R}_A and $E\tilde{R}_A = E\tilde{R}_B$. From our previous work (Theorem 4.4), we know this can be made precise by stating that \tilde{R}_A SSD \tilde{R}_B or that \tilde{R}_B is a mean-preserving-spread of \tilde{R}_A . In other words, one can write $\tilde{R}_B = \tilde{R}_A + \tilde{\varepsilon}$ where $\tilde{\varepsilon}$ is a zero mean random variable uncorrelated with \tilde{R}_A . Let s_A and s_B be, respectively, the savings out of Y_0 corresponding to the return distributions \tilde{R}_A and \tilde{R}_B . The issue is whether s_A is larger than s_B or if the converse is true. In other words can one predict that a representative risk averse agent *will save more or less* when confronted with riskier returns on his or her savings?

Let us try to think intuitively about this issue. On the one hand, one may expect that more risk will mean a decrease in savings because "a bird in the hand is worth two in the bush". This can be viewed as a substitution effect: A

⁶Note that this $U(c)$ is in principle different from the (indirect) utility of wealth function considered earlier in the chapter. Assuming a single consumption good with price identically equal to one, however, these notions are equivalent.

riskier return can be likened to an increase in the cost of future consumption. A rational individual may then opt to consume more to-day. On the other hand, a risk averse individual may want to increase savings in the face of uncertainty as a precautionary measure, in order to guarantee a minimum standard of living in the future. This reaction indeed is associated with the notion of “precautionary savings”. The reader is invited to verify that this ambiguity is resolved in a mean-variance world in favor of the first argument. In that context, riskier returns imply a decrease in the RHS of Equation (5.6), or a decrease in the expected future marginal utility of consumption weighted by the gross return. For the equality to be restored, consumption today must increase and, consequently, savings must decrease.

It is important to realize, however, that the mean-variance response is not representative of the reactions of all risk averse agents. Indeed, observers seeking to explain the increase in savings registered in many Western countries in the first half of the nineties have regularly pointed to the rising uncertainties surrounding the macroeconomic situation in general and the pace of economic growth in particular⁷. As our discussion suggests, the key technical issue is whether the RHS of equation (5.6) is increased or decreased by an increase in risk. Applying reasoning similar to that used when discussing risk aversion (see Section 5.2), it is easy to see that this issue, in fact, revolves around whether the RHS of (5.6), (i.e., $\delta U'(sR)R \equiv \delta g(R)$), is convex (in which case it increases) or concave (in which case it decreases) in R .

Suppose, to take an extreme case, that the latter is linear in R ; we know that linearity means that the RHS of (5.6) can be written as $\delta E(g(R)) = \delta g(ER)$. But since R_A and R_B have the same mean, this implies that the RHS of Equation (5.6), and consequently optimal savings, are unaffected by an increase in risk. If on the other hand, $g(R)$ is concave, then $E(g(R)) < g(E(R))$; this is the previous case with a concave $g(\cdot)$, upwards deviations from the mean produce smaller changes in the values attained by the function than downwards deviations. It is then the case that the RHS will be decreases as a result of the mean preserving spread on R . The reverse is true if $g(\cdot)$ is convex.

Note that in the all important case where $U(c) = \ln(c)$, $g(R)$ is in fact a constant function of R , with the obvious result that the savings decision is not affected by the increase in the riskiness of the return distribution. This difference in results between two of the workhorses of finance (mean variance and log utility) is worth emphasizing.

Let us now compute the second derivative of $g(R)$ ⁸:

$$g''(R) = 2U''(sR)s + s^2RU'''(sR) \quad (5.7)$$

Using 5.7 one can express the sign of g'' in terms of the relative rate of risk aversion as in

Theorem 5.7 (Rothschild and Stiglitz, 1971):

⁷Which, if they are right, would tend to suggest that “the world is not mean-variance”.

⁸ $g(R) = U'(sR)R \Rightarrow g'(R) = U''(sR)sR + U'(sR)$ and $g''(R)$ as in Equation (5.6). In the log case we get $g(R) = 1/s$ and thus $g'(R) = g''(R) = 0$

Let \tilde{R}_A, \tilde{R}_B be two return distributions with identical means such that \tilde{R}_A SSD \tilde{R}_B , and let s_A and s_B be, respectively, the savings out of Y_0 corresponding to the return distributions \tilde{R}_A and \tilde{R}_B .

If $R'_R(Y) \leq 0$ and $R_R(Y) > 1$, then $s_A < s_B$;

If $R'_R(Y) \geq 0$ and $R_R(Y) < 1$, then $s_A > s_B$.

Proof:

To prove this assertion we need the following Lemma 5.7.

Lemma 5.7 :

$R'_R(Y)$ has the same sign as $-[U'''(Y)Y + U''(Y)(1 + R_R(Y))]$.

Proof : Since $R_R(Y) = \frac{-YU''(Y)}{U'(Y)}$,

$$R'_R(Y) = \frac{[-U'''(Y)Y - U''(Y)]U'(Y) - [-U''(Y)Y]U''(Y)}{[U'(Y)]^2}.$$

Since $U'(Y) > 0$, $R'_R(Y)$ has the same sign as

$$\begin{aligned} & \frac{[-U'''(Y)Y - U''(Y)]U'(Y) - [-U''(Y)Y]U''(Y)}{U'(Y)} \\ &= -U'''(Y)Y - U''(Y) - \left[\frac{-U''(Y)Y}{U'(Y)} \right] U''(Y) \\ &= -\{U'''(Y)Y + U''(Y)[1 + R_R(Y)]\} . \square \end{aligned}$$

Now we can proceed with the theorem. We'll show only the first implication as the second follows similarly. By the lemma, since $R'_R(Y) < 0$,

$$\begin{aligned} -\{U'''(Y)Y + U''(Y)[1 + R_R(Y)]\} &< 0, \text{ or} \\ \{U'''(Y)Y + U''(Y)[1 + R_R(Y)]\} &> 0 \end{aligned}$$

In addition, since $U''(Y) < 0$, and $R_R(Y) > 1$,

$$U'''(Y)Y + U''(Y)(2) > \{U'''(Y)Y + U''(Y)[1 + R_R(Y)]\} > 0.$$

This is true for all Y ; hence

$2U''(sR) + sRU'''(sR) > 0$. Multiplying left and right by $s > 0$, one gets $2U''(sR)s + s^2RU'''(sR) > 0$, which by equation (5.6) implies

$$g''(R) > 0.$$

But by the earlier remarks, this means that $s_A < s_B$ as required. \square

Theorem 5.7 implies that for the class of constant relative risk aversion utility functions, i.e. functions of the form

$$U(c) = (1 - \gamma)^{-1} c^{1-\gamma}$$

($0 < \gamma \neq 1$), an increase in risk increases savings if $\gamma > 1$ and decreases it if $\gamma < 1$, with the $U(c) = \ln(c)$ case being the watershed for which savings is unaffected. For broader classes of utility functions, this theorem provides a partial characterization only, suggesting different investors react differently according to whether they display declining or increasing relative risk aversion.

A more complete characterization of the issue of interest is afforded if we introduce the concept of prudence, first proposed by Kimball (1990). Let

$\mathbf{P}(c) = \frac{-U'''(c)}{U''(c)}$ be a measure of Absolute Prudence, while by analogy with risk aversion,

$\mathbf{P}(c)c = \frac{-cU'''(c)}{U''(c)}$ then measures Relative Prudence. Theorem 5.7 can now be restated as Theorem 5.8:

Theorem 5.8:

Let \tilde{R}_A, \tilde{R}_B be two return distributions such that \tilde{R}_A SSD \tilde{R}_B , and let s_A and s_B be, respectively, the savings out of Y_0 corresponding to the return distributions \tilde{R}_A and \tilde{R}_B . Then,

$$s_A \geq s_B \text{ iff } c \mathbf{P}(c) \leq 2,$$

and conversely,

$$s_A < s_B \text{ iff } c \mathbf{P}(c) > 2$$

i.e., risk averse individuals with Relative Prudence lower than 2 decrease savings while those with Relative Prudence above 2 increase savings in the face of an increase in the riskiness of returns.

Proof:

We have seen that $s_A < s_B$ if and only if $g''(R) > 0$. From Equation (5.6), this means

$$\begin{aligned} sRU'''(sR)/U''(sR) &< s - 2, \text{ or} \\ c\mathbf{P}(c) = \frac{sRU'''(sR)}{-U''(sR)} &> 2, \end{aligned}$$

as claimed. The other part of the proposition is proved similarly. \square

5.6.2 Illustrating Prudence

The relevance of the concept of prudence can be illustrated in the simplest way if we turn to a slightly different problem, where one ignores uncertainty in returns (assuming, in effect, that the net return is identically zero) while asking how savings in period zero is affected by uncertain labor income in period 1. Our remarks in this context are drawn from Kimball (1990).

Let us write the agent's second period labor income, Y , as $Y = \bar{Y} + \tilde{Y}$ where \bar{Y} is the mean labor income and \tilde{Y} measures deviations from the mean

(of course, $E\tilde{Y} = 0$). The simplest form of the decision problem facing the agent is thus:

$$\max_s E\{U(Y_0 - s) + \beta U(s + \bar{Y} + \tilde{Y})\},$$

where $s = s_i$ satisfies the first order condition

$$(i) \quad U'(Y_0 - s_i) = \beta E\{U'(s_i + \bar{Y} + \tilde{Y})\}.$$

It will be of interest to compare the solution s_i to the above FOC with the solution to the analogous decision problem, denoted s_{ii} , in which the uncertain labor income component is absent. The latter FOC is simply

$$(ii) \quad U'(Y_0 - s_{ii}) = \beta U'(s_{ii} + \bar{Y}).$$

The issue once again is whether and to what extent s_i differs from s_{ii} .

One approach to this question, which gives content to the concept of prudence is to ask what the agent would need to be paid (what compensation is required in terms of period 2 income) to ignore labor income risk; in other words, for his first-period consumption and savings decision to be unaffected by uncertainty in labor income. The answer to this question leads to the definition of the *compensating precautionary premium* $\psi = \psi(\bar{Y}, \tilde{Y}, s)$ as the amount of additional second period wealth (consumption) that must be given to the agent in order that the solution to (i) coincides with the solution to (ii). That is, the compensatory precautionary premium $\psi(\bar{Y}, \tilde{Y}, s)$ is defined as the solution of

$$U'(Y_0 - s_{ii}) = \beta E\{U'(s_{ii} + \bar{Y} + \tilde{Y} + \psi(\bar{Y}, \tilde{Y}, s))\}$$

Kimball (1990) proves the following two results.

Theorem 5.9:

Let $U(\cdot)$ be three times continuously differentiable and $\mathbf{P}(s)$ be the index of Absolute Prudence. Then

$$(i) \quad \psi(\bar{Y}, \tilde{Y}, s) \approx 1/2\sigma_{\tilde{Y}}^2 \mathbf{P}(s + \bar{Y})$$

Furthermore, let $U_1(\cdot)$ and $U_2(\cdot)$ be two second period utility functions for which

$$\mathbf{P}_1(s) = \frac{-U_1'''(s)}{U_1''(s)} < \frac{-U_2'''(s)}{U_2''(s)} = \mathbf{P}_2(s), \text{ for all } s.$$

Then

$$(ii) \quad \psi_2(\bar{Y}, \tilde{Y}, s) > \psi_1(\bar{Y}, \tilde{Y}, s) \text{ for all } s, \bar{Y}, \tilde{Y}.$$

Theorem 5.9 (i) shows that investors' precautionary premia are directly proportional to the product of their prudence index and the variance of their uncertain income component, a result analogous to the characterization of the measure of absolute risk aversion obtained in Section 4.3. The result of Theorem 5.9 (ii) confirms the intuition that the more "prudent" the agent, the greater the compensating premium.

5.6.3 The Joint Saving-portfolio Problem

Although for conceptual reasons we have so far distinguished the consumption-savings and the portfolio allocation decisions, it is obvious that the two decisions should really be considered jointly. We now formalize the consumption/savings/portfolio allocation problem:

$$\max_{\{a,s\}} U(Y_0 - s) + \delta E U(s(1 + r_f) + a(\tilde{r} - r_f)), \quad (5.8)$$

where s denotes the total amount saved and a is the amount invested in the risky asset. Specializing the utility function to the form $U(Y) = \frac{Y^{1-\gamma}}{1-\gamma}$ first order conditions for this joint decision problem are

$$\begin{aligned} s &: (Y_0 - s)^{-\gamma}(-1) + \delta E [s(1 + r_f) + a(\tilde{r} - r_f)]^{-\gamma}(1 + r_f) = 0 \\ a &: E [(s(1 + r_f) + a(\tilde{r} - r_f))^{-\gamma}(\tilde{r} - r_f)] = 0 \end{aligned}$$

The first equation spells out the condition to be satisfied at the margin for the savings level – and by corollary, consumption – to be optimal. It involves comparing the marginal utility today with the expected marginal utility tomorrow, with the rate of transformation between consumption today and consumption tomorrow being the product of the discount factor by the gross risk free return. This FOC needs not occupy us any longer here. The interesting element is the solution to the second first order condition: it has the exact same form as Equation (5.2) with the endogenous (optimal) s replacing the exogenous initial wealth level Y_0 . Let us rewrite this equation as

$$s^{-\gamma} E \left[\left((1 + r_f) + \frac{a}{s}(\tilde{r} - r_f) \right)^{-\gamma} (\tilde{r} - r_f) \right] = 0,$$

which implies:

$$E \left[\left((1 + r_f) + \frac{a}{s}(\tilde{r} - r_f) \right)^{-\gamma} (\tilde{r} - r_f) \right] = 0.$$

This equation confirms the lessons of Equations (5.3) and (5.4): For the selected utility function, the proportion of savings invested in the risky asset is independent of s , the amount saved. This is an important result, which does not generalize to other utility functions, but opens up the possibility of a straightforward extension of the savings-portfolio problem to a many period problem. We pursue this important extension in Chapter 14.

5.7 Separating Risk and Time Preferences

In the context of a standard consumption savings problem such as (5.5), let us suppose once again that the agent's period utility function has been specialized to have the standard CRRA form,

$$U(c) = \frac{c^{1-\gamma}}{1-\gamma}, \quad \gamma > 0.$$

For this utility function, the single parameter γ captures not only the agent's sensitivity to risk, but also his sensitivity to consumption variation across time periods and, equivalently, his willingness to substitute consumption in one period for consumption in another. A high γ signals a strong desire for a very smooth intertemporal consumption profile and, simultaneously, a strong reluctance to substitute consumption in one period for consumption in another. To see this more clearly, consider a deterministic version of Problem (5.5) where $\delta < 1$, $\tilde{R} \equiv 1$:

$$\max_{0 \leq s \leq Y_0} \{U(Y_0 - s) + \delta U(s)\}$$

The necessary and sufficient first-order condition is

$$\begin{aligned} -(Y_0 - s)^{-\gamma} + \delta s^{-\gamma} &= 0 \text{ or} \\ \left(\frac{1}{\delta}\right)^{\frac{1}{\gamma}} &= \left(\frac{Y_0 - s}{s}\right). \end{aligned}$$

With $\delta < 1$, as the agent becomes more and more risk averse ($\gamma \mapsto \infty$), $\left(\frac{Y_0 - s}{s}\right) \mapsto 1$; i.e., $c_0 \approx c_1$. For this preference structure, a highly risk averse agent will also seek an intertemporal consumption profile that is very smooth.

We have stressed repeatedly the pervasiveness of the preference for smooth consumption whether across time or across states of nature, and its relationship with the notion of risk aversion. It is time to recognize that while in an atemporal setting a desire for smooth consumption (across states of nature) is the very definition of risk aversion, in a multiperiod environment, risk aversion and intertemporal consumption smoothing need not be equated. After all, one may speak of intertemporal consumption smoothing in a no-risk, deterministic setting, and one may speak of risk aversion in an uncertain, a-temporal environment. The situation considered so far where the same parameter determines both is thus restrictive. Empirical studies, indeed, tend to suggest that typical individuals are more averse to intertemporal substitution (they desire very smooth consumption intertemporally) than they are averse to risk *per se*. This latter fact cannot be captured in the aforementioned, single parameter setting.

Is it possible to generalize the standard utility specification and break this coincidence of time and risk preferences? Epstein and Zin (1989, 1991) answer positively and propose a class of utility functions which allows each dimension to be parameterized separately while still preserving the time consistency property discussed in Section 3.7 of Chapter 3. They provide, in particular, the axiomatic basis for preferences over lotteries leading to the Kreps and Porteus (1978)-like utility representation (see Chapter 3):

$$U_t = U(c_t, c_{t+1}, c_{t+2}, \dots) = W(c_t, CE(\tilde{U}_{t+1})), \quad (5.9)$$

where $CE(\tilde{U}_{t+1})$ denotes the certainty equivalent in terms of period t consumption of the uncertain utility in all future periods. Epstein and Zin (1991) and others (e.g., Weil (1989)) explore the following CES-like specialized version:

$$U(c_t, CE_{t+1}) = \left[(1-\delta)c_t^{\frac{1-\gamma}{\theta}} + \delta CE_{t+1}^{\frac{1-\gamma}{\theta}} \right]^{\frac{\theta}{1-\gamma}}, \quad (5.10)$$

$$\text{with } \theta = \frac{1-\gamma}{1-\frac{1}{\rho}}, 0 < \delta < 1, 1 \neq \gamma > 0, \rho > 0; \text{ or}$$

$$U(c_t, CE_{t+1}) = (1-\delta) \log c_t + \delta \log CE_{t+1}, \gamma = 1, \quad (5.11)$$

where $CE_{t+1} = CE(\tilde{U}_{t+1})$ is the certainty equivalent of future utility and is calculated according to

$$\left[CE(\tilde{U}_{t+1}) \right]^{1-\gamma} = E_t(\tilde{U}_{t+1})^{1-\gamma}, 1 \neq \gamma > 0, \text{ or} \quad (5.12)$$

$$\log CE(\tilde{U}_{t+1}) = E_t(\log \tilde{U}_{t+1}), \gamma = 1. \quad (5.13)$$

Epstein and Zin (1989) show that γ can be viewed as the agent's coefficient of risk aversion. Similarly, when the time preference parameter ρ becomes smaller, the agent becomes less willing to substitute consumption intertemporally. If $\gamma = \frac{1}{\rho}$, recursive substitution to eliminate U_t yields

$$U_t = \left[(1-\delta)E_t \sum_{j=0}^{\infty} \delta^j c_{t+j}^{1-\gamma} \right]^{\frac{1}{1-\gamma}}$$

which represents the same preference as

$$E_t \sum_{j=0}^{\infty} \delta^j c_{t+j}^{1-\gamma},$$

and is thus equivalent to the usual time separable case with CRRA utility.

Although seemingly complex, this utility representation turns out to be surprisingly easy to work with in consumption-savings contexts. We will provide an illustration of its use in Chapters 9 and 14. Note, however, that (5.10) to (5.13) do not lead to an expected utility representation as the probabilities do not enter linearly. If one wants to extricate time and risk preferences, the expected utility framework must be abandoned.

5.8 Conclusions

We have considered, in a very simple context, the relationship between an investor's degree of risk aversion, on the one hand, his desire to save and the composition of his portfolio on the other. Most of the results were intuitively acceptable and that, in itself, makes us more confident of the VNM representation.

Are there any lessons here for portfolio managers to learn? At least three are suggested.

(1) Irrespective of the level of risk, some investment in risky assets is warranted, even for the most risk averse clients (provided $E\tilde{r} > r_f$). This is the substance of Theorem 5.1.

(2) As the value of a portfolio changes significantly, the asset allocation (proportion of wealth invested in each asset class) and the risky portfolio composition should be reconsidered. How that should be done depends critically on the client's attitudes towards risk. This is the substance of Theorems 5.4 to 5.6.

(3) Investors are willing, in general, to pay to reduce income (consumption) risk, and would like to enter into mutually advantageous transactions with institutions less risk averse than themselves. The extreme case of this is illustrated in Section 5.5.

We went on to consider how greater return uncertainty influences savings behavior. On this score and in some other instances, this chapter has illustrated the fact that, somewhat surprisingly, risk aversion is not always a sufficient hypothesis to recover intuitive behavior in the face of risk. The third derivative of the utility function often plays a role. The notion of prudence permits an elegant characterization in these situations.

In many ways, this chapter has aimed at providing a broad perspective allowing to place Modern Portfolio Theory and its underlying assumptions in their proper context. We are now ready to revisit this pillar of modern finance.

References

- Arrow, K. J. (1971), *Essays in the Theory of Risk Bearing*, Markham, Chicago.
- Becker, G., Mulligan, C. (1997), "The Endogenous Determination of Time Preferences," *Quarterly Journal of Economics*, 112, 729-758.
- Cass, D., Stiglitz, J.E. (1970), "The Structure of Investor Preference and Asset Returns and Separability in Portfolio Allocation: A Contribution to the Pure Theory of Mutual Funds", *Journal of Economic Theory*, 2, 122-160.
- Epstein, L.G., Zin, S.E. (1989), "Substitution, Risk Aversion, and the Temporal Behavior of Consumption Growth and Asset Returns I: Theoretical Framework," *Econometrica*, 57, 937-969.
- Epstein, L.G., Zin, S.E. (1991), "Substitution, Risk Aversion, and the Temporal Behavior of Consumption Growth and Asset Returns II: An Empirical Analysis," *Journal of Empirical Economics*, 99, 263-286.
- Hartwick, J. (2000), "Labor Supply Under Wage Uncertainty," *Economics Letters*, 68, 319-325.
- Kimball, M. S. (1990), "Precautionary Savings in the Small and in the Large," *Econometrica*, 58, 53-73.

- Kreps, D., Portier (1978), "Temporal Resolution of Uncertainty and Dynamic Choice Theory", *Econometrica*, 46, 185-200.
- Rothschild, M., Stiglitz, J. (1971), "Increasing Risk II: Its Economic Consequences," *Journal of Economic Theory*, 3, 66-85.
- Weil, Ph. (1989), "The Equity Premium Puzzle and the Riskfree Rate Puzzle," *Journal of Monetary Economics*, 24, 401-421.

Chapter 6: Risk Aversion and Investment Decisions, Part II: Modern Portfolio Theory

6.1 Introduction

In the context of the previous chapter, we encountered the following canonical portfolio problem:

$$\max_a EU(\tilde{Y}_1) = \max_a EU[Y_0(1 + r_f) + a(\tilde{r} - r_f)]. \quad (6.1)$$

Here the portfolio choice is limited to allocating investable wealth, Y_0 , between a risk-free and a risky asset, a being the amount invested in the latter.

Slightly more generally, we can admit N risky assets, with returns $(\tilde{r}_1, \tilde{r}_2, \dots, \tilde{r}_N)$, as in the Cass-Stiglitz theorem. The above problem in this case becomes:

$$\begin{aligned} & \max_{\{a_1, a_2, \dots, a_N\}} EU(Y_0(1 + r_f) + \sum_{i=1}^N a_i(\tilde{r}_i - r_f)) \\ &= \max_{\{w_1, w_2, \dots, w_N\}} EU(Y_0(1 + r_f) + \sum_{i=1}^N w_i Y_0(\tilde{r}_i - r_f)) \end{aligned} \quad (6.2)$$

Equation (6.2) re-expresses the problem with $w_i = \frac{a_i}{Y_0}$, the proportion of wealth invested in the risky asset i , being the key decision variable rather than a_i the money amount invested. The latter expression may further be written as

$$\max_{\{w_1, w_2, \dots, w_N\}} EU \left\{ Y_0 \left[(1 + r_f) + \sum_{i=1}^N w_i(\tilde{r}_i - r_f) \right] \right\} = EU \{ Y_0 [1 + \tilde{r}_P] \} = EU \{ \tilde{Y}_1 \} \quad (6.3)$$

where \tilde{Y}_1 denotes the end of period wealth and \tilde{r}_P the rate of return on the overall portfolio of assets held.

Modern Portfolio Theory (MPT) explores the details of a portfolio choice such as problem 6.3, (i) under the mean-variance utility hypothesis, and (ii) for an arbitrary number of risky investments, with or without a risk-free asset. The goal of this chapter is to review the fundamentals underlying this theory. We first draw the connection between the mean-variance utility hypothesis and our earlier utility development.

6.2 More About Utility Functions

What provides utility? As noted in Chapter 3, financial economics assumes that the ultimate source of consumer's satisfaction is the consumption of the goods and services he is able to purchase.¹ Preference relations and utility functions

¹Of course this doesn't mean that nothing else in life provides utility or satisfaction (!) but the economist's inquiry is normally limited to the realm of market phenomena and economic choices.

are accordingly defined on bundles of consumption goods:

$$u(c_1, c_2, \dots, c_n), \quad (6.4)$$

where the indexing $i = 1, \dots, n$ is across date-state (contingent) commodities: goods characterized not only by their identity as a product or service but also by the time and state in which they may be consumed. States of nature, however, are mutually exclusive. For each date and state of nature (θ) there is a traditional budget constraint

$$p_{1\theta}c_{1\theta} + p_{2\theta}c_{2\theta} + \dots + p_{m\theta}c_{m\theta} \leq Y_\theta \quad (6.5)$$

where the indexing runs across goods for a given state θ ; in other words, the m quantities $c_{i\theta}, i = 1, \dots, m$, and the m prices $p_{i\theta}, i = 1, \dots, m$ correspond to the m goods available in state of nature θ , while Y_θ is the (“end of period”) wealth level available in that same state. We quite naturally assume that the number of goods available in each state is constant.²

In this context, and in some sense summarizing what we did in our last chapter, it is quite natural to think of an individual’s decision problem as being undertaken sequentially, in three steps.

Step 1: The Consumption-Savings Decision

Here, the issue is deciding how much to consume versus how much to save today: how to split period zero income Y_0 between current consumption now C_0 and saving S_0 for consumption in the future where

$$C_0 + S_0 = Y_0.$$

Step 2: The Portfolio Problem

At this second step, the problem is to choose assets in which to invest one’s savings so as to obtain the desired pattern of end-of-period wealth across the various states of nature. This means, in particular, allocating $(Y_0 - C_0)$ between the risk-free and the N risky assets with $(1 - \sum_{i=1}^N w_i)(Y_0 - C_0)$ representing the investment in the risk-free asset, and $(w_1(Y_0 - C_0), w_2(Y_0 - C_0), \dots, w_N(Y_0 - C_0))$, representing the vector of investments in the various risky assets.

Step 3: Tomorrow’s Consumption Choice

Given the realized state of nature and the wealth level obtained, there remains the issue of choosing consumption bundles to maximize the utility function [Equation (6.4)] subject to Equation (6.5) where

$$Y_\theta = (Y_0 - C_0) \left[(1 + r_f) + \sum_{i=1}^N w_i(r_{i\theta} - r_f) \right]$$

and $r_{i\theta}$ denotes the ex-post return to asset i in state θ .

²This is purely formal: if a good is not available in a given state of nature, it is said to exist but with a total economy-wide endowment of the good being zero.

In such problems, it is fruitful to work by backward induction, starting from the end (step 3). Step 3 is a standard microeconomic problem and for our purpose its solution can be summarized by a utility-of-money function $U(Y_\theta)$ representing the (maximum) level of utility that results from optimizing in step 3 given that the wealth available in state θ is Y_θ .

In other words,

$$U(Y_\theta) \equiv_{def} \max_{(c_{1\theta}, \dots, c_{m\theta})} u(c_{1\theta}, \dots, c_{m\theta})$$

$$\text{s.t. } p_{1\theta}c_{1\theta} + \dots + p_{m\theta}c_{m\theta} \leq Y_\theta.$$

Naturally enough, maximizing the expected utility of Y_θ across all states of nature becomes the objective of step 2:

$$\max_{\{w_1, w_2, \dots, w_N\}} EU(\tilde{Y}) = \sum_{\theta} \pi_{\theta} U(Y_{\theta}).$$

Here π_{θ} is the probability of state of nature θ . The end-of-period wealth (a random variable) can now be written as $\tilde{Y} = (Y_0 - C_0)(1 + \tilde{r}_P)$, with $(Y_0 - C_0)$ the initial wealth net of date 0 consumption and $\tilde{r}_P = r_f + \sum_{i=1}^N w_i(\tilde{r}_i - r_f)$ the rate of return on the portfolio of assets in which $(Y_0 - C_0)$ is invested. This brings us back to Equation (6.3).

Clearly with an appropriate redefinition of the utility function,

$$\max EU(\tilde{Y}) = \max EU((Y_0 - C_0)(1 + \tilde{r}_P)) \equiv_{def} \max \hat{U}(\tilde{r}_P)$$

where in all cases the decision variables are portfolio proportions (or amounts) invested in the different available assets. The level of investable wealth, $(Y_0 - C_0)$, becomes a parameter of the $\hat{U}(\cdot)$ representation. Note that restrictions on the form of the utility function do not have the same meaning when imposed on $U(\cdot)$ or on $\hat{U}(\cdot)$, or for that matter on $u(\cdot)$ [(as in Equation (6.4)].

Finally, given the characteristics (e.g., expected return, standard deviation) of the optimally chosen portfolio, the optimal consumption and savings levels can be selected. We are back in step 1 of the decision problem.

From now on in this chapter we shall work with utility functions defined on the overall portfolio's rate of return \tilde{r}_P . This utility index can be further constrained to be a function of the *mean* and *variance* (or standard deviation) of the probability distribution of \tilde{r}_P . This latter simplification can be accepted either as a working approximation or it can be seen as resulting from two further (alternative) hypotheses made within the expected utility framework: It must be assumed that the decision maker's utility function is quadratic or that asset returns are normally distributed.

The main justification for using a mean-variance approximation is its tractability. As already noted, probability distributions are cumbersome to manipulate and difficult to estimate empirically. Summarizing them by their first two moments is appealing and leads to a rich set of implications that can be tested empirically.

Using a simple Taylor series approximation, one can also see that the mean and variance of an agent's wealth distribution are critical to the determination of his expected utility for any distribution. Let \tilde{Y} denote an investor's end period wealth, an uncertain quantity, and $U(\cdot)$ his utility-of-money function. The Taylor series approximation for his utility of wealth around $E(\tilde{Y})$ yields:

$$EU(\tilde{Y}) = U[E(\tilde{Y})] + U'[E(\tilde{Y})][\tilde{Y} - E(\tilde{Y})] + \frac{1}{2}U''[E(\tilde{Y})][\tilde{Y} - E(\tilde{Y})]^2 + H_3 \quad (6.6)$$

$$\text{where } H_3 = \sum_{j=3}^{\infty} \frac{1}{j!} U^{(j)}[E(\tilde{Y})][\tilde{Y} - E(\tilde{Y})]^j.$$

Now let us compute expected utility using this approximation:

$$\begin{aligned} EU(\tilde{Y}) &= U[E(\tilde{Y})] + U'[E(\tilde{Y})] \underbrace{[E(\tilde{Y}) - E(\tilde{Y})]}_{=0} \\ &\quad + \frac{1}{2}U''[E(\tilde{Y})] \underbrace{E[\tilde{Y} - E(\tilde{Y})]^2}_{=\sigma^2(\tilde{Y})} + EH_3 \\ &= U[E(\tilde{Y})] + \frac{1}{2}U''[E(\tilde{Y})]\sigma^2(\tilde{Y}) + EH_3 \end{aligned}$$

Thus if EH_3 is small, at least to a first approximation, $E(\tilde{Y})$ and $\sigma^2(\tilde{Y})$ are central to determining $EU(\tilde{Y})$.

If $U(Y)$ is quadratic, U'' is a constant and, as a result, $EH_3 \equiv 0$, so $E(\tilde{Y})$ and $\sigma^2(\tilde{Y})$ are all that matter. If \tilde{Y} is normally distributed, EH_3 can be expressed in terms of $E(\tilde{Y})$ and $\sigma^2(\tilde{Y})$, so the approximation is exact in this case as well. These well-known assertions are detailed in Appendix 6.1 where it is also shown that, under either of the above hypotheses, indifference curves in the mean-variance space are increasing and convex to the origin.

Assuming the utility objective function is quadratic, however, is not fully satisfactory since the preference representation would then possess an attribute we deemed fairly implausible in Chapter 4, increasing absolute risk aversion (IARA). On this ground, supposing all or most investors have a quadratic utility function is very restrictive. The normality hypothesis on the rate of return processes is easy to verify directly, but we know it cannot, in general, be satisfied exactly. Limited liability instruments such as stocks can pay at worst a negative return of -100% (complete loss of the investment). Even more clearly at odds with the normality hypothesis, default-free (government) bonds always yield a positive return (abstracting from inflation). Option-based instruments, which are increasingly prevalent, are also characterized by asymmetric probability distributions. These remarks suggest that our analysis to follow must be viewed as a (useful and productive) approximation.

Box 6.1 About the Probability Distribution on Returns

As noted in the text, the assumption that period returns (e.g., daily, monthly, annual) are normally distributed is inconsistent with the limited liability feature of most financial instruments; i.e., $\tilde{r}_{it} \geq -1$ for most securities i . It furthermore presents a problem for computing compounded cumulative returns: The product of normally distributed random variables (returns) is not itself normally distributed.

These objections are made moot if we assume that it is the continuously compounded rate of return, r_{it}^c , that is normally distributed where $\tilde{r}_{it}^c = \log(1 + \tilde{r}_{it})$.

This is consistent with limited liability since $Y_0 e^{\tilde{r}_{it}^c} \geq 0$ for any $\tilde{r}_{it}^c \in (-\infty, +\infty)$. It has the added feature that cumulative continuously compounded returns are normally distributed since the sum of normally distributed random variables is normally distributed.

The working assumption in empirical financial economics is that continuously compounded equity returns are i.i.d. normal; in other words, for all times t ,

$$\tilde{r}_{it}^c \approx N(\mu_i, \sigma_i).$$

By way of language we say that the discrete period returns \tilde{r}_{it}^c are lognormally distributed because their logarithm is normally distributed. There is substantial statistical evidence to support this assumption, subject to a number of qualifications, however.

1. First, while the normal distribution is perfectly symmetric about its mean, daily stock returns are frequently skewed to the right. Conversely, the returns to certain stock indices appear skewed to the left.³

2. Second, the sample daily return distributions for many individual stocks exhibit “excess kurtosis” or “fat tails”; i.e., there is more probability in the tails than would be justified by the normal distribution. The same is true of stock indices. The extent of this excess kurtosis diminishes substantially, however, when monthly data is used.⁴

Figure 6.1 illustrates for the returns on the Dow Jones and the S&P500. Both indices display negative skewness and a significant degree of kurtosis.

There is one further complication. Even if individual stock returns are log-normally distributed, the returns to a portfolio of such stocks need not be log-normal (log of a sum is not equal to the sum of the logs). The extent of the error introduced by assuming lognormal portfolio returns is usually not great if the return period is short (e.g., daily). \square

³Skewness: The extent to which a distribution is “pushed left or right” off symmetry is measured by the skewness statistic $S(\tilde{r}_{it})$, defined by $S(\tilde{r}_{it}) = E \frac{(r_{it} - \mu_i)^3}{\sigma_i^3}$. $S(\tilde{r}_{it}) \equiv 0$ if \tilde{r}_{it} is normally distributed. $S(\tilde{r}_{it}) > 0$ suggests a rightward bias, and conversely if $S(\tilde{r}_{it}) < 0$.

⁴Kurtosis is measured as the normalized fourth moment: $K(\tilde{r}_{it}) = E \frac{(r_{it} - \mu_i)^4}{\sigma_i^4}$. If \tilde{r}_{it} is normal then $K(\tilde{r}_{it}) = 3$, but fat-tailed distribution with extra probability weight in the tail areas have higher kurtosis measures.

Note also that in this chapter we are assuming the investor knows and takes as exogenously given the vector of return random variables $(\tilde{r}_1, \tilde{r}_2, \dots, \tilde{r}_N)$. We have not yet discussed how these rate of return distributions are determined. We are at the stage of identifying demand curves, and not yet attempting to describe equilibrium prices or returns.

Insert Figure 6.1 about here

6.3 Description of the Opportunity Set in the Mean-Variance Space: The Gains from Diversification and the Efficient Frontier

The main idea of this section is as follows: The expected return to a portfolio is the *weighted* average of the expected returns of the assets composing the portfolio. The same result is not generally true for the variance. The variance of a portfolio is generally *smaller* than the weighted average of the variances of individual asset returns corresponding to this portfolio. Therein lies the gain from diversification.

Let us illustrate this assertion, starting with the case of a portfolio of two assets only. The typical investor's objective is to maximize a function $U(\mu_P, \sigma_P)$, where $U_1 > 0$ and $U_2 < 0$: he likes expected return (μ_P) and dislikes standard deviation (σ_P). In this context, one recalls that an asset (or portfolio) A is said to **mean-variance dominate** an asset (or portfolio) B if $\mu_A \geq \mu_B$ and simultaneously $\sigma_A < \sigma_B$, or if $\mu_A > \mu_B$ while $\sigma_A \leq \sigma_B$. We can then define the **efficient frontier** as the locus of all non-dominated portfolios in the mean-standard deviation space. By definition, no ("rational") mean-variance investor would choose to hold a portfolio not located on the efficient frontier. The shape of the efficient frontier is thus of primary interest.

Let us examine the efficient frontier in the two-asset case for a variety of possible asset return correlations. The basis for the results of this section is the formula for the variance of a portfolio of two assets, 1 and 2, defined by their respective expected returns, \bar{r}_1, \bar{r}_2 , standard deviations, σ_1 and σ_2 , and their correlation $\rho_{1,2}$:

$$\sigma_P^2 = w_1^2 \sigma_1^2 + (1 - w_1)^2 \sigma_2^2 + 2w_1(1 - w_1) \sigma_1 \sigma_2 \rho_{1,2},$$

where w_i is the proportion of the portfolio allocated to asset i . The following results, detailed in Appendix 6.2, are of importance.

Case 1 (Reference).

In the case of two risky assets with perfectly positively correlated returns, the efficient frontier is linear. In that extreme case the two assets are essentially identical, there is no gain from diversification, and the portfolio's standard deviation is nothing other than the average of the standard deviations of the

component assets:

$$\sigma_R = w_1\sigma_1 + (1 - w_1)\sigma_2.$$

As a result, the equation of the efficient frontier is

$$\mu_R = \bar{r}_1 + \frac{\bar{r}_2 - \bar{r}_1}{\sigma_2 - \sigma_1} (\sigma_R - \sigma_1),$$

as depicted in Figure 6.2. It assumes positive amounts of both assets are held.

Insert Figure 6.2

Case 2.

In the case of two risky assets with imperfectly correlated returns, the standard deviation of the portfolio is necessarily smaller than it would be if the two component assets were perfectly correlated. By the previous result, one must have $\sigma_P < w_1\sigma_1 + (1 - w_1)\sigma_2$, provided the proportions are not 0 or 1. Thus, the efficient frontier must stand left of the straight line in Figure 6.2. This is illustrated in Figure 6.3 for different values of $\rho_{1,2}$.

Insert Figure 6.3

The smaller the correlation (the further away from +1), the more to the left is the efficient frontier as demonstrated formally in Appendix 6.2. Note that the diagram makes clear that in this case, some portfolios made up of assets 1 and 2 are, in fact, dominated by other portfolios. Unlike in Case 1, not all portfolios are efficient. In view of future developments, it is useful to distinguish the **minimum variance frontier** from the efficient frontier. In the present case, all portfolios between *A* and *B* belong to the minimum variance frontier, that is, they correspond to the combination of assets with minimum variance for all arbitrary levels of expected returns. However, certain levels of expected returns are not efficient targets since higher levels of returns can be obtained for identical levels of risk. Thus portfolio *C* is minimum variance, but it is not efficient, being dominated by portfolio *D*, for instance. Figure 6.3 again assumes positive amounts of both assets (*A* and *B*) are held.

Case 3.

If the two risky assets have returns that are perfectly negatively correlated, one can show that the minimum variance portfolio is risk free while the frontier is once again linear. Its graphical representation in that case is in Figure 6.4, with the corresponding demonstration placed in Appendix 6.2.

Insert Figure 6.4

Case 4.

If one of the two assets is risk free, then the efficient frontier is a straight line originating on the vertical axis at the level of the risk-free return. In the

absence of a short sales restriction, that is, if it is possible to borrow at the risk-free rate to leverage one's holdings of the risky asset, then, intuitively enough, the overall portfolio can be made riskier than the riskiest among the existing assets. In other words, it can be made riskier than the one risky asset and it must be that the efficient frontier is projected to the right of the (\bar{r}_2, σ_2) point (defining asset 1 as the risk-free asset). This situation is depicted in Figure 6.5 with the corresponding results demonstrated in Appendix 6.2.

Insert Figure 6.5

Case 5 (n risky assets).

It is important to realize that a portfolio is also an asset, fully defined by its expected return, its standard deviation, and its correlation with other existing assets or portfolios. Thus, the previous analysis with two assets is more general than it appears: It can easily be repeated with one of the two assets being a portfolio. In that way, one can extend the analysis from two to three assets, from three to four, etc. If there are n risky, imperfectly correlated assets, then the efficient frontier will have the bullet shape of Figure 6.6. Adding an extra asset to the two-asset framework implies that the diversification possibilities are improved and that, in principle, the efficient frontier is displaced to the left.

Case 6.

If there are n risky assets and a risk-free one, the efficient frontier is a straight line once again. To arrive at this conclusion, let us arbitrarily pick one portfolio on the efficient frontier when there are n risky assets only, say portfolio E in Figure 6.6, and make up all possible portfolios combining E and the risk-free asset.

Insert Figure 6.6

What we learned above tells us that the set of such portfolios is the straight line joining the point $(0, r_f)$ to E . Now we can quickly check that all portfolios on this line are dominated by those we can create by combining the risk-free asset with portfolio F . Continuing our reasoning in this way and searching for the highest similar line joining $(0, r_f)$ with the risky asset bullet-shaped frontier, we obtain, as the truly efficient frontier, the straight line originating from $(0, r_f)$ that is tangent to the risky asset frontier. Let T be the tangency portfolio. As before, if we allow short position in the risk-free asset, the efficient frontier extends beyond T ; it is represented by the broken line in Figure 6.6.

Formally, with n assets (possibly one of them risk free), the efficient frontier is obtained as the relevant (non-dominated) portion of the minimum variance frontier, the latter being the solution, for all possible expected returns μ , to the following quadratic program (QP):

$$\begin{aligned}
(QP) \quad & \min_{w_i} \sum_i \sum_j w_i w_j \sigma_{ij} \\
& s.t. \sum_i w_i \bar{r}_i = \mu \\
& \sum_i w_i = 1
\end{aligned}$$

In (QP) we search for the vector of weights that minimizes the variance of the portfolio (verify that you understand the writing of the portfolio variance in the case of n assets) under the constraint that the expected return on the portfolio must be μ . This defines one point on the minimum variance frontier. One can then change the fixed value of μ equating it successively to all plausible levels of portfolio expected return; in this way one effectively draws the minimum variance frontier⁵. Program (QP) is the simplest version of a family of similar quadratic programs used in practice. This is because (QP) includes the minimal set of constraints. The first is only an artifice in that it defines the expected return to be reached in a context where μ is a parameter; the second constraint is simply the assertion that the vector of w_i 's defines a portfolio (and thus that they add up to one).

Many other constraints can be added to customize the portfolio selection process without altering the basic structure of problem (QP) . Probably the most common implicit or explicit constraint for an investor involves limiting her investment universe. The well-known *home bias puzzle* reflects the difficulty in explaining, from the MPT viewpoint, why investors do not invest a larger fraction of their portfolios in stocks quoted “away from home,” that is, in international, or emerging markets. This can be viewed as the result of an unconscious limitation of the investment universe considered by the investor. Self-limitation may also be fully conscious and explicit as in the case of “ethical” mutual funds that exclude arms manufacturers or companies with a tarnished ecological record from their investment universe. These constraints are easily accommodated in our setup, as they simply appear or do not appear in the list of the N assets under consideration.

Other common constraints are non-negativity constraints ($w_i \geq 0$), indicating the impossibility of short selling some or all assets under consideration. Short selling may be impossible for feasibility reasons (exchanges or brokers may not allow it for certain instruments) or, more frequently, for regulatory reasons applying to specific types of investors, for example, pension funds.

An investor may also wish to construct an efficient portfolio subject to the constraint that his holdings of some stocks should not, in value terms, fall below a certain level (perhaps because of potential tax liabilities or because ownership of a large block of this stock affords some degree of managerial control). This requires a constraint of the form

$$w_j \geq \frac{V_j}{V_P},$$

⁵While in principle one could as well maximize the portfolio's expected return for given levels of standard deviation, it turns out to be more efficient computationally to do the reverse.

where V_j is the current value of his holdings of stock j and V_P is the overall value of his portfolio.

Other investors may wish to obtain the lowest risk subject to a required expected return constraint and/or be subject to a constraint that limits the number of stocks in their portfolio (in order, possibly, to economize on transaction costs). An investor may, for example, wish to hold at most 3 out of a possible 10 stocks, yet to hold those 3 which give the minimum risk subject to a required return constraint. With certain modifications, this possibility can be accommodated into (QP) as well. Appendix 6.3 details how Microsoft Excel[®] can be used to construct the portfolio efficient frontier under these and other constraints.

6.4 The Optimal Portfolio: A Separation Theorem

The optimal portfolio is naturally defined as that portfolio maximizing the investor's (mean-variance) utility; in other words, that portfolio for which he is able to reach the highest indifference curve, which we know to be increasing and convex from the origin. If the efficient frontier has the shape described in Figure 6.5, that is, if there is a risk-free asset, then all tangency points must lie on the same efficient frontier, irrespective of the rate of risk aversion of the investor. Let there be two investors sharing the same perceptions as to expected returns, variances, and return correlations but differing in their willingness to take risks. The relevant efficient frontier will be identical for these two investors, although their optimal portfolios will be represented by different points on the same line: with differently shaped indifference curves the tangency points must differ. See Figure 6.7.

Insert Figure 6.7

However, it is a fact that our two investors will invest in the same two funds, the risk-free asset on the one hand, and the risky portfolio (T) identified by the tangency point between the straight line originating from the vertical axis and the bullet-shaped frontier of risky assets, on the other. This is the **two-fund theorem**, also known as the **separation theorem**, because it implies the optimal portfolio of risky assets can be identified separately from the knowledge of the risk preference of an investor. This result will play a significant role in the next Chapter when constructing the Capital Asset Pricing Model.

6.5 Conclusions

First, it is important to keep in mind that everything said so far applies regardless of the (possibly normal) probability distributions of returns representing the *subjective* expectations of the particular investor upon whom we are focusing. Market equilibrium considerations are next.

Second, although initially conceived in the context of descriptive economic theories, the success of portfolio theory arose primarily from the possibility of giving it a normative interpretation; that is, of seeing the theory as providing a guide on how to proceed to identify a potential investor's optimal portfolio. In particular it points to the information requirements to be fulfilled (ideally). Even if we accept the restrictions implied by mean-variance analysis, one cannot identify an optimal portfolio without spelling out expectations on mean returns, standard deviations of returns, and correlations among returns. One can view the role of the financial analyst as providing plausible figures for the relevant statistics or offering alternative scenarios for consideration to the would-be investor. This is the first step in the search for an optimal portfolio. The computation of the efficient frontier is the second step and it essentially involves solving the quadratic programming problem (QP) possibly in conjunction with constraints specific to the investor. The third and final step consists of defining, at a more or less formal level, the investor's risk tolerance and, on that basis, identifying his optimal portfolio.

References

- Markowitz, H .M. (1952), "Portfolio Selection," *Journal of Finance* 7, 77–91.
- Tobin, J. (1958), "Liquidity Preference as Behavior Towards Risk," *Review of Economic Studies* 26, 65–86.

Appendix 6.1: Indifference Curves Under Quadratic Utility or Normally Distributed Returns

In this appendix we demonstrate more rigorously that if his utility function is quadratic or if returns are normally distributed, the investor's expected utility of the portfolio's rate of return is a function of the portfolio's mean return and standard deviation only (Part I). We subsequently show that in either case, investor's indifference curves are convex to the origin (Part II).

Part I

If the utility function is **quadratic**, it can be written as:

$U(\tilde{r}_P) = a + b\tilde{r}_P + c\tilde{r}_P^2$ where \tilde{r}_P denotes a portfolio's rate of return. Let the constant $a = 0$ in what follows since it does not play any role. For this function to make sense we must have $b > 0$ and $c < 0$. The first and second derivatives are, respectively,

$$U'(\tilde{r}_P) = b + 2c\tilde{r}_P, \text{ and } U''(\tilde{r}_P) = 2c < 0$$

Expected utility is then of the following form:

$$E(U(\tilde{r}_P)) = bE(\tilde{r}_P) + c(E(\tilde{r}_P^2)) = b\mu_P + c\mu_P^2 + c\sigma_P^2,$$

that is, of the form $g(\sigma_P, \mu_P)$. As showed in Figure (A6.1). this function is strictly concave. But it must be restricted to ensure positive marginal utility:

$\tilde{r}_P < -b/2c$. Moreover, the coefficient of absolute risk aversion is increasing ($R'_A > 0$). These two characteristics are unpleasant and they prevent a more systematic use of the quadratic utility function.

Insert Figure A6.1

Alternatively, if the individual asset returns \tilde{r}_i are normally distributed, $\tilde{r}_P = \sum_i w_i \tilde{r}_i$ is normally distributed as well. Let \tilde{r}_P have density $f(\tilde{r}_P)$, where

$$f(\tilde{r}_P) = N(\tilde{r}_P; \mu_P, \sigma_P)$$

The standard normal variate \tilde{Z} is defined by:

$$\tilde{Z} = \frac{\tilde{r}_P - \mu_P}{\sigma_P} \sim N(\tilde{Z}; 0, 1)$$

$$\text{Thus, } \tilde{r}_P = \sigma_P \tilde{Z} + \mu_P \quad (6.7)$$

$$E(U(\tilde{r}_P)) = \int_{-\infty}^{+\infty} U(r_P) f(r_P) dr_P = \int_{-\infty}^{+\infty} U(\sigma_P Z + \mu_P) N(Z; 0, 1) dZ. \quad (6.8)$$

The quantity $E(U(\tilde{r}_P))$ is again a function of σ_P and μ_P only. Maximizing $E(U(\tilde{r}_P))$ amounts to choosing w_i so that the corresponding σ_P and μ_P maximize the integral (6.8).

Part II

Construction of indifference curves in the mean-variance space. There are again two cases.

U is quadratic.

An indifference curve in the mean-variance space is defined as the set:

$$\{(\sigma_P, \mu_P) | E(U(\tilde{r}_P)) = b\mu_P + c\mu_P^2 + c\sigma_P^2 = k\}, \text{ for some utility level } k.$$

This can be rewritten as

$$\sigma_P^2 + \mu_P^2 + \frac{b}{c}\mu_P + \frac{b^2}{4c^2} = \frac{k}{c} + \frac{b^2}{4c^2}$$

$$\sigma_P^2 + \left(\mu_P + \frac{b}{2c}\right)^2 = \frac{k}{c} + \frac{b^2}{4c^2}$$

This equation defines the set of points (σ_P, μ_P) located in the circle of radius $\sqrt{\frac{k}{c} + \frac{b^2}{4c^2}}$ and of center $(0, -\frac{b}{2c})$ as in Figure A6.2.

Insert Figure A6.2.

In the relevant portion of the (σ_P, μ_P) space, indifference curves thus have positive slope and are convex to the origin.

The Distribution of \mathbf{R} is Normal.

One wants to describe

$$\{(\sigma_P, \mu_P) \mid \int_{-\infty}^{+\infty} U(\sigma_P Z + \mu_P) N(Z; 0, 1) dZ = \bar{U}\}$$

Differentiating totally yields:

$$\begin{aligned} 0 &= \int_{-\infty}^{+\infty} U'(\sigma_P Z + \mu_P) (Z d\sigma_P + d\mu_P) N(Z; 0, 1) dZ, \text{ or} \\ \frac{d\mu_P}{d\sigma_P} &= - \frac{\int_{-\infty}^{+\infty} U'(\sigma_P Z + \mu_P) Z N(Z; 0, 1) dZ}{\int_{-\infty}^{+\infty} U'(\sigma_P Z + \mu_P) N(Z; 0, 1) dZ}. \end{aligned}$$

If $\sigma_P = 0$ (at the origin),

$$\frac{d\mu_P}{d\sigma_P} = - \frac{\int_{-\infty}^{+\infty} Z N(Z; 0, 1) dZ}{\int_{-\infty}^{+\infty} N(Z; 0, 1) dZ} = 0.$$

If $\sigma_P > 0$, $d\mu_P/d\sigma_P > 0$.

Indeed, the denominator is positive since $U'(\cdot)$ is positive by assumption, and $N(Z; 0, 1)$ is a probability density function, hence it is always positive.

The expression $\int_{-\infty}^{+\infty} U'(\sigma_P Z + \mu_P) Z N(Z; 0, 1) dZ$ is negative under the hypothesis that the investor is risk averse; in other words, that $U(\cdot)$ is strictly concave. If this hypothesis is verified, the marginal utility associated with each negative value of Z is larger than the marginal utility associated with positive values. Since this is true for all pairs of $\pm Z$, the integral on the numerator is negative. See Figure A.6.3 for an illustration.

Proof of the Convexity of Indifference Curves

Let two points (σ_P, μ_P) and $(\sigma_{P'}, \mu_{P'})$ lie on the same indifference curve offering the same level of expected utility \bar{U} . Let us consider the point $(\sigma_{P''}, \mu_{P''})$ where: $\sigma_{P''} = \alpha\sigma_P + (1 - \alpha)\sigma_{P'}$ and $\mu_{P''} = \alpha\mu_P + (1 - \alpha)\mu_{P'}$.

One would like to prove that: $E(U(\sigma_{P''}\tilde{Z} + \mu_{P''})) > \alpha E(U(\sigma_P\tilde{Z} + \mu_P)) + (1 - \alpha)E(U(\sigma_{P'}\tilde{Z} + \mu_{P'})) = \bar{U}$.

By the strict concavity of U , the inequality

$$U(\sigma_{P''}\tilde{Z} + \mu_{P''}) > \alpha U(\sigma_P\tilde{Z} + \mu_P) + (1 - \alpha)U(\sigma_{P'}\tilde{Z} + \mu_{P'})$$

is verified for all (σ_P, μ_P) and $(\sigma_{P'}, \mu_{P'})$.
 One may thus write:

$$\begin{aligned}
 & \int_{-\infty}^{+\infty} U(\sigma_{P''}Z + \mu_{P''})N(Z; 0, 1)dZ > \\
 & \alpha \int_{-\infty}^{+\infty} U(\sigma_PZ + \mu_P)N(Z; 0, 1)dZ + (1 - \alpha) \int_{-\infty}^{+\infty} U(\sigma_{P'}Z + \mu_{P'})N(Z; 0, 1)dZ, \text{ or} \\
 & E(U(\sigma_{P''}\tilde{Z} + \mu_{P''})) > \alpha E(U(\sigma_P\tilde{Z} + \mu_P)) + (1 - \alpha)E(U(\sigma_{P'}\tilde{Z} + \mu_{P'})), \text{ or} \\
 & E(U(\sigma_{P''}\tilde{Z} + \mu_{P''})) > \alpha\bar{U} + (1 - \alpha)\bar{U} = \bar{U}
 \end{aligned}$$

See Figure A.6.4 for an illustration. \square

Insert Figure A.6.4

Appendix 6.2: The Shape of the Efficient Frontier; Two Assets; Alternative Hypotheses

Perfect Positive Correlation (Figure 6.2)

$$\rho_{12} = 1$$

$\sigma_P = w_1\sigma_1 + (1 - w_1)\sigma_2$, the weighed average of the SD's of individuals asset returns

$$\begin{aligned}\rho_{1,2} &= 1 \\ \mu_P &= w_1\bar{r}_1 + (1 - w_1)\bar{r}_2 = \bar{r}_1 + (1 - w_1)(\bar{r}_2 - \bar{r}_1) \\ \sigma_P^2 &= w_1^2\sigma_1^2 + (1 - w_1)^2\sigma_2^2 + 2w_1w_2\sigma_1\sigma_2\rho_{1,2} \\ &= w_1^2\sigma_1^2 + (1 - w_1)^2\sigma_2^2 + 2w_1w_2\sigma_1\sigma_2 \\ &= (w_1\sigma_1 + (1 - w_1)\sigma_2)^2 \text{ [perfect square]}\end{aligned}$$

$$\begin{aligned}\sigma_P &= \pm(w_1\sigma_1 + (1 - w_1)\sigma_2) \Rightarrow w_1 = \frac{\sigma_P - \sigma_2}{\sigma_1 - \sigma_2}; 1 - w_1 = \frac{\sigma_1 - \sigma_P}{\sigma_1 - \sigma_2} \\ \mu_P &= \bar{r}_1 + \frac{\sigma_1 - \sigma_P}{\sigma_1 - \sigma_2}(\bar{r}_2 - \bar{r}_1) = \bar{r}_1 + \frac{\bar{r}_2 - \bar{r}_1}{\sigma_2 - \sigma_1}(\sigma_P - \sigma_1)\end{aligned}$$

Imperfectly Correlated Assets (Figure 6.3)

$$-1 < \rho_{12} < 1$$

$$\begin{aligned}\text{Reminder: } \mu_P &= w_1\bar{r}_1 + (1 - w_1)\bar{r}_2 \\ \sigma_P^2 &= w_1^2\sigma_1^2 + (1 - w_1)^2\sigma_2^2 + 2w_1w_2\sigma_1\sigma_2\rho_{1,2}\end{aligned}$$

Thus,

$$\frac{\partial \sigma_P^2}{\partial \rho_{1,2}} = 2w_1w_2\sigma_1\sigma_2 > 0$$

which implies: $\sigma_P < w_1\sigma_1 + (1 - w_1)\sigma_2$; σ_P is smaller than the weighted average of the σ 's, there are gains from diversifying.

Fix μ_P , hence w_1 , and observe: as one decreases $\rho_{1,2}$ (from +1 to -1), σ_P^2 diminishes (and thus also σ_P). Hence the opportunity set for $\rho = \bar{\rho} < 1$ must be to the left of the line AB ($\rho_{1,2} = 1$) except for the extremes.

$$\begin{aligned}w_1 = 0 &\Rightarrow \mu_P = \bar{r}_2 \text{ et } \sigma_P^2 = \sigma_2^2 \\ w_1 = 1 &\Rightarrow \mu_P = \bar{r}_1 \text{ et } \sigma_P^2 = \sigma_1^2\end{aligned}$$

Perfect Negative Correlation (Figure 6.4)

$$\begin{aligned}
\rho_{1,2} &= -1 \\
\sigma_P^2 &= w_1^2 \sigma_1^2 + (1 - w_1)^2 \sigma_2^2 - 2w_1 w_2 \sigma_1 \sigma_2 \text{ with } (w_2 = (1 - w_1)) \\
&= (w_1 \sigma_1 - (1 - w_1) \sigma_2)^2 \text{ [perfect square again]} \\
\sigma_P &= \pm [w_1 \sigma_1 - (1 - w_1) \sigma_2] = \pm [w_1 (\sigma_1 + \sigma_2) - \sigma_2] \\
w_1 &= \frac{\pm \sigma_P + \sigma_2}{\sigma_1 + \sigma_2} \\
\sigma_P &= 0 \Leftrightarrow w_1 = \frac{\sigma_2}{\sigma_1 + \sigma_2} \\
\mu_P &= \frac{\pm \sigma_P + \sigma_2}{\sigma_1 + \sigma_2} \bar{r}_1 + \left(1 - \frac{\pm \sigma_P + \sigma_2}{\sigma_1 + \sigma_2}\right) \bar{r}_2 \\
&= \frac{\pm \sigma_P + \sigma_2}{\sigma_1 + \sigma_2} \bar{r}_1 + \frac{\sigma_1 \pm \sigma_P}{\sigma_1 + \sigma_2} \bar{r}_2 \\
&= \frac{\sigma_2}{\sigma_1 + \sigma_2} \bar{r}_1 + \frac{\sigma_1}{\sigma_1 + \sigma_2} \bar{r}_2 \pm \frac{\bar{r}_1 - \bar{r}_2}{\sigma_1 + \sigma_2} \sigma_P
\end{aligned}$$

One Riskless and One Risky Asset (Figure 6.5)

Asset 1: $\bar{r}_1, \sigma_1 = 0$

Asset 2: \bar{r}_2, σ_2

$\bar{r}_1 < \bar{r}_2$

$$\begin{aligned}
\mu_P &= w_1 \bar{r}_1 + (1 - w_1) \bar{r}_2 \\
\sigma_P^2 &= w_1^2 \sigma_1^2 + (1 - w_1)^2 \sigma_2^2 + 2w_1 (1 - w_1) \text{cov}_{1,2} \\
&= (1 - w_1)^2 \sigma_2^2 \text{ since } \sigma_1^2 = 0 \text{ and } \text{cov}_{1,2} = \rho_{1,2} \sigma_1 \sigma_2 = 0; \text{ thus,} \\
\sigma_P &= (1 - w_1) \sigma_2, \text{ and} \\
w_1 &= 1 - \frac{\sigma_P}{\sigma_2}
\end{aligned}$$

Appendix 6.3: Constructing the Efficient Frontier

In this Appendix we outline how Excel's "SOLVER" program may be used to construct an efficient frontier using historical data on returns. Our method does not require the explicit computation of means, standard deviations, and return correlations for the various securities under consideration; they are implicitly obtained from the data directly.

The Basic Portfolio Problem

Let us, for purposes of illustration, assume that we have assembled a time series of four data points (monthly returns) for each of three stocks, and let us further assume that these four realizations fully describe the relevant return distributions. We also assign equal probability to the states underlying these realizations.

Table A6.1 Hypothetical Return Data

	Prob	Stock 1	Stock 2	Stock 3
State 1	.25	6.23%	5.10%	7.02%
State 2	.25	-.68%	4.31%	.79%
State 3	.25	5.55%	-1.27%	-.21%
State 4	.25	-1.96%	4.52%	10.30%

Table A6.1 presents this hypothetical data.

Following our customary notation, let w_i represent the fraction of wealth invested in asset i , $i = 1, 2, 3$, and let r_{P,θ_j} represent the return for a portfolio of these assets in the case of event θ_j , $j = 1, 2, 3, 4$. The Excel formulation analogous to problem (QP) of the text is found in Table A6.2. where (A1) through (A4) define the portfolio's return in each of the four states; (A5) defines the portfolio's average return; (A6) places a bound on the expected return; by varying μ , it is possible to trace out the efficient frontier; (A7) defines the standard deviation when each state is equally probable; and (A8) is the *budget constraint*.

Table A6.2: The Excel Formulation of the (QP) Problem

$$\begin{aligned}
& \min_{\{w_1, w_2, w_3, w_4\}} SD \\
& \text{(minimize portfolio standard deviation)} \\
& \text{Subject to:} \\
& \text{(A1) } r_{P,\theta_1} = 6.23w_1 + 5.10w_2 + 7.02w_3 \\
& \text{(A2) } r_{P,\theta_2} = -.68w_1 + 4.31w_2 + .79w_3 \\
& \text{(A3) } r_{P,\theta_3} = 5.55w_1 - 1.27w_2 - .21w_3 \\
& \text{(A4) } r_{P,\theta_4} = -1.96w_1 + 4.52w_2 + 10.30w_3 \\
& \text{(A5) } \bar{r}_P = .25r_1^P + .25r_2^P + .25r_3^P + .25r_4^P \\
& \text{(A6) } \bar{r}_P \geq \mu = 3 \\
& \text{(A7) } SD = SQRT(SUMPRODUCT(r_{P,\theta_1}, r_{P,\theta_2}, r_{P,\theta_3}, r_{P,\theta_4})) \\
& \text{(A8) } w_1 + w_2 + w_3 = 1
\end{aligned}$$

The Excel-based solution to this problem is

$$w_1 = .353$$

$$w_2 = .535$$

$$w_3 = .111,$$

when μ is fixed at $\mu = 3.0\%$. The corresponding portfolio mean and standard deviation are $\bar{r}_P = 3.00$, and $\sigma_P = 1.67$. Screen 1 describes the Excel setup for this case.

Insert Figure A6.5 about here

Notice that this approach does not require the computation of individual security expected returns, variances, or correlations, but it is fundamentally no

different than problem (QP) in the text which does require them. Notice also that by recomputing “min SD” for a number of different values of μ , the efficient frontier can be well approximated.

Generalizations

The approach described above is very flexible and accommodates a number of variations, all of which amount to specifying further constraints.

Non-Negativity Constraints

These amount to restrictions on short selling. It is sufficient to specify the additional constraints

$$\begin{aligned}w_1 &\geq 0 \\w_2 &\geq 0 \\w_3 &\geq 0.\end{aligned}$$

The functioning of SOLVER is unaffected by these added restrictions (although more constraints must be added), and for the example above the solution remains unchanged. (This is intuitive since the solutions were all positive.) See Screen 2.

Insert Figure A6.6 about here

Composition Constraints

Let us enrich the scenario. Assume the market prices of stocks 1, 2, and 3 are, respectively, \$25, \$32, and \$17, and that the current composition of the portfolio consists of 10,000 shares of stock 1, 10,000 shares of stock 2, and 30,000 shares of stock 3, with an aggregate market value of \$1,080,000. You wish to obtain the lowest SD for a given expected return subject to the constraints that you retain 10,000 shares of stock 1 and 10,000 shares of stock 3. Equivalently, you wish to constrain portfolio proportions as follows:

$$\begin{aligned}w_1 &\geq \frac{10,000 \times \$25}{\$1,080,000} = .23 \\w_3 &\geq \frac{10,000 \times \$17}{\$1,080,000} = .157;\end{aligned}$$

while w_2 is free to vary. Again SOLVER easily accommodates this. We find $w_1 = .23$, $w_2 = .453$, and $w_3 = .157$, yielding $\bar{r}_P = 3.03\%$ and $\sigma_P = 1.70\%$. Both constraints are binding. See Screen 3.

Insert Figure A6.7 about here

Adjusting the Data (Modifying the Means)

On the basis of the information in Table A6.1,

$$\begin{aligned}\bar{r}_1 &= 2.3\% \\ \bar{r}_2 &= 3.165\% \\ \bar{r}_3 &= 4.47\%.\end{aligned}$$

Suppose, either on the basis of fundamental analysis or an SML-style calculation, other information becomes available suggesting that, over the next portfolio holding period, the returns on stocks 1 and 2 would be 1% higher than their historical mean and the return on stock 3 would be 1% lower. This supplementary information can be incorporated into min SD by modifying Table A6.1. In particular, each return entry for stocks 1 and 2 must be increased by 1% while each entry of stock 3 must be decreased by 1%. Such changes do not in any way alter the standard deviations or correlations implicit in the data. The new input table for SOLVER is found in Table A6.3.

Table A6.3 Modified Return Data

	Prob	Stock 1	Stock 2	Stock 3
Event 1	.25	7.23%	6.10%	6.02%
Event 2	.25	.32%	5.31%	-.21%
Event 3	.25	6.55%	-.27%	-1.21%
Event 4	.25	-.96%	5.52%	9.30%

Solving the same problem, min SD without additional constraints yields $w_1 = .381$, $w_2 = .633$, and $w_3 = -0.013$, yielding $\bar{r}_P = 3.84$ and $\sigma_P = 1.61$. See Screen 4.

Insert Figure A6.8 about here

Constraints on the Number of Securities in the Portfolio

Transactions costs may be substantial. In order to economize on these costs, suppose an investor wished to solve min SD subject to the constraint that his portfolio would contain at most two of the three securities. To accommodate this change, it is necessary to introduce three new binary variables that we will denote x_1, x_2, x_3 , corresponding to stocks 1, 2, and 3, respectively. For all $x_i, i = 1, 2, 3, x_i \in \{0, 1\}$. The desired result is obtained by adding the following constraints to the problem min SD:

$$\begin{aligned}
 w_1 &\leq x_1 \\
 w_2 &\leq x_2 \\
 w_3 &\leq x_3 \\
 x_1 + x_2 + x_3 &\leq 2, \\
 x_1, x_2, x_3 &\text{ are binary}
 \end{aligned}$$

In the previous example the solution is to include only securities one and two with proportions $w_1 = .188$, and $w_2 = .812$. See Screen 5.

Insert Figure A6.9 about here

Part III

Equilibrium Pricing

Chapter 7: The Capital Asset Pricing Model: Another View About Risk

7.1 Introduction

The CAPM is an equilibrium theory built on the premises of Modern Portfolio Theory. It is, however, an equilibrium theory with a somewhat peculiar structure. This is true for a number of reasons:

1. First, the CAPM is a theory of financial equilibrium only. Investors take the various statistical quantities — means, variances, covariances — that characterize a security's return process as given. There is no attempt within the theory to link the return processes with events in the *real* side of the economy. In future model contexts we shall generalize this feature.

2. Second, as a theory of financial equilibrium it makes the assumption that the supply of existing assets is equal to the demand for existing assets and, as such, that the currently observed asset prices are equilibrium ones. There is no attempt, however, to compute asset supply and demand functions explicitly. Only the equilibrium price vector is characterized. Let us elaborate on this point.

Under the CAPM, portfolio theory informs us about the demand side. If individual i invests a fraction w_{ij} of his initial wealth Y_{0i} in asset j , the value of his asset j holding is $w_{ij}Y_{0i}$. Absent any information that he wishes to alter these holdings, we may interpret the quantity $w_{ij}Y_{0i}$ as his demand for asset j at the prevailing price vector. If there are I individuals in the economy, the total value of all holdings of asset j is $\sum_i w_{ij}Y_{0i}$; by the same remark we may interpret this quantity as aggregate demand. At equilibrium one must have $\sum_i w_{ij}Y_{0i} = p_j Q_j$ where p_j is the prevailing equilibrium price per share of asset j , Q_j is the total number of shares outstanding and, consequently, $p_j Q_j$ is the market capitalization of asset j . The CAPM derives the implications for prices by assuming that the actual economy-wide asset holdings are investors' aggregate optimal asset holdings.

3. Third, the CAPM expresses equilibrium in terms of relationships between the return distributions of individual assets and the return characteristics of the portfolio of all assets. We may view the CAPM as informing us, via modern Portfolio Theory, as to what asset return interrelationships must be in order for equilibrium asset prices to coincide with the observed asset prices.

In what follows we first present an overview of the traditional approach to the CAPM. This is followed by a more general presentation that permits at once a more complete and more general characterization.

7.2 The Traditional Approach to the CAPM

To get useful results in this complex world of many assets we have to make simplifying assumptions. The CAPM approach essentially hypothesizes (1) that all

agents have the *same beliefs* about future returns (i.e., homogenous expectations), and, in its simplest form, (2) that there is a risk-free asset, paying a safe return r_f . These assumptions guarantee (Chapter 6) that the mean-variance efficient frontier is the same for every investor, and furthermore, by the separation theorem, that all investors' optimal portfolios have an identical structure: a fraction of initial wealth is invested in the risk-free asset, the rest in the (identical) tangency portfolio (two-fund separation). It is then possible to derive a few key characteristics of equilibrium asset and portfolio returns without detailing the underlying equilibrium structure, that is, the demand for and supply of assets, or discussing their prices.

Because all investors acquire shares in the same risky tangency portfolio T , and make no other risky investments, all existing risky assets must belong to T by the definition of an equilibrium. Indeed, if some asset k were not found in T , there would be no demand for it; yet, it is assumed to exist in positive supply. Supply would then exceed demand, which is inconsistent with assumed financial market equilibrium. The same reasoning implies that the share of any asset j in portfolio T must correspond to the ratio of the market value of that asset $p_j Q_j$ to the market value of all assets $\sum_{j=1}^J p_j Q_j$. This, in turn, guarantees that tangency portfolio T must be nothing other than the market portfolio M , the portfolio of all existing assets where each asset appears in a proportion equal to the ratio of its market value to the total market capitalization.

Insert Figure 7.1

This simple reasoning leads to a number of useful conclusions:

- a. The market portfolio is efficient since it is on the efficient frontier.
- b. All individual optimal portfolios are located on the half-line originating at point $(0, r_f)$ and going through (\bar{r}_M, σ_M) , which is also the locus of all efficient portfolios (see Figure 7.1). This locus is usually called the **Capital Market Line** or CML.
- c. The slope of the CML is $\frac{\bar{r}_M - r_f}{\sigma_M}$. It tells us that an investor considering a marginally riskier efficient portfolio would obtain, in exchange, an increase in expected return of $\frac{\bar{r}_M - r_f}{\sigma_M}$. This is the price of, or reward for, risk taking—the price of risk as applicable to efficient portfolios. In other words, for efficient portfolios, we have the simple linear relationship in Equation (7.1).

$$\bar{r}_p = r_f + \frac{\bar{r}_M - r_f}{\sigma_M} \sigma_p \quad (7.1)$$

The CML applies only to efficient portfolios. What can be said of an arbitrary asset j not belonging to the efficient frontier? To discuss this essential part of

the CAPM we first rely on Equation (7.2), formally derived in Appendix 7.1, and limit our discussion to its intuitive implications:

$$\bar{r}_j = r_f + (\bar{r}_M - r_f) \frac{\sigma_{jM}}{\sigma_M^2} \quad (7.2)$$

Let us define $\beta_j = \frac{\sigma_{jM}}{\sigma_M^2}$, that is the ratio of the covariance between the returns on asset j and the returns on the market portfolio over the variance of the market returns. We can thus rewrite Equation (7.2) as Equation (7.3).

$$\bar{r}_j = r_f + \left(\frac{\bar{r}_M - r_f}{\sigma_M} \right) \beta_j \sigma_M = r_f + \left(\frac{\bar{r}_M - r_f}{\sigma_M} \right) \rho_{jM} \sigma_j \quad (7.3)$$

Comparing Equations (7.1) and (7.3), we obtain one of the major lessons of the CAPM: Only a portion of the total risk of an asset j , σ_j , is remunerated by the market. Indeed, the risk premium on a given asset is the market price of risk, $\frac{\bar{r}_M - r_f}{\sigma_M}$, multiplied by the relevant measure of the quantity of risk for that asset. In the case of an inefficient asset or portfolio j , this measure of risk differs from σ_j . The portion of total risk that is priced is measured by $\beta_j \sigma_M$ or $\rho_{jM} \sigma_j$ ($\leq \sigma_j$). This is the **systematic risk** of asset j (also referred to as market risk or undiversifiable risk). The intuition for this fundamental result is as follows.

Every investor holds the market portfolio ($T = M$). The relevant risk for the investor is thus the variance of the market portfolio. Consequently, what is important to him is the contribution of asset j to the risk of the market portfolio; that is, the extent to which the inclusion of asset j into the overall portfolio increases the latter's variance. This marginal contribution of asset j to the overall portfolio risk is appropriately measured by $\beta_j \sigma_M$ ($= \beta_j \sigma_M$). Equation (7.3) says that investors must be compensated to persuade them to hold an asset with high covariance with the market, and that this compensation takes the form of a higher expected return.

The comparison of Equations (7.1) and (7.3) also leads us to conclude that an efficient portfolio is one for which all diversifiable risks have been eliminated. For an efficient portfolio, total risk and systematic risk are thus one and the same. This result is made clear from writing, without loss of generality, the return on asset j as a linear function of the market return with a random error term that is independent of the market return,¹

$$\tilde{r}_j = \alpha + \beta_j \tilde{r}_M + \varepsilon_j \quad (7.4)$$

Looking at the implication of this general regression equation for variances,

$$\sigma_j^2 = \beta_j^2 \sigma_M^2 + \sigma_{\varepsilon_j}^2, \quad (7.5)$$

we obtain the justification for the “beta” label. The standard regression estimator of the market return coefficient in Equation (7.4) will indeed be of the

¹The “market model” is based on this same regression equation. The market model is reviewed in Chapter 13.

form

$$\hat{\beta}_j = \frac{\hat{\sigma}_{jM}}{\hat{\sigma}_M^2}.$$

Equation (7.3) can equivalently be rewritten as

$$\bar{r}_j - r_f = (\bar{r}_M - r_f) \beta_j \quad (7.6)$$

which says that the expected excess return or the risk premium on an asset j is proportional to its β_j . Equation (7.6) defines the **Security Market Line** or SML. It is depicted in Figure 7.2.

The SML has two key features. The beta of asset j , β_j , is the sole specific determinant of the excess return on asset j . Adopting a terminology that we shall justify later, we can say that the beta is the unique explanatory factor model. Furthermore, the relation between excess returns on different assets and their betas is linear.

Insert Figure 7.2

7.3 Valuing risky cash flows with the CAPM

We are now in position to make use of the CAPM not only to price assets but also to value non-traded risky cash flows such as those arising from an investment project. The traditional approach to this problem proposes to value an investment project at its present value price, i.e., at the appropriately discounted sum of the expected future cash flows. The logic is straightforward: To value a project equal to the present value of its expected future cash flows discounted at a particular rate is to price the project in a manner such that, at its present value price, it is expected to earn that discount rate. The appropriate rate, in turn, must be the analyst's estimate of the rate of return on *other* financial assets that represent title to cash flows *similar* in risk and timing to that of the project in question. This strategy has the consequence of pricing the project to pay the prevailing competitive rate for its risk class.

Enter the CAPM which makes a definite statement regarding on the appropriate discount factor to be used or, equivalently, on the risk premium that should be applied to discount expected future cash flows. Strictly speaking, the CAPM is a one-period model; it is thus formally appropriate to use it only for one-period cash flows or projects. In practice its use is more general and a multi-period cash-flow is typically viewed as the sum of one-period cash flows, each of which can be evaluated with the approach we now describe.

Consider some project j with cash flow pattern

t	$t + 1$
$-p_{j,t}$	$\bar{CF}_{j,t+1}$

The link with the CAPM is immediate once we define the rate of return on project j . For a financial asset we would naturally write $\tilde{r}_{j,t+1} = \frac{\tilde{p}_{j,t+1} + \tilde{d}_{j,t} - p_{j,t}}{p_{j,t}}$, where $\tilde{d}_{j,t}$ is the dividend or any flow payment associated with the asset between date t and $t + 1$. Similarly, if the initial value of the project with cash flow $\tilde{CF}_{j,t+1}$ is $p_{j,t}$, the return on the project is $\tilde{r}_{j,t+1} = \frac{\tilde{CF}_{j,t+1} - p_{j,t}}{p_{j,t}}$.

One thus has

$$1 + E(\tilde{r}_j) = E\left(\frac{\tilde{CF}_{j,t+1}}{p_{j,t}}\right) = \frac{E(\tilde{CF}_{j,t+1})}{p_{j,t}}, \text{ and by the CAPM,}$$

$$\begin{aligned} E(\tilde{r}_j) &= r_f + \beta_j (E(\tilde{r}_M) - r_f), \text{ or} \\ 1 + E(\tilde{r}_j) &= 1 + (r_f + \beta_j (E(\tilde{r}_M) - r_f)), \text{ or} \\ \frac{E(\tilde{CF}_{j,t+1})}{p_{j,t}} &= 1 + r_f + \beta_j (E(\tilde{r}_M) - r_f). \text{ Thus,} \\ p_{j,t} &= \frac{E(\tilde{CF}_{j,t+1})}{1 + r_f + \beta_j (E(\tilde{r}_M) - r_f)}. \end{aligned}$$

According to the CAPM, the project is thus priced at the present value of its expected cash flows discounted at the risk-adjusted rate appropriate to its risk class (β_j).

As discussed in Chapter 1, there is another potential approach to the pricing problem. It consists in altering the *numerator* of the pricing equations (the sum of expected cash flows) so that it is permissible to discount at the risk-free rate. This approach is based on the concept of certainty equivalent, which we discussed in Chapter 3. The idea is simple: If we replace each element of the future cash flow by its *CE*, it is clearly permissible to discount at the risk-free rate. Since we are interested in equilibrium valuations, however, we need a market certainty equivalent rather than an individual investor one. It turns out that this approach raises exactly the same set of issues as the more common one just considered: an equilibrium asset pricing model is required to tell us what market risk premium it is appropriate to deduct from the expected cash flow to obtain its *CE*. Again the CAPM helps solve this problem.²

In the case of a one-period cash flow, transforming period-by-period cash flows into their market certainty equivalents can be accomplished in a straightforward fashion by applying the CAPM equation to the rate of return expected on the project. With $\tilde{r}_j = \frac{\tilde{CF}_j^{t+1}}{p_{j,t}} - 1$, the CAPM implies

$$E\left(\frac{\tilde{CF}_{j,t+1}}{p_{j,t}} - 1\right) = r_f + \beta_j (E\tilde{r}_M - r_f) = r_f + \frac{\text{cov}\left(\frac{\tilde{CF}_{j,t+1}}{p_{j,t}} - 1, \tilde{r}_M\right)}{\sigma_M^2} (E\tilde{r}_M - r_f),$$

or

$$E\left(\frac{\tilde{CF}_{j,t+1}}{p_{j,t}} - 1\right) = r_f + \frac{1}{p_{j,t}} \text{cov}(\tilde{CF}_{j,t+1}, \tilde{r}_M) \left[\frac{E(\tilde{r}_M) - r_f}{\sigma_M^2}\right].$$

²Or, similarly, the APT.

Solving for $p_{j,t}$ yields

$$p_{j,t} = \frac{E(\tilde{C}F_{j,t+1}) - \text{cov}(\tilde{C}F_{j,t+1}, \tilde{r}_M) \left[\frac{E\tilde{r}_M - r_f}{\sigma_M^2} \right]}{1 + r_f},$$

which one may also write

$$p_{j,t} = \frac{E(\tilde{C}F_{j,t+1}) - p_{j,t}\beta_j[E\tilde{r}_M - r_f]}{1 + r_f}.$$

Thus by appropriately transforming the expected cash flows, that is, by subtracting what we have called an insurance premium (in Chapter 3), one can discount at the risk-free rate. The equilibrium certainty equivalent can thus be defined using the CAPM relationship. Note the information requirements in the procedure: if what we are valuing is indeed a one-off, non-traded cash flow, the estimation of the β_j , or of $\text{cov}(\tilde{C}F_{j,t+1}, \tilde{r}_M)$, is far-from-straightforward; in particular, it cannot be based on historical data since they are none for the project at hand. It is here that the standard prescription calls for identifying a traded asset that can be viewed as similar in the sense of belonging to the same risk class. The estimated β for that traded asset is then to be used as an approximation in the above valuation formulas.

In the sections that follow, we first generalize the analysis of the efficient frontier presented in Chapter 6 to the $N \geq 2$ asset case. Such a generalization will require the use of elementary matrix algebra and is one of those rare situations in economic science where a more general approach yields a greater specificity of results. We will, for instance, be able to detail a version of the CAPM without risk-free asset. This is then followed by the derivation of the standard CAPM where a risk-free asset is present.

As noted in the introduction, the CAPM is essentially an interpretation that we are able to apply to the efficient frontier. Not surprisingly, therefore, we begin this task with a return to characterizing that frontier.

7.4 The Mathematics of the Portfolio Frontier: Many Risky Assets and No Risk-Free Asset

Notation. Assume $N \geq 2$ risky assets; assume further that no asset has a return that can be expressed as a linear combination of the returns to a subset of the other assets, (the returns are linearly independent). Let V denote the variance-covariance matrix, in other words, $V_{ij} = \text{cov}(r_i, r_j)$; by construction V is symmetric. Linear independence in the above sense implies that V^{-1} exists. Let w represent a column vector of portfolio weights for the N assets. The expression $w^T V w$ then represents the portfolio's return variance: $w^T V w$ is always positive (i.e., V is positive definite).

Let us illustrate this latter assertion in the two-asset case

$$w^T V w = \begin{pmatrix} w_1 & w_2 \end{pmatrix} \begin{pmatrix} \sigma_1^2 & \sigma_{12} \\ \sigma_{21} & \sigma_2^2 \end{pmatrix} \begin{pmatrix} w_1 \\ w_2 \end{pmatrix} = \begin{pmatrix} w_1 \sigma_1^2 + w_2 \sigma_{21} & w_1 \sigma_{12} + w_2 \sigma_2^2 \end{pmatrix} \begin{pmatrix} w_1 \\ w_2 \end{pmatrix}$$

$$\begin{aligned}
&= w_1^2 \sigma_1^2 + w_1 w_2 \sigma_{21} + w_1 w_2 \sigma_{12} + w_2^2 \sigma_2^2 \\
&= w_1^2 \sigma_1^2 + w_2^2 \sigma_2^2 + 2w_1 w_2 \sigma_{12} \geq 0
\end{aligned}$$

since $\sigma_{12} = \rho_{12} \sigma_1 \sigma_2 \geq -\sigma_1 \sigma_2$.

Definition 7.1 formalizes the notion of a portfolio lying on the efficient frontier. Note that every portfolio is ultimately defined by the weights that determine its composition.

Definition 7.1:

A frontier portfolio is one that displays minimum variance among all feasible portfolios with the same $E(\tilde{r}_p)$.

A portfolio p , characterized by w_p , is a frontier portfolio, if and only if w_p solves.³

$$\begin{aligned}
&\min_w \frac{1}{2} w^T V w \\
(\lambda) \quad &\text{s.t. } w^T e = E \quad \left(\sum_{i=1}^N w_i E(\tilde{r}_i) = E(\tilde{r}_p) = E \right) \\
(\gamma) \quad &w^T \mathbf{1} = 1 \quad \left(\sum_{i=1}^N w_i = 1 \right)
\end{aligned}$$

where the superscript T stands for *transposed*, i.e., transforms a column vector into a line vector and reciprocally, e denotes the column vector of expected returns to the N assets, $\mathbf{1}$ represents the column vector of ones, and λ, γ are Lagrange multipliers. Short sales are permitted (no non-negativity constraints are present). The solution to this problem can be characterized as the solution to $\min_{\{w, \lambda, \gamma\}}$ where L is the Lagrangian:

$$L = \frac{1}{2} w^T V w + \lambda (E - w^T e) + \gamma (1 - w^T \mathbf{1}) \quad (7.7)$$

Under these assumptions, w_p , λ and γ must satisfy Equations (7.8) through (7.10), which are the necessary and sufficient first order conditions:

$$\frac{\partial L}{\partial w} = Vw - \lambda e - \gamma \mathbf{1} = 0 \quad (7.8)$$

$$\frac{\partial L}{\partial \lambda} = E - w^T e = 0 \quad (7.9)$$

$$\frac{\partial L}{\partial \gamma} = 1 - w^T \mathbf{1} = 0 \quad (7.10)$$

In the lines that follow, we manipulate these equations to provide an intuitive characterization of the optimal portfolio proportions (7.16). From (7.8), $Vw_p = \lambda e + \gamma \mathbf{1}$, or

³The problem below is, in vector notation, problem (QP) of Chapter 5.

$$w_p = \lambda V^{-1}e + \gamma V^{-1}\mathbf{1}, \text{ and} \quad (7.11)$$

$$e^T w_p = \lambda (e^T V^{-1}e) + \gamma (e^T V^{-1}\mathbf{1}). \quad (7.12)$$

Since $e^T w_p = w_p^T e$, we also have, from Equation (7.9), that

$$E(\tilde{r}_p) = \lambda (e^T V^{-1}e) + \gamma (e^T V^{-1}\mathbf{1}). \quad (7.13)$$

From Equation (7.12), we have:

$$\begin{aligned} \mathbf{1}^T w_p &= w_p^T \mathbf{1} = \lambda (\mathbf{1}^T V^{-1}e) + \gamma (\mathbf{1}^T V^{-1}\mathbf{1}) \\ &= 1 \text{ [by Equation(7.10)]} \\ 1 &= \lambda (\mathbf{1}^T V^{-1}e) + \gamma (\mathbf{1}^T V^{-1}\mathbf{1}) \end{aligned} \quad (7.14)$$

Notice that Equations (7.13) and (7.14) are two *scalar* equations in the unknowns λ and γ (since such terms as $e^T V^{-1}e$ are pure numbers!). Solving this system of two equations in two unknowns, we obtain:

$$\lambda = \frac{CE - A}{D} \text{ and } \gamma = \frac{B - AE}{D} \quad (7.15)$$

where

$$\begin{aligned} A &= \mathbf{1}^T V^{-1}e = e^T V^{-1}\mathbf{1} \\ B &= e^T V^{-1}e > 0 \\ C &= \mathbf{1}^T V^{-1}\mathbf{1} \\ D &= BC - A^2 \end{aligned}$$

Here we have used the fact that the inverse of a positive definite matrix is itself positive definite. It can be shown that D is also strictly positive. Substituting Equations (7.15) into Equation (7.11) we obtain:

$$\begin{aligned} w_p &= \underbrace{\frac{CE - A}{D}}_{\substack{\lambda \\ \text{scalar}}} \underbrace{V^{-1}e}_{\text{vector}} + \underbrace{\frac{B - AE}{D}}_{\substack{\gamma \\ \text{scalar}}} \underbrace{V^{-1}\mathbf{1}}_{\text{vector}} \\ &= \frac{1}{D} [B (V^{-1}\mathbf{1}) - A (V^{-1}e)] + \frac{1}{D} [C (V^{-1}e) - A (V^{-1}\mathbf{1})] E \\ w_p &= \underbrace{g}_{\text{vector}} + \underbrace{h}_{\text{vector}} \underbrace{E}_{\text{scalar}} \end{aligned} \quad (7.16)$$

Since the FOCs [Equations (7.8) through (7.10)] are a necessary and sufficient characterization for w_p to represent a frontier portfolio with expected

return equal to E , any frontier portfolio can be represented by Equation (7.16). This is a very nice expression; pick the desired expected return E and it straightforwardly gives the weights of the corresponding frontier portfolio with E as its expected return. The portfolio's variance follows as $\sigma_p^2 = w_p^T V w_p$, which is also straightforward. Efficient portfolios are those for which E exceeds the expected return on the minimum risk, risky portfolio. Our characterization thus applies to efficient portfolios as well: Pick an efficient E and Equation (7.16) gives its exact composition. See Appendix 7.2 for an example.

Can we further identify the vectors g and h in Equation (7.16); in particular, do they somehow correspond to the weights of easily recognizable portfolios? The answer is positive. Since, if $E = 0$, $g = w_p$, g then represents the weights that define the frontier portfolio with $E(\tilde{r}_p) = 0$. Similarly, $g + h$ corresponds to the weights of the frontier portfolio with $E(\tilde{r}_p) = 1$, since $w_p = g + hE(\tilde{r}_p) = g + h1 = g + h$.

The simplicity of the relationship in Equation (7.16) allows us to make two claims.

Proposition 7.1:

The entire set of frontier portfolios can be generated by (are affine combinations of) g and $g + h$.

Proof:

To see this, let q be an arbitrary frontier portfolio with $E(\tilde{r}_q)$ as its expected return. Consider portfolio weights (proportions) $\pi_g = 1 - E(\tilde{r}_q)$ and $\pi_{g+h} = E(\tilde{r}_q)$; then, as asserted,

$$[1 - E(\tilde{r}_q)]g + E(\tilde{r}_q)(g + h) = g + hE(\tilde{r}_q) = w_q. \square$$

The prior remark is generalized in Proposition 7.2.

Proposition 7.2:

The portfolio frontier can be described as affine combinations of any two frontier portfolios, not just the frontier portfolios g and $g + h$.

Proof:

To confirm this assertion, let p_1 and p_2 be any two distinct frontier portfolios; since the frontier portfolios are different, $E(\tilde{r}_{p_1}) \neq E(\tilde{r}_{p_2})$. Let q be an arbitrary frontier portfolio, with expected return equal to $E(\tilde{r}_q)$. Since $E(\tilde{r}_{p_1}) \neq E(\tilde{r}_{p_2})$, there must exist a unique number α such that

$$E(\tilde{r}_q) = \alpha E(\tilde{r}_{p_1}) + (1 - \alpha) E(\tilde{r}_{p_2}) \quad (7.17)$$

Now consider a portfolio of p_1 and p_2 with weights $\alpha, 1 - \alpha$, respectively, as determined by Equation (7.17). We must show that $w_q = \alpha w_{p_1} + (1 - \alpha) w_{p_2}$.

$$\begin{aligned} \alpha w_{p_1} + (1 - \alpha) w_{p_2} &= \alpha [g + hE(\tilde{r}_{p_1})] + (1 - \alpha) [g + hE(\tilde{r}_{p_2})] \\ &= g + h [\alpha E(\tilde{r}_{p_1}) + (1 - \alpha) E(\tilde{r}_{p_2})] \\ &= g + hE(\tilde{r}_q) \\ &= w_q, \text{ since } q \text{ is a frontier portfolio.} \square \end{aligned}$$

What does the set of frontier portfolios, which we have calculated so conveniently, look like? Can we identify, in particular, the minimum variance portfolio? Locating that portfolio is surely key to a description of the set of all frontier portfolios. Fortunately, given our results thus far, the task is straightforward. For any portfolio on the frontier,

$$\sigma^2(\tilde{r}_p) = [g + hE(\tilde{r}_p)]^T V [g + hE(\tilde{r}_p)], \text{ with } g \text{ and } h \text{ as defined earlier.}$$

Multiplying all this out (very messy), yields:

$$\sigma^2(\tilde{r}_p) = \frac{C}{D} \left(E(\tilde{r}_p) - \frac{A}{C} \right)^2 + \frac{1}{C}, \quad (7.18)$$

where A , C , and D are the constants defined earlier. We can immediately identify the following: since $C > 0, D > 0$,

- (i) the expected return of the minimum variance portfolio is A/C ;
- (ii) the variance of the minimum variance portfolio is given by $\frac{1}{C}$;
- (iii) Equation (7.18) is the equation of a parabola with vertex $(\frac{1}{C}, \frac{A}{C})$ in the expected return/variance space and of a hyperbola in the expected return/standard deviation space. See Figures 7.3 and 7.4.

Insert Figure 7.3

Insert Figure 7.4

The extended shape of this set of frontier portfolios is due to the allowance for short sales as underlined in Figure 7.5.

Insert Figure 7.5

What has been accomplished thus far? First and foremost, we have a much richer knowledge of the set of frontier portfolios: Given a level of desired expected return, we can easily identify the relative proportions of the constituent assets that must be combined to create a portfolio with that expected return. This was illustrated in Equation (7.16), and it is key. We then used it to identify the minimum risk portfolio and to describe the graph of all frontier portfolios.

All of these results apply to portfolios of any arbitrary collection of assets. So far, nothing has been said about financial market equilibrium. As a next step toward that goal, however, we need to identify the set of frontier portfolios that is efficient. Given Equation (7.16) this is a straightforward task.

7.5 Characterizing Efficient Portfolios (No Risk-Free Assets)

Our first order of business is a definition.

Definition 7.2: Efficient portfolios are those frontier portfolios for which the expected return exceeds A/C , the expected return of the minimum variance portfolio.

Since Equation (7.16) applies to all frontier portfolios, it applies to efficient ones as well. Fortunately, we also know the expected return on the minimum variance portfolio. As a first step, let us prove the converse of Proposition 7.2.

Proposition 7.3:

Any convex combination of frontier portfolios is also a frontier portfolio.

Proof:

Let $(\bar{w}_1 \dots \bar{w}_N)$, define N frontier portfolios (\bar{w}_i represents the vector defining the composition of the i th portfolio) and α_i , $i = 1, \dots, N$ be real numbers such that $\sum_{i=1}^N \alpha_i = 1$. Lastly, let $E(\tilde{r}_i)$ denote the expected return of the portfolio with weights \bar{w}_i .

We want to show that $\sum_{i=1}^N \alpha_i \bar{w}_i$ is a frontier portfolio with $E(\bar{r}) = \sum_{i=1}^N \alpha_i E(\tilde{r}_i)$.

The weights corresponding to a linear combination of the above N portfolios are:

$$\begin{aligned} \sum_{i=1}^N \alpha_i \bar{w}_i &= \sum_{i=1}^N \alpha_i (g + hE(\tilde{r}_i)) \\ &= \sum_{i=1}^N \alpha_i g + h \sum_{i=1}^N \alpha_i E(\tilde{r}_i) \\ &= g + h \left[\sum_{i=1}^N \alpha_i E(\tilde{r}_i) \right] \end{aligned}$$

Thus $\sum_{i=1}^N \alpha_i \bar{w}_i$ is a frontier portfolio with $E(\bar{r}) = \sum_{i=1}^N \alpha_i E(\tilde{r}_i)$. \square

A corollary to the previous result is:

Proposition 7.4:

The set of efficient portfolios is a convex set. ⁴

Proof:

Suppose each of the N portfolios under consideration was efficient; then $E(\tilde{r}_i) \geq \frac{A}{C}$, for every portfolio i . However, $\sum_{i=1}^N \alpha_i E(\tilde{r}_i) \geq \sum_{i=1}^N \alpha_i \frac{A}{C} = \frac{A}{C}$; thus, the convex combination is efficient as well. So the set of efficient portfolios, *as characterized by their portfolio weights*, is a convex set. \square

It follows from Proposition 7.4 that if every investor holds an efficient portfolio, the market portfolio, being a weighted average of all individual portfolios, is also efficient. This is a key result.

⁴This does not mean, however, that the frontier of this set is convex-shaped in the risk-return space.

The next section further refines our understanding of the set of frontier portfolios and, more especially, the subset of them that is efficient. Observe, however, that as yet we have said nothing about equilibrium.

7.6 Background for Deriving the Zero-Beta CAPM: Notion of a Zero Covariance Portfolio

Proposition 7.5:

For any frontier portfolio p , except the minimum variance portfolio, there exists a unique frontier portfolio with which p has zero covariance.

We will call this portfolio the *zero covariance portfolio relative to p* , and denote its vector of portfolio weights by $ZC(p)$.

Proof:

To prove this claim it will be sufficient to exhibit the (unique) portfolio that has this property. As we shall demonstrate shortly [see Equation (7.24) and the discussion following it], the covariance of any two frontier portfolios p and q is given by the following general formula:

$$\text{cov}(\tilde{r}_p, \tilde{r}_q) = \frac{C}{D} \left[E(\tilde{r}_p) - \frac{A}{C} \right] \left[E(\tilde{r}_q) - \frac{A}{C} \right] + \frac{1}{C} \quad (7.19)$$

where A, C , and D are uniquely defined by e , the vector of expected returns and V , the matrix of variances and covariances for portfolio p . These are, in fact, the same quantities A, C , and D defined earlier. If it exists, $ZC(p)$ must therefore satisfy,

$$\text{cov}(\tilde{r}_p, \tilde{r}_{ZC(p)}) = \frac{C}{D} \left[E(\tilde{r}_p) - \frac{A}{C} \right] \left[E(\tilde{r}_{ZC(p)}) - \frac{A}{C} \right] + \frac{1}{C} = 0 \quad (7.20)$$

Since A, C , and D are all numbers, we can solve for $E(\tilde{r}_{ZC(p)})$

$$E(\tilde{r}_{ZC(p)}) = \frac{A}{C} - \frac{\frac{D}{C^2}}{E(\tilde{r}_p) - \frac{A}{C}} \quad (7.21)$$

Given $E(\tilde{r}_{ZC(p)})$, we can use Equation (7.16) to uniquely define the portfolio weights corresponding to it. \square

From Equation (7.21), since $A > 0, C > 0, D > 0$, if $E(\tilde{r}_p) > \frac{A}{C}$ (i.e., is efficient), then $E(\tilde{r}_{ZC(p)}) < \frac{A}{C}$ (i.e., is inefficient), and vice versa. The portfolio $ZC(p)$ will turn out to be crucial to what follows. It is possible to give a more complete geometric identification to the zero covariance portfolio if we express the frontier portfolios in the context of the $E(\tilde{r}) - \sigma^2(\tilde{r})$ space (Figure 7.6).

Insert Figure 7.6

The equation of the line through the chosen portfolio p and the minimum variance portfolio can be shown to be the following [it has the form ($y = b + mx$)]:

$$E(\tilde{r}) = \frac{A}{C} - \frac{\frac{D}{C^2}}{E(\tilde{r}_p) - \frac{A}{C}} + \frac{E(\tilde{r}_p) - \frac{A}{C}}{\sigma^2(\tilde{r}_p) - \frac{1}{C}} \sigma^2(\tilde{r}).$$

If $\sigma^2(\tilde{r}) = 0$, then

$$E(\tilde{r}) = \frac{A}{C} - \frac{\frac{D}{C^2}}{E(\tilde{r}_p) - \frac{A}{C}} = E(\tilde{r}_{ZC(p)})$$

[by Equation (7.21)].

That is, the intercept of the line joining p and the minimum variance portfolio is the expected return on the zero-covariance portfolio. This identifies the zero-covariance portfolio to p geometrically. We already know how to determine its precise composition.

Our next step is to describe the expected return on any portfolio in terms of frontier portfolios. After some manipulations this will yield Equation (7.27). The specialization of this relationship will give the zero-beta CAPM, which is a version of the CAPM when there is no risk-free asset. Recall that thus far we have not included a risk-free asset in our collection of assets from which we construct portfolios. Let q be any portfolio (which might not be on the portfolio frontier) and let p be any frontier portfolio.

$$\text{cov}(\tilde{r}_p, \tilde{r}_q) \underbrace{=}_{\text{by definition}} w_p^T V w_q$$

$$\begin{aligned} &= [\lambda V^{-1}e + \gamma V^{-1}\mathbf{1}]^T V w_q \\ &= \lambda e^T V^{-1} V w_q + \gamma \mathbf{1}^T V^{-1} V w_q \\ &= \lambda e^T w_q + \gamma \text{ (since } \mathbf{1}^T w_q = \sum_{i=1}^N w_q^i \equiv 1) \end{aligned} \quad (7.22)$$

$$= \lambda E(\tilde{r}_q) + \gamma \text{ (since } e^T w_q = \sum_{i=1}^N E(\tilde{r}_i) w_q^i \equiv E(\tilde{r}_q)) \quad (7.23)$$

where $\lambda = \frac{CE(\tilde{r}_p) - A}{D}$ and $\gamma = \frac{B - AE(\tilde{r}_p)}{D}$, as per earlier definitions.

Substituting these expressions into Equation (7.23) gives

$$\text{cov}(\tilde{r}_p, \tilde{r}_q) = \frac{CE(\tilde{r}_p) - A}{D} E(\tilde{r}_q) + \frac{B - AE(\tilde{r}_p)}{D}. \quad (7.24)$$

Equation (7.24) is a short step from Equation (7.19): Collect all terms involving expected returns, add and subtract $\frac{A^2 C}{DC^2}$ to get the first term in Equation (7.19) with a remaining term equal to $+\frac{1}{C}(\frac{BC}{D} - \frac{A^2}{D})$. But the latter is simply $1/C$ since $D = BC - A^2$.

Let us go back to Equation (7.23) and apply it to the case where q is $ZC(p)$; one gets

$$0 = \text{cov}(\tilde{r}_p, \tilde{r}_{ZC(p)}) = \lambda E(\tilde{r}_{ZC(p)}) + \gamma \text{ or } \gamma = -\lambda E(\tilde{r}_{ZC(p)}); \quad (7.25)$$

hence Equation (7.23) becomes

$$\text{cov}(\tilde{r}_p, \tilde{r}_q) = \lambda [E(\tilde{r}_q) - E(\tilde{r}_{ZC(p)})]. \quad (7.26)$$

Apply the later to the case $p = q$ to get

$$\sigma_p^2 = \text{cov}(\tilde{r}_p, \tilde{r}_p) = \lambda [E(\tilde{r}_p) - E(\tilde{r}_{ZC(p)})]; \quad (7.27)$$

and divide Equation (7.26) by Equation (7.27) and rearrange to obtain

$$E(\tilde{r}_q) = E(\tilde{r}_{ZC(p)}) + \beta_{pq} [E(\tilde{r}_p) - E(\tilde{r}_{ZC(p)})]. \quad (7.28)$$

This equation bears more than a passing resemblance to the *Security Market Line* (SML) implication of the capital asset pricing model. But as yet it is simply a statement about the various portfolios that can be created from arbitrary collections of assets: (1) pick any frontier portfolio p ; (2) this defines an associated zero-covariance portfolio $ZC(p)$; (3) any other portfolio q 's expected return can be expressed in terms of the returns to those portfolios and the covariance of q with the arbitrarily chosen frontier portfolio. Equation (7.28) would *very closely* resemble the security market line if, in particular, we could choose $p = M$, the market portfolio of existing assets. The circumstances under which it is possible to do this form the subject to which we now turn.

7.7 The Zero-Beta Capital Asset Pricing Model

We would like to explain asset expected returns *in equilibrium*. The relationship in Equation (7.28), however, is not the consequence of an equilibrium theory because it was derived for a *given* particular vector of expected asset returns, e , and a given covariance-variance matrix, V . In fact, it is the vector of returns e that we would like, in equilibrium, to understand. We need to identify a particular portfolio as being a frontier portfolio without specifying a priori the (expected) return vector and variance-covariance matrix of its constituent assets. The zero-beta CAPM tells us that under certain assumptions, this desired portfolio can be identified as the market portfolio M .

We may assume one of the following:

(i) agents maximize expected utility with increasing and strictly concave utility of money functions and asset returns are multivariate normally distributed, or

(ii) each agent chooses a portfolio with the objective of maximizing a derived utility function of the form $W(e, \sigma^2)$, $W_1 > 0$, $W_2 < 0$, W concave.

In addition, we assume that all investors have a common time horizon and homogeneous beliefs about e and V .

Under either set of assumptions, investors will only hold mean-variance efficient frontier portfolios⁵. But this implies that, **in equilibrium**, the market portfolio, which is a convex combination of individual portfolios is also on the efficient frontier.⁶

⁵Recall the demonstration in Section 6.3

⁶Note that, in the standard version of the CAPM, the analogous claim crucially depended on the existence of a risk-free asset.

Therefore, in Equation (7.22), p can be chosen to be M , the portfolio of all risky assets, and Equation (7.28) can, therefore, be expressed as:

$$E(\tilde{r}_q) = E(\tilde{r}_{ZC(M)}) + \beta_{Mq} [E(\tilde{r}_M) - E(\tilde{r}_{ZC(M)})] \quad (7.29)$$

The relationship in Equation (7.29) holds for any portfolio q , whether it is a frontier portfolio or not. This is the zero-beta CAPM.

An individual asset j is also a portfolio, so Equation (7.29) applies to it as well:

$$E(\tilde{r}_j) = E(\tilde{r}_{ZC(M)}) + \beta_{Mj} [E(\tilde{r}_M) - E(\tilde{r}_{ZC(M)})] \quad (7.30)$$

The zero-beta CAPM (and the more familiar *Sharpe-Lintner-Mossin* CAPM)⁷ is an equilibrium theory: The relationships in Equations (7.29) and (7.30) hold in equilibrium. In equilibrium, investors will not be maximizing utility unless they hold efficient portfolios. Therefore, the market portfolio is efficient; we have identified one efficient frontier portfolio, and we can apply Equation (7.29). By contrast, Equation (7.28) is a pure mathematical relationship with no economic content; it simply describes relationships between frontier portfolio returns and the returns from any other portfolio of the same assets.

As noted in the introduction, the zero-beta CAPM does not, however, describe the process to or by which equilibrium is achieved. In other words, the process by which agents buy and sell securities in their desire to hold efficient portfolios, thereby altering security prices and thus expected returns, and requiring further changes in portfolio composition is not present in the model. When this process ceases and all agents are optimizing given the prevailing prices, then all will be holding efficient portfolios given the equilibrium expected returns e and covariance-variance matrix V . Thus M is also efficient.

Since, in equilibrium, agents desired holdings of securities coincide with their actual holdings, we can identify M as the actual portfolio of securities held in the market place. There are many convenient approximations to M —the S&P 500 index of stocks being the most popular in the United States. The usefulness of these approximations, which are needed to give empirical content to the CAPM, is, however, debatable as discussed in our concluding comments.

As a final remark, let us note that the name “zerobeta CAPM” comes from the fact that $\beta_{ZC(M),M} = \frac{\text{cov}(\tilde{r}_M, \tilde{r}_{ZC(M)})}{\sigma_{ZC(M)}^2} = 0$, by construction of $ZC(M)$; in other words, the beta of $ZC(M)$ is zero.

7.8 The Standard CAPM

Our development thus far did not admit the option of a risk-free asset. We need to add this if we are to achieve the standard form CAPM. On a purely formal basis, of course, a risk-free asset has zero covariance with M and thus $r_f = E(\tilde{r}_{ZC(M)})$. Hence we could replace $E(\tilde{r}_{ZC(M)})$ with r_f in Equation

⁷Sharpe (1964), Lintner (1965), and Mossin (1966).

(7.30) to obtain the standard representation of the CAPM, the SML. But this approach is not entirely appropriate since the derivation of Equation (7.30) presumed the absence of any such risk-free asset.

More formally, the addition of a risk-free asset substantially alters the shape of the set of frontier portfolios in the $[E(\tilde{r}), \sigma(\tilde{r})]$ space. Let us briefly outline the development here, which closely resembles what is done above. Consider N risky assets with expected return vector e , and one risk-free asset, with *expected* return $\equiv r_f$. Let p be a frontier portfolio and let w_p denote the N vector of portfolio weights on the risky assets of p ; w_p in this case is the solution to:

$$\begin{aligned} \min_w & \frac{1}{2} w^T V w \\ \text{s.t.} & w^T e + (1 - w^T \mathbf{1}) r_f = E \end{aligned}$$

Solving this problem gives

$$w_p = V^{-1}(e - r_f \mathbf{1}) \frac{E - r_f}{H}$$

where $H = B - 2Ar_f + Cr_f^2$ and A, B, C are defined as before.

Let us examine this expression for w_p more carefully:

$$w_p = \underbrace{\underbrace{V^{-1}}_{n \times n} \underbrace{(e - r_f \mathbf{1})}_{n \times 1}}_{n \times 1} \underbrace{\frac{E(\tilde{r}_p) - r_f}{H}}_{a \text{ number}} \quad (7.31)$$

This expression tells us that if we wish to have a higher expected return, we should invest proportionally the same amount more in each risky asset so that the relative proportions of the risky assets remain unchanged. These proportions are defined by the $V^{-1}(e - r_f \mathbf{1})$ term. This is exactly the result we were intuitively expecting: Graphically, we are back to the linear frontier represented in Figure 7.1.

The weights w_p uniquely identify the tangency portfolio T . Also,

$$\sigma^2(\tilde{r}_p) = w_p^T V w_p = \frac{[E(\tilde{r}_p) - r_f]^2}{H}, \text{ and} \quad (7.32)$$

$$\text{cov}(\tilde{r}_q, \tilde{r}_p) = w_q^T V w_p = \frac{[E(\tilde{r}_q) - r_f][E(\tilde{r}_p) - r_f]}{H} \quad (7.33)$$

for any portfolio q and any frontier portfolio p . Note how all this parallels what we did before. Solving Equation (7.33) for $E(\tilde{r}_q)$ gives:

$$E(\tilde{r}_q) - r_f = \frac{H \text{cov}(\tilde{r}_q, \tilde{r}_p)}{E(\tilde{r}_p) - r_f} \quad (7.34)$$

Substituting for H via Equation (7.32) yields

$$E(\tilde{r}_q) - r_f = \frac{\text{cov}(\tilde{r}_q, \tilde{r}_p)}{E(\tilde{r}_p) - r_f} \frac{[E(\tilde{r}_p) - r_f]^2}{\sigma^2(\tilde{r}_p)}$$

or

$$E(\tilde{r}_q) - r_f = \frac{\text{cov}(\tilde{r}_q, \tilde{r}_p)}{\sigma^2(\tilde{r}_p)} [E(\tilde{r}_p) - r_f] \quad (7.35)$$

Again, since T is a frontier portfolio, we can choose $p \equiv T$. But in equilibrium $T = M$; in this case, Equation (7.35) gives:

$$E(\tilde{r}_q) - r_f = \frac{\text{cov}(\tilde{r}_q, \tilde{r}_M)}{\sigma^2(\tilde{r}_M)} [E(\tilde{r}_M) - r_f],$$

or

$$E(\tilde{r}_q) = r_f + \beta_{qM} [E(\tilde{r}_M) - r_f] \quad (7.36)$$

for any asset (or portfolio) q . This is the standard CAPM.

Again, let us review the flow of logic that led to this conclusion. First, we identified the efficient frontier of risk-free and risky assets. This efficient frontier is fully characterized by the risk-free asset and a specific tangency frontier portfolio. The latter is identified in Equation (7.31). We then observed that all investors, in equilibrium under homogeneous expectations, would hold combinations of the risk-free asset and that portfolio. Thus it must constitute the *market* — the portfolio of all risky assets. It is these latter observations that give the CAPM its empirical content.

7.9 Conclusions

Understanding and identifying the determinants of equilibrium asset returns is inherently an overwhelmingly complex problem. In order to make some progress, we have made, in the present chapter, a number of simplifying assumptions that we will progressively relax in the future.

1. Rather than deal with fully described probability distributions on returns, we consider only the first two moments, $E(\tilde{r}_p)$ and $\sigma^2(\tilde{r}_p)$. When returns are at least approximately normally distributed, it is natural to think first of characterizing return distributions by their means and variances, since $\text{Prob}(\mu_r - 2\sigma_r \leq \tilde{r} \leq \mu_r + 2\sigma_r) = 0.95$, for the normal distribution: plus or minus two standard deviations from the mean will encompass nearly all of the probability. It is also natural to try to estimate these distributions and their moments from historical data. To do this naively would be to assign equal probability to each past observation. Yet, we suspect that more recent observations should contain more relevant information concerning the true distribution than observations in the distant past. Indeed the entire distribution may be shifting through time; that is, it may be nonstationary. Much current research is devoted to studying what and how information can be extracted from historical data in this setting.

2. The model is static; in other words, only one period of returns are measured and analyzed. The defined horizon is assumed to be equally relevant for all investors.

3. Homogeneous expectations: all investors share the same information. We know this assumption cannot, in fact, be true. Anecdotally, different security analysts produce reports on the same stock that are wildly different. More objectively, the observed volume of trade on the stock exchanges is much higher than is predicted by trading models with assumed homogeneous expectations.

The CAPM is at the center of modern financial analysis. As with modern portfolio theory, its first and foremost contribution is conceptual: It has played a major role in helping us to organize our thoughts on the key issue of equilibrium asset pricing. Beyond that, it is, in principle, a testable theory, and indeed, a huge amount of resources has been devoted to testing it. Since the above-mentioned assumptions do not hold up in practice, it is not surprising that empirical tests of the CAPM come up short.⁸ One example is the Roll (1977) critique. Roll reminds us that the CAPM's view of the market portfolio is that it contains every asset. Yet data on asset returns is not available for many assets. For example, no systematic data is available on real estate and, in the United States at least, one estimates that approximately one-half of total wealth is invested in real estate. Thus it is customary to use proxies for the true M in conducting tests of CAPM. Roll demonstrates, however, that even if two potential proxies for M are correlated greater than 0.9, the beta estimates obtained using each may be very different. This suggests that the empirical implications of the model are very sensitive to the choice of proxy. With no theory to inform us as to what proxy to use, the applicability of the theory is suspect.

Furthermore, beginning in the late 1970s and continuing to the present, more and more evidence has come to light suggesting that firm characteristics beyond beta may provide explanatory power for mean equity returns. In particular, various studies have demonstrated that a firm's average equity returns are significantly related to its size (as measured by the aggregate market value of equity), the ratio of book value per share to market value per share for its common equity, its equity price to earnings ratio, its cash flow per share to price per share ratio, and its historical sales growth. These relationships contradict strict CAPM, which argues that only a stock's systematic risk should matter for its returns; as such they are referred to as anomalies. In addition, even in models that depart from the CAPM assumptions, there is little theoretical evidence as to why these particular factors should be significant.

We close this chapter by illustrating these ideas with brief summaries of two especially prominent recent papers. Fama and French (1992) showed that the relationship between market betas and average returns is essentially flat for their sample period (1963 to 1990). In other words, their results suggest that the single factor CAPM can no longer explain the cross-sectional variation in

⁸We have chosen not to systematically review this literature. Standard testing procedures and their results are included in the introductory finance manuals which are pre-requisites for the present text. Advanced issues properly belong to financial econometrics courses. The student wishing to invest in this area should consult Jensen (1979) and Friend, Westerfield and Granito (1979) for early surveys ; Ferson and Jagannathan (1996) and Shanken (1996) for more recent ones.

equity returns. They also find that for this sample period the univariate (single factor) relationship between average stock returns and size (market value of equity), leverage, earnings-to-price ratio, and book-to-market value of equity per share are strong. More specifically, there is a negative relationship between size and average return, which is robust to the inclusion of other variables. There is also a consistent positive relationship between average returns and book-to-market ratio, which is not swamped by the introduction of other variables. They find that the combination of size and the book-to-market ratio as explanatory variables appears, for their sample period, to subsume the explanatory roles of leverage and the price-to-earnings ratio.

In a related paper, Fama and French (1993) formalize their size and book-to-market ratio factors more precisely by artificially constructing two-factor portfolios to which they assign acronyms, HML (high-medium-low) and SMB (small-medium-big). Both portfolios consist of a joint long and short position and have net asset value zero. The HML portfolio represents a combination of a long position in high book-to-market stocks with a short position in low book-to-market stocks. The SMB portfolio is one consisting of a long position in small capitalization stocks and a short position in large capitalization stocks. These designations are, of course, somewhat arbitrary.⁹ In conjunction with the excess (above the risk-free rate) return on a broad-based index, Fama and French study the ability of these factors to explain cross-sectional stock returns. They find that their explanatory power is highly significant.

References

- Fama, E., French, K. (1992), "The Cross Section of Expected Stock Returns," *Journal of Finance* 47, 427-465.
- Fama, E., French, K. (1993), "Common Risk Factors in the Returns on Stocks and Bonds," *Journal of Financial Economics* 33, 3-56.
- Ferson, W. E., Jagannathan, R. (1996), "Econometric Evaluation of Asset Pricing Models," in *Statistical Methods in Finance, Handbook of Statistics*, Maddala, G.S., Rao, C.R., eds., Amsterdam: North Holland, 14.

⁹Short description on how to construct SMB, HML:

In June each year, all NYSE stocks are ranked by size. The median NYSE size is used to split all NYSE, AMEX, NASDAQ firms into two groups, small and big. All NYSE, AMEX, NASDAQ stocks are also broken into three BE/ME equity groups based on the breakpoints for the bottom 30% (low), middle 40% (medium), and top 30% (high) of the ranked values of BE/ME (Book value of Equity/Market value of Equity) for NYSE stocks. Fama and French (1993) then construct six portfolios (S/L, S/M, S/H, B/L, B/M, B/H) from the intersection of the two ME groups and the three BE/ME groups. SMB is the difference between the simple average of returns on three small stocks portfolio (S/L, S/M, S/H) and the three big stock portfolios (B/L, B/M, B/H). SMB mimics the risk factor in return related to size. Accordingly, HML is the difference of the simple average of the returns on the two high BE/ME portfolios (S/H, B/H) and the average of the returns on the two low BE/ME portfolios (S/L, B/L). HML mimics the risk factor in returns that is related to BE/ME.

- Friend, I., Westerfield, R., Granito, M. (1979) “New Evidence on the Capital Asset Pricing Model,” in *Handbook of Financial Economics*, Bicksler, J.L., ed., North Holland.
- Jensen, M. (1979), “Tests of Capital Market Theory and Implications of the Evidence,” in *Handbook of Financial Economics*, Bicksler, J.L., ed., North Holland.
- Lintner, J. (1965), “The Valuation of Risky Assets and the Selection of Risky Investments in Stock Portfolios and Capital Budgets,” *Review of Economics and Statistics* 47(1), 13—37.
- Mossin, J. (1966), “Equilibrium in a Capital Asset Market,” *Econometrica* 34(4), 768—783.
- Roll, R. (1977), “A Critique of the Asset Pricing Theory’s Test—Part I: On Past and Potential Testability of the Theory,” *Journal of Financial Economics* 4, 129-76.
- Shanken, J. (1996), “Statistical Methods in Tests of Portfolio Efficiency: A Synthesis” in *Statistical Methods in Finance, Handbook of Statistics*, Madala, G.S., Rao, C.R., eds., Amsterdam: North Holland, 14.
- Sharpe, W. F. (1964), “Capital Asset Prices: A Theory of Market Equilibrium under Conditions of Risk,” *Journal of Finance*, 19(3).

Appendix 7.1:

Proof of the CAPM Relationship

Refer to Figure 7.1. Consider a portfolio with a fraction $1 - \alpha$ of wealth invested in an arbitrary security j and a fraction α in the market portfolio.

$$\begin{aligned}\bar{r}_p &= \alpha \bar{r}_M + (1 - \alpha) \bar{r}_j \\ \sigma_p^2 &= \alpha^2 \sigma_M^2 + (1 - \alpha)^2 \sigma_j^2 + 2\alpha(1 - \alpha) \sigma_{jM}\end{aligned}$$

As α varies we trace a locus that

- passes through M
- (- and through j)
- cannot cross the CML (why?)
- hence must be tangent to the CML at M

Tangency = $\frac{d\bar{r}_p}{d\sigma_p} \big|_{\alpha=1}$ = slope of the locus at M = slope of CML = $\frac{\bar{r}_M - r_f}{\sigma_M}$

$$\begin{aligned}
\frac{d\bar{r}_p}{d\sigma_p} &= \frac{d\bar{r}_p/d\alpha}{d\sigma_p/d\alpha} \\
\frac{d\bar{r}_p}{d\alpha} &= \bar{r}_M - \bar{r}_j \\
2\sigma_p \frac{d\sigma_p}{d\alpha} &= 2\alpha\sigma_M^2 - 2(1-\alpha)\sigma_j^2 + 2(1-2\alpha)\sigma_{jM} \\
\frac{d\bar{r}_p}{d\sigma_p} &= \frac{(\bar{r}_M - \bar{r}_j)\sigma_p}{\alpha\sigma_M^2 - (1-\alpha)\sigma_j^2 + (1-2\alpha)\sigma_{jM}} \\
\left. \frac{d\bar{r}_p}{d\sigma_p} \right|_{\alpha=1} &= \frac{(\bar{r}_M - \bar{r}_j)\sigma_M}{\sigma_M^2 - \sigma_{jM}} = \frac{\bar{r}_M - \bar{r}_f}{\sigma_M} \\
(\bar{r}_M - \bar{r}_j) &= \frac{(\bar{r}_M - r_f)(\sigma_M^2 - \sigma_{jM})}{\sigma_M^2} \\
(\bar{r}_M - \bar{r}_j) &= (\bar{r}_M - r_f) \left(1 - \frac{\sigma_{jM}}{\sigma_M^2} \right) \\
\bar{r}_j &= r_f + (\bar{r}_M - r_f) \frac{\sigma_{jM}}{\sigma_M^2}
\end{aligned}$$

Appendix 7.2:

The Mathematics of the Portfolio Frontier: An Example

Assume $e = \begin{pmatrix} \bar{r}_1 \\ \bar{r}_2 \end{pmatrix} = \begin{pmatrix} 1 \\ 2 \end{pmatrix}$; $V = \begin{pmatrix} 1 & -1 \\ -1 & 4 \end{pmatrix}$ i.e., $\rho_{12} = \rho_{21} = -1/2$

Therefore,

$$V^{-1} = \begin{pmatrix} 4/3 & 1/3 \\ 1/3 & 1/3 \end{pmatrix};$$

check:

$$\begin{pmatrix} 1 & -1 \\ -1 & 4 \end{pmatrix} \begin{pmatrix} 4/3 & 1/3 \\ 1/3 & 1/3 \end{pmatrix} = \begin{pmatrix} 4/3 - 1/3 & 1/3 - 1/3 \\ -4/3 + 4/3 & -1/3 + 4/3 \end{pmatrix} = \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix}$$

$$\begin{aligned}
A &= 1^T V^{-1} e = \begin{pmatrix} 1 & 1 \end{pmatrix} \begin{pmatrix} 4/3 & 1/3 \\ 1/3 & 1/3 \end{pmatrix} \begin{pmatrix} 1 \\ 2 \end{pmatrix} = \begin{pmatrix} \underbrace{4/3 + 1/3}_{5/3} & \underbrace{1/3 + 1/3}_{2/3} \end{pmatrix} \begin{pmatrix} 1 \\ 2 \end{pmatrix} \\
&= 5/3 + 2(2/3) = 3
\end{aligned}$$

$$B = e^T V^{-1} e = \begin{pmatrix} 1 & 2 \end{pmatrix} \begin{pmatrix} 4/3 & 1/3 \\ 1/3 & 1/3 \end{pmatrix} \begin{pmatrix} 1 \\ 2 \end{pmatrix} = \begin{pmatrix} 4/3 + 2/3 & 1/3 + 2/3 \end{pmatrix} \begin{pmatrix} 1 \\ 2 \end{pmatrix} = 4$$

$$C = 1^T V^{-1} 1 = \begin{pmatrix} 1 & 1 \end{pmatrix} \begin{pmatrix} 4/3 & 1/3 \\ 1/3 & 1/3 \end{pmatrix} \begin{pmatrix} 1 \\ 1 \end{pmatrix} = \begin{pmatrix} 5/3 & 2/3 \end{pmatrix} \begin{pmatrix} 1 \\ 1 \end{pmatrix} = 7/3$$

$$D = BC - A^2 = 4(7/3) - 9 = 28/3 - 27/3 = 1/3$$

Now we can compute g and h :

1.

$$\begin{aligned}
g &= \frac{1}{D} [B(V^{-1}1) - A(V^{-1}e)] \\
&= \frac{1}{1/3} \left[4 \begin{pmatrix} 4/3 & 1/3 \\ 1/3 & 1/3 \end{pmatrix} \begin{pmatrix} 1 \\ 1 \end{pmatrix} - 3 \begin{pmatrix} 4/3 & 1/3 \\ 1/3 & 1/3 \end{pmatrix} \begin{pmatrix} 1 \\ 2 \end{pmatrix} \right] \\
&= 3 \left[4 \begin{pmatrix} 5/3 \\ 2/3 \end{pmatrix} - 3 \begin{pmatrix} 6/3 \\ 3/3 \end{pmatrix} \right] = 3 \left[\begin{pmatrix} 20/3 \\ 8/3 \end{pmatrix} - \begin{pmatrix} 18/3 \\ 9/3 \end{pmatrix} \right] \\
&= \left[\begin{pmatrix} 20 \\ 8 \end{pmatrix} - \begin{pmatrix} 18 \\ 9 \end{pmatrix} \right] = \begin{pmatrix} 2 \\ -1 \end{pmatrix}
\end{aligned}$$

2.

$$\begin{aligned}
h &= \frac{1}{D} [C(V^{-1}e) - A(V^{-1}1)] \\
&= \frac{1}{1/3} \left[\frac{7}{3} \begin{pmatrix} 4/3 & 1/3 \\ 1/3 & 1/3 \end{pmatrix} \begin{pmatrix} 1 \\ 2 \end{pmatrix} - 3 \begin{pmatrix} 4/3 & 1/3 \\ 1/3 & 1/3 \end{pmatrix} \begin{pmatrix} 1 \\ 2 \end{pmatrix} \right] \\
&= 3 \left[\frac{7}{3} \begin{pmatrix} 2 \\ 1 \end{pmatrix} - 3 \begin{pmatrix} 5/3 \\ 2/3 \end{pmatrix} \right] = 7 \begin{pmatrix} 2 \\ 1 \end{pmatrix} - 3 \begin{pmatrix} 5/3 \\ 2/3 \end{pmatrix} = \begin{pmatrix} 14 \\ 7 \end{pmatrix} - \begin{pmatrix} 15 \\ 6 \end{pmatrix} = \begin{pmatrix} -1 \\ 1 \end{pmatrix}
\end{aligned}$$

Check by recovering the two initial assets; suppose $E(\tilde{r}_p) = 1$:

$$\begin{pmatrix} w_1 \\ w_2 \end{pmatrix} = \begin{pmatrix} 2 \\ -1 \end{pmatrix} + \begin{pmatrix} -1 \\ 1 \end{pmatrix} E(\tilde{r}_p) = \begin{pmatrix} 2 \\ -1 \end{pmatrix} + \begin{pmatrix} -1 \\ 1 \end{pmatrix} = \begin{pmatrix} 1 \\ 0 \end{pmatrix} \Rightarrow \text{OK}$$

suppose $E(\tilde{r}_p) = 2$:

$$\begin{pmatrix} w_1 \\ w_2 \end{pmatrix} = \begin{pmatrix} 2 \\ -1 \end{pmatrix} + \begin{pmatrix} -1 \\ 1 \end{pmatrix} 2 = \begin{pmatrix} 2 \\ -1 \end{pmatrix} + \begin{pmatrix} -2 \\ 2 \end{pmatrix} = \begin{pmatrix} 0 \\ 1 \end{pmatrix} \Rightarrow \text{OK}$$

The equation corresponding to Equation (7.16) thus reads:

$$\begin{pmatrix} w_1^p \\ w_2^p \end{pmatrix} = \begin{pmatrix} 2 \\ -1 \end{pmatrix} + \begin{pmatrix} -1 \\ 1 \end{pmatrix} E(\tilde{r}_p).$$

Let us compute the minimum variance portfolio for these assets.

$$E(\tilde{r}_{p, \min \text{ var}}) = \frac{A}{C} = \frac{9}{7}$$

$$\sigma^2(\tilde{r}_{p, \min \text{ var}}) = \frac{1}{C} = \frac{3}{7} < \min \{ 1, 4 \}$$

$$w^p = \begin{pmatrix} 2 \\ -1 \end{pmatrix} + \begin{pmatrix} -1 \\ 1 \end{pmatrix} \frac{9}{7} = \begin{pmatrix} 2 \\ -1 \end{pmatrix} + \begin{pmatrix} -9/7 \\ 9/7 \end{pmatrix} = \begin{pmatrix} 14/7 \\ -7/7 \end{pmatrix} + \begin{pmatrix} -9/7 \\ 9/7 \end{pmatrix} = \begin{pmatrix} 5/7 \\ 2/7 \end{pmatrix}$$

Let's check $\sigma^2(\tilde{r}_p)$ by computing it another way:

$$\sigma_p^2 = \begin{pmatrix} 5/7 & 2/7 \end{pmatrix} \begin{pmatrix} 1 & -1 \\ -1 & 4 \end{pmatrix} \begin{pmatrix} 5/7 \\ 2/7 \end{pmatrix} = \begin{pmatrix} 3/7 & 3/7 \end{pmatrix} \begin{pmatrix} 5/7 \\ 2/7 \end{pmatrix} = 3/7 \Rightarrow \text{OK}$$

Chapter 8 : Arrow-Debreu Pricing, Part I

8.1 Introduction

As interesting and popular as it is, the CAPM is a very limited theory of equilibrium pricing and we will devote the next chapters to reviewing alternative theories, each of which goes beyond the CAPM in one direction or another. The Arrow-Debreu pricing theory discussed in this chapter is a full general equilibrium theory as opposed to the partial equilibrium static view of the CAPM. Although also static in nature, it is applicable to a multi-period setup and can be generalized to a broad set of situations. In particular, it is free of any preference restrictions, and of distributional assumptions on returns. The Consumption CAPM considered subsequently (Chapter 9) is a fully dynamic construct. It is also an equilibrium theory, though of a somewhat specialized nature. With the Risk Neutral Valuation Model and the Arbitrage Pricing Theory (APT), taken up in Chapters 11 to 13, we will be moving into the domain of arbitrage-based theories, after observing, however, that the Arrow-Debreu pricing theory itself may also be interpreted in the arbitrage perspective (Chapter 10).

The Arrow-Debreu model takes a more standard equilibrium view than the CAPM: It is explicit in stating that equilibrium means supply equals demand in every market. It is a very general theory accommodating production and, as already stated, very broad hypotheses on preferences. Moreover, no restriction on the distribution of returns is necessary. We will not, however, fully exploit the generality of the theory: In keeping with the objective of this text, we shall often limit ourselves to illustrating the theory with examples.

We will be interested to apply it to the equilibrium pricing of securities, especially the pricing of complex securities that pay returns in many different time periods and states of nature, such as common stocks or 30-year government coupon bonds. The theory will, as well, enrich our understanding of project valuation because of the formal equivalence, underlined in Chapter 2, between a project and an asset. In so doing we will be moving beyond a pure equilibrium analysis and start using the concept of arbitrage. It is in the light of a set of no-arbitrage relationships that the Arrow-Debreu pricing takes its full force. This perspective on the Arrow-Debreu theory will be developed in Chapter 10.

8.2 Setting: An Arrow-Debreu Economy

In the basic setting that we shall use, the following parameters apply:

1. There are two dates: 0, 1. This setup, however, is fully generalizable to multiple periods; see the later remark.
2. There are N possible states of nature at date 1, which we index by $\theta = 1, 2, \dots, N$ with probabilities π_θ ;
3. There is one perishable (non-storable) consumption good.

4. There are K agents, indexed by $k = 1, \dots, K$, with preferences:

$$U_0^k(c_0^k) + \delta^k \sum_{\theta=1}^N \pi_\theta U^k(c_\theta^k);$$

5. Agent k 's endowment is described by the vector $\{e_0^k, (e_\theta^k)_{\theta=1,2,\dots,N}\}$.

In this description, c_θ^k denotes agent k 's consumption of the sole consumption good in state θ , U is the real-valued utility representation of agent k 's period preferences, and δ^k is the agent's time discount factor. In fact, the theory allows for more general preferences than the time-additive expected utility form. Specifically, we could adopt the following representation of preferences:

$$u^k(c_0^k, c_{\theta_1}^k, c_{\theta_2}^k, \dots, c_{\theta_N}^k).$$

This formulation allows not only for a different way of discounting the future (implicit in the relative taste for present consumption relative to all future consumptions), but it also permits heterogeneous, subjective views on the state probabilities (again implicit in the representation of relative preference for, say, $c_{\theta_2}^k$ vs. $c_{\theta_3}^k$). In addition, it assumes neither time-additivity, nor an expected utility representation. Since our main objective is not generality, we choose to work with the less general, but easier to manipulate time-additive expected utility form.

In this economy, the only traded securities are of the following type: One unit of security θ , with price q_θ , pays one unit of consumption if state θ occurs and nothing otherwise. Its payout can thus be summarized by a vector with all entries equal to zero except for column θ where the entry is 1: $(0, \dots, 0, 1, 0, \dots, 0)$. These primitive securities are called **Arrow-Debreu securities**,¹ or *state contingent claims* or simply *state claims*. Of course, the consumption of any individual k if state θ occurs equals the number of units of security θ that he holds. This follows from the fact that buying the relevant contingent claim is the only way for a consumer to secure purchasing power at a future date-state (recall that the good is perishable). An agent's decision problem can then be characterized by:

$$\begin{aligned} & \max_{(c_0^k, c_1^k, \dots, c_N^k)} U_0^k(c_0^k) + \delta^k \sum_{\theta=1}^N \pi_\theta U^k(c_\theta^k) \\ & \text{s.t.} \\ & c_0^k + \sum_{\theta=1}^N q_\theta c_\theta^k \leq e_0^k + \sum_{\theta=1}^N q_\theta e_\theta^k \\ & c_0^k, c_1^k, \dots, c_N^k \geq 0 \end{aligned} \tag{P}$$

The first inequality constraint will typically hold with equality in a world of non-satiation. That is, the total value of goods and security purchases made by

¹So named after the originators of modern equilibrium theory: see Arrow (1951) and Debreu (1959).

the agent (the left-hand side of the inequality) will exhaust the total value of his endowments (the right-hand side).

Equilibrium for this economy is a set of contingent claim prices (q_1, q_2, \dots, q_N) such that

1. at those prices (c_0^k, \dots, c_N^k) solve problem (P), for all k , and
2. $\sum_{k=1}^K c_0^k = \sum_{k=1}^K e_0^k$, $\sum_{k=1}^K c_\theta^k = \sum_{k=1}^K e_\theta^k$, for every θ .

Note that here the agents are solving for desired future and present consumption holdings rather than holdings of Arrow-Debreu securities. This is justified because, as just noted, there is a one-to-one relationship between the amount consumed by an individual in a given state θ and his holdings of the Arrow-Debreu security corresponding to that particular state θ , the latter being a promise to deliver one unit of the consumption good if that state occurs.

Note also that there is nothing in this formulation that inherently restricts matters to two periods, if we define our notion of a state, somewhat more richly, as a *date-state* pair. Consider three periods, for example. There are N possible states in date one and J possible states in date two, irrespective of the state achieved in date 1. Define $\hat{\theta}$ new states to be of the form $\hat{\theta}_s = (j, \theta_k^j)$, where j denotes the state in date 1 and θ_k^j denotes the state k in date 2, conditional that state j was observed in date 1 (Refer to Figure 8.1). So $(1, \theta_5^1)$ would be a state and $(2, \theta_3^2)$ another state. Under this interpretation, the number of *states* expands to $1 + NJ$, with:

- | | | |
|-----|---|-----------------------------|
| 1 | : | the date 0 state |
| N | : | the number of date-1 states |
| J | : | the number of date-2 states |

Insert Figure 8.1

With minor modifications, we can thus accommodate many periods and states. In this sense, our model is fully general and can represent as complex an environment as we might desire. In this model, the real productive side of the economy is in the background. We are, in effect, viewing that part of the economy as invariant to securities trading. The unusual and unrealistic aspect of this economy is that all trades occur at $t = 0$.² We will relax this assumption in Chapter 9.

8.3 Competitive Equilibrium and Pareto Optimality Illustrated

Let us now develop an example. The essentials are found in Table 8.1.

²Interestingly, this is less of a problem for project valuation than for asset pricing.

Table 8.1: Endowments and Preferences In Our Reference Example

Agents	Endowments		Preferences
	$t = 0$	$t = 1$	
		$\theta_1 \quad \theta_2$	
Agent 1	10	1 2	$\frac{1}{2}c_0^1 + 0.9 \left(\frac{1}{3} \ln(c_1^1) + \frac{2}{3} \ln(c_2^1) \right)$
Agent 2	5	4 6	$\frac{1}{2}c_0^2 + 0.9 \left(\frac{1}{3} \ln(c_1^2) + \frac{2}{3} \ln(c_2^2) \right)$

There are two dates and, at the future date, two possible states of nature with probabilities $1/3$ and $2/3$. It is an exchange economy and the issue is to share the existing endowments between two individuals. Their (identical) preferences are linear in date 0 consumption with constant marginal utility equal to $1/2$. This choice is made for ease of computation, but great care must be exercised in interpreting the results obtained in such a simplified framework. Date 1 preferences are concave and identical. The discount factor is $.9$. Let q_1 be the price of a unit of consumption in date 1 state 1, q_2 the price of one unit of the consumption good in date 1 state 2. We will solve for optimal consumption directly, knowing that this will define the equilibrium holdings of the securities. The prices of these consumption goods coincide with the prices of the corresponding state-contingent claims; period 0 consumption is taken as the numeraire and its price is 1. This means that all prices are expressed in units of period 0 consumption: q_1, q_2 are prices for the consumption good at date 1, in states 1 and 2, respectively, measured in units of date 0 consumption. They can thus be used to add up or compare units of consumption at different dates and in different states, making it possible to add different date cash flows, with the q_i being the appropriate weights. This, in turn, permits computing an individual's wealth. Thus, in the previous problem, agent 1's wealth, which equals the present value of his current and future endowments, is $10 + 1q_1 + 2q_2$ while agent 2's wealth is $5 + 4q_1 + 6q_2$.

The respective agent problems are:

$$\begin{aligned} \text{Agent 1:} \quad & \max \frac{1}{2} (10 + 1q_1 + 2q_2 - c_1^1 q_1 - c_2^1 q_2) + 0.9 \left(\frac{1}{3} \ln(c_1^1) + \frac{2}{3} \ln(c_2^1) \right) \\ \text{s.t.} \quad & c_1^1 q_1 + c_2^1 q_2 \leq 10 + q_1 + 2q_2, \text{ and } c_1^1, c_2^1 \geq 0 \end{aligned}$$

$$\begin{aligned} \text{Agent 2:} \quad & \max \frac{1}{2} (5 + 4q_1 + 6q_2 - c_1^2 q_1 - c_2^2 q_2) + 0.9 \left(\frac{1}{3} \ln(c_1^2) + \frac{2}{3} \ln(c_2^2) \right) \\ \text{s.t.} \quad & c_1^2 q_1 + c_2^2 q_2 \leq 5 + 4q_1 + 6q_2 \text{ and } c_1^2, c_2^2 \geq 0 \end{aligned}$$

Note that in this formation, we have substituted out for the date 0 consumption; in other words, the first term in the max expression stands for $1/2(c_0)$ and we have substituted for c_0 its value obtained from the constraint: $c_0 + c_1^1 q_1 + c_2^1 q_2 = 10 + 1q_1 + 2q_2$. With this trick, the only constraints remaining are the non-negativity constraints requiring consumption to be nonnegative in all date-states.

The FOCs state that the intertemporal rate of substitution between future (in either state) and present consumption (i.e. the ratio of the relevant marginal utilities) should equal the price ratio. The latter is effectively measured by the

price of the Arrow-Debreu security, the date 0 price of consumption being the numeraire. These FOCs (assuming interior solutions) are

$$\text{Agent 1 : } \begin{cases} c_1^1 : \frac{q_1}{2} = 0.9 \left(\frac{1}{3}\right) \frac{1}{c_1^1} \\ c_2^1 : \frac{q_2}{2} = 0.9 \left(\frac{2}{3}\right) \frac{1}{c_2^1} \end{cases} \quad \text{Agent 2 : } \begin{cases} c_1^2 : \frac{q_1}{2} = 0.9 \left(\frac{1}{3}\right) \frac{1}{c_1^2} \\ c_2^2 : \frac{q_2}{2} = 0.9 \left(\frac{2}{3}\right) \frac{1}{c_2^2} \end{cases}$$

while the market clearing conditions read: $c_1^1 + c_1^2 = 5$ and $c_2^1 + c_2^2 = 8$. Each of the FOCs is of the form $\frac{q_\theta}{1} = \frac{(.9)(\pi_\theta) \frac{1}{c_\theta^k}}{1/2}$, $k, \theta = 1, 2$, or

$$q_\theta = \frac{\delta \pi_\theta \frac{\partial U^k}{\partial c_\theta^k}}{\frac{\partial U_0^k}{\partial c_0^k}}, \quad k, \theta = 1, 2. \quad (8.1)$$

Together with the market clearing conditions, Equation (8.1) reveals the determinants of the equilibrium Arrow-Debreu security prices. It is of the form:

$$\frac{\text{price of the good if state } \theta \text{ is realized}}{\text{price of the good today}} = \frac{MU_\theta^k}{MU_0^k},$$

in other words, the ratio of the price of the Arrow-Debreu security to the price of the date 0 consumption good must equal (at an interior solution) the ratio of the marginal utility of consumption tomorrow if state θ is realized to the marginal utility of today's consumption (the latter being constant at $1/2$). This is the *marginal rate of substitution* between the contingent consumption in state θ and today's consumption. From this system of equations, one clearly obtains $c_1^1 = c_1^2 = 2.5$ and $c_2^1 = c_2^2 = 4$ from which one, in turn, derives:

$$\begin{aligned} q_1 &= \frac{1}{2} (0.9) \left(\frac{1}{3}\right) \left(\frac{1}{c_1^1}\right) = 2 (0.9) \left(\frac{1}{3}\right) \left(\frac{1}{2.5}\right) = (0.9) \left(\frac{1}{3}\right) \left(\frac{4}{5}\right) = 0.24 \\ q_2 &= \frac{1}{2} (0.9) \left(\frac{2}{3}\right) \left(\frac{1}{c_2^1}\right) = 2 (0.9) \left(\frac{2}{3}\right) \left(\frac{1}{4}\right) = (0.9) \left(\frac{2}{3}\right) \left(\frac{4}{8}\right) = 0.3 \end{aligned}$$

Notice how the Arrow-Debreu state-contingent prices reflect probabilities, on the one hand, and marginal rates of substitution (taking the time discount factor into account and computed at consumption levels compatible with market clearing) and thus relative scarcities, on the other. The prices computed above differ in that they take account of the different state probabilities ($1/3$ for state 1, $2/3$ for state 2) and because the marginal utilities differ as a result of the differing total quantities of the consumption good available in state 1 (5 units) and in state 2 (8 units). In our particular formulation, the total amount of goods available at date 0 is made irrelevant by the fact that date 0 marginal utility is constant. Note that if the date 1 marginal utilities were constant, as would be the case with linear (risk neutral) utility functions, the goods endowments would not influence the Arrow-Debreu prices, which would then be exactly proportional to the state probabilities.

Box 8.1:

Interior vs. Corner Solutions

We have described the *interior* solution to the maximization problem. By that restriction we generally mean the following: The problem under maximization is constrained by the condition that consumption at all dates should be nonnegative. There is no interpretation given to a negative level of consumption, and, generally, even a zero consumption level is precluded. Indeed, when we make the assumption of a log utility function, the marginal utility at zero is infinity, meaning that by construction the agent will do all that is in his power to avoid that situation. Effectively an equation such as Equation (8.1) will never be satisfied for finite and nonzero prices with log utility and period one consumption level equal to zero; that is, it will never be optimal to select a zero consumption level. Such is not the case with the linear utility function assumed to prevail at date 0. Here it is conceivable that, no matter what, the marginal utility in either state at date 1 [the numerator in the RHS of Equation (8.1)] be larger than 1/2 times the Arrow-Debreu price [the denominator of the RHS in Equation (8.1) multiplied by the state price]. Intuitively, this would be a situation where the agent derives more utility from the good tomorrow than from consuming today, even when his consumption level today is zero. Fundamentally, the interior optimum is one where he would like to consume less than zero today to increase even further consumption tomorrow, something which is impossible. Thus the only solution is at a corner, that is at the boundary of the feasible set, with $c_0^k = 0$ and the condition in Equation (8.1) taking the form of an inequality.

In the present case we can argue that corner solutions cannot occur with regard to future consumption (because of the log utility assumption). The full and complete description of the FOCs for problem (P) spelled out in Section 8.2 is then

$$q_\theta \frac{\partial U_0^k}{\partial c_0^k} \leq \delta \pi_\theta \frac{\partial U^k}{\partial c_\theta^k}, \text{ if } c_0^k > 0, \text{ and } k, \theta = 1, 2. \quad (8.2)$$

In line with our goal of being as transparent as possible, we will often, in the sequel, satisfy ourselves with a description of interior solutions to optimizing problems, taking care to ascertain, ex post, that the solutions do indeed occur at the interior of the choice set. This can be done in the present case by verifying that the optimal c_0^k is strictly positive for both agents at the interior solutions, so that Equation (8.1) must indeed apply. \square

The date 0 consumptions, at those equilibrium prices, are given by

$$\begin{aligned} c_0^1 &= 10 + 1(.24) + 2(.3) - 2.5(.24) - 4(.3) = 9.04 \\ c_0^2 &= 5 + 4(.24) + 6(.3) - 2.5(.24) - 4(.3) = 5.96 \end{aligned}$$

The post-trade equilibrium consumptions are found in Table 8.2.

This allocation is the best each agent can achieve at the given prices $q_1 = .24$ and $q_2 = .3$. Furthermore, at those prices, supply equals demand in each market, in every state and time period. These are the characteristics of a (general) competitive equilibrium.

Table 8.2: Post-Trade Equilibrium Consumptions

	$t = 0$	$t = 1$	
		θ_1	θ_2
Agent 1	9.04	2.5	4
Agent 2	5.96	2.5	4
Total	15.00	5.0	8

In light of this example, it is interesting to return to some of the concepts discussed in our introductory chapter. In particular, let us confirm the (Pareto) optimality of the allocation emerging from the competitive equilibrium. Indeed, we have assumed as many markets as there are states of nature, so assumption H1 is satisfied. We have *de facto* assumed competitive behavior on the part of our two consumers (they have taken prices as given when solving their optimization problems), so H2 is satisfied (of course, in reality such behavior would not be privately optimal if indeed there were only two agents. Our example would not have changed materially had we assumed a large number of agents, but the notation would have become much more cumbersome).

In order to guarantee the existence of an equilibrium, we need hypotheses H3 and H4 as well. H3 is satisfied in a weak form (no curvature in date 0 utility). Finally, ours is an exchange economy where H4 does not apply (or, if one prefers, it is trivially satisfied). Once the equilibrium is known to exist, as is the case here, H1 and H2 are sufficient to guarantee the optimality of the resulting allocation of resources. Thus, we expect to find that the above competitive allocation is Pareto optimal (PO), that is, it is impossible to rearrange the allocation of consumptions so that the utility of one agent is higher without diminishing the utility of the other agent.

One way to verify the optimality of the competitive allocation is to establish the precise conditions that must be satisfied for an allocation to be Pareto optimal in the exchange economy context of our example. It is intuitively clear that the above Pareto superior reallocations will be impossible if the initial allocation maximizes the weighted sum of the two agents' utilities. That is, an allocation is optimal in our example if, for some weight λ it solves the following maximization problem.³

$$\begin{aligned}
 & \max_{\{c_0^1, c_1^1, c_2^1\}} u^1(c_0^1, c_1^1, c_2^1) + \lambda u^2(c_0^2, c_1^2, c_2^2) \\
 & \text{s.t.} \\
 & c_0^1 + c_0^2 = 15; c_1^1 + c_1^2 = 5; c_2^1 + c_2^2 = 8, \\
 & c_0^1, c_1^1, c_2^1, c_0^2, c_1^2, c_2^2 \geq 0
 \end{aligned}$$

This problem can be interpreted as the problem of a benevolent central planner constrained by an economy's total endowment (15, 5, 8) and weighting the two agents utilities according to a parameter λ , possibly equal to 1. The decision

³It is just as easy here to work with the most general utility representation.

variables at his disposal are the consumption levels of the two agents in the two dates and the two states. With u_i^k denoting the derivative of agent k 's utility function with respect to c_i^k ($i = 1, 2, 3$), the FOCs for an interior solution to the above problem are found in Equation (8.3).

$$\frac{u_0^1}{u_0^2} = \frac{u_1^1}{u_1^2} = \frac{u_2^1}{u_2^2} = \lambda \quad (8.3)$$

This condition states that, in a Pareto optimal allocation, the ratio of the two agents' marginal utilities with respect to the three goods (i.e., the consumption good at date 0, the consumption good at date 1 if state 1, and the consumption good at date 1 if state 2) should be identical.⁴ In an exchange economy this condition, properly extended to take account of the possibility of corner solution, together with the condition that the agents' consumption adds up to the endowment in each date-state, is necessary and sufficient.

It remains to check that Equation (8.3) is satisfied at the equilibrium allocation. We can rewrite Equation (8.3) for the parameters of our example:

$$\frac{1/2}{1/2} = \frac{(0.9)^{\frac{1}{3}} \frac{1}{c_1^1}}{(0.9)^{\frac{1}{3}} \frac{1}{c_1^2}} = \frac{(0.9)^{\frac{2}{3}} \frac{1}{c_2^1}}{(0.9)^{\frac{2}{3}} \frac{1}{c_2^2}}$$

It is clear that the condition in Equation (8.3) is satisfied since $c_1^1 = c_1^2$; $c_2^1 = c_2^2$ at the competitive equilibrium, which thus corresponds to the Pareto optimum with equal weighting of the two agents' utilities: $\lambda = 1$ and all three ratios of marginal utilities are equal to 1. Note that other Pareto optima are feasible, for example one where $\lambda = 2$. In that case, however, only the latter two equalities can be satisfied: the date 0 marginal utilities are constant which implies that no matter how agent consumptions are redistributed by the market or by the central planner, the first ratio of marginal utilities in Equation (8.3) cannot be made equal to 2. This is an example of a corner solution to the maximization problem leading to equation (8.3).

In this example, agents are able to purchase consumption in any date-state of nature. This is the case because there are enough Arrow-Debreu securities; specifically there is an Arrow-Debreu security corresponding to each state of nature. If this were not the case, the attainable utility levels would decrease: at least one agent, possibly both of them, would be worse off. If we assume that only the state 1 Arrow-Debreu security is available, then there are no ways to make the state 2 consumption of the agents differ from their endowments. It is easy to check that this constraint does not modify their demand for the state 1 contingent claim, nor its price. The post-trade allocation, in that situation, is found in Table 8.3.

The resulting post-trade utilities are

$$\begin{aligned} \text{Agent 1} & : 1/2(9.64) + .9(1/3 \ln(2.5) + 2/3 \ln(2)) = 5.51 \\ \text{Agent 2} & : 1/2(5.36) + .9(1/3 \ln(2.5) + 2/3 \ln(6)) = 4.03 \end{aligned}$$

⁴Check that Equation (8.3) implies that the MRS between any two pair of goods is the same for the 2 agents and refer to the definition of the contract curve (the set of PO allocations) in the appendix to Chapter 1.

Table 8.3 : The Post-Trade Allocation

	$t = 0$	$t = 1$	
		θ_1	θ_2
Agent 1	9.64	2.5	2
Agent 2	5.36	2.5	6
Total	15.00	5.0	8

In the case with two state-contingent claim markets, the post-trade utilities are both higher (illustrating a reallocation of resources that is said to be *Pareto superior* to the no-trade allocation):

$$\text{Agent 1} : 1/2(9.04) + .9(1/3 \ln(2.5) + 2/3 \ln(4)) = 5.62$$

$$\text{Agent 2} : 1/2(5.96) + .9(1/3 \ln(2.5) + 2/3 \ln(4)) = 4.09.$$

When there is an Arrow-Debreu security corresponding to each state of nature, one says that the securities markets are complete.

8.4 Pareto Optimality and Risk Sharing

In this section and the next we further explore the nexus between a competitive equilibrium in an Arrow-Debreu economy and Pareto optimality. We first discuss the risk-sharing properties of a Pareto optimal allocation. We remain in the general framework of the example of the previous two sections but start with a different set of parameters. In particular, let the endowment matrix for the two agents be as shown in Table 8.4.

Table 8.4: The New Endowment Matrix

	$t = 0$	$t = 1$	
		θ_1	θ_2
Agent 1	4	1	5
Agent 2	4	5	1

Assume further that each state is now equally likely with probability 1/2. As before, consumption in period 0 cannot be stored and carried over into period 1. In the absence of trade, agents clearly experience widely differing consumption and utility levels in period 1, depending on what state occurs (see Table 8.5).

How could agents' utilities be improved? By concavity (risk aversion), this must be accomplished by reducing the spread of the date 1 income possibilities, in other words, lowering the risk associated with date 1 income. Because of symmetry, all date 1 income fluctuations can, in fact, be eliminated if agent 2 agrees to transfer 2 units of the good in state 1 against the promise to receive 2 units from agent 1 if state 2 is realized (see Table 8.6).

Table 8.5: Agents' Utility In The Absence of Trade

	State-Contingent Utility		Expected Utility in Period 1
	θ_1	θ_2	
Agent 1	$\ln(1) = 0$	$\ln(5) = 1.609$	$1/2 \ln(1) + 1/2 \ln(5) = .8047$
Agent 2	$\ln(5) = 1.609$	$\ln(1) = 0$	$1/2 \ln(1) + 1/2 \ln(5) = .8047$

Table 8.6: The Desirable Trades And Post-Trade Consumptions

Date 1	Endowments Pre-Trade		Consumption Post-Trade	
	θ_1	θ_2	θ_1	θ_2
Agent 1	1	5 \Downarrow 2	3	3
Agent 2	5 \Uparrow 2	1	3	3

Now we can compare expected second period utility levels before and after trade for both agents:

Before	After
.8047	$1/2 \ln(3) + 1/2 \ln(3) = 1.099 \cong 1.1$

in other words, expected utility has increased quite significantly as anticipated.⁵

This feasible allocation is, in fact, *Pareto optimal*. In conformity with Equation (8.3), the ratios of the two agents' marginal utilities are indeed equalized across states. More is accomplished in this perfectly symmetrical and equitable allocation: Consumption levels and MU are equated across agents and states, but this is a coincidence resulting from the symmetry of the initial endowments.

Suppose the initial allocation was that illustrated in Table 8.7.

Once again there is no aggregate risk: The total date 1 endowment is the same in the two states, but one agent is now richer than the other. Now consider the plausible trade outlined in Table 8.8.

Check that the new post-trade allocation is also Pareto optimal: Although consumption levels and marginal utilities are not identical, the ratio of marginal utilities is the same across states (except at date 0 where, as before, we have a corner solution since the marginal utilities are given constants). Note that this PO allocation features perfect risk sharing as well. By that we mean that the two agents have constant date 1 consumption (2 units for agent 1, 4 units for agent 2) independent of the realized state. This is a general characteristic of PO allocations in the absence of aggregate risk (and with risk-averse agents).

⁵With the selected utility function, it has increased by 37%. Such quantification is not, however, compatible with the observation that expected utility functions are defined only up to a linear transformation. Instead of using $\ln c$ for the period utility function, we could equally well have used $(b + \ln c)$ to represent the same preference ordering. The quantification of the increased in utility pre- and post-trade would be affected.

Table 8.7 : Another Set of Initial Allocations

	$t = 0$	$t = 1$	
		θ_1	θ_2
Agent 1	4	1	3
Agent 2	4	5	3

Table 8.8 : Plausible Trades And Post-Trade Consumptions

Date 1	Endowments Pre-trade		Consumption Post-trade	
	θ_1	θ_2	θ_1	θ_2
Agent 1	1	3 $[\Downarrow 1]$	2	2
Agent 2	5 $[\Uparrow 1]$	3	4	4

If there is no aggregate risk, all PO allocations necessarily feature full mutual insurance.

This statement can be demonstrated, using the data of our problem. Equation (8.3) states that the ratio of the two agents' marginal utilities should be equated across states. This also implies, however, that the Marginal Rate of Substitution (MRS) between state 1 and state 2 consumption must be the same for the two agents. In the case of log period utility:

$$\frac{1/c_1^1}{1/c_1^2} = \frac{1/c_2^1}{1/c_2^2} = \frac{1/c_1^1}{1/c_2^1} = \frac{1/c_1^2}{1/c_2^2}$$

The latter equality has the following implications:

1. If one of the two agents is fully insured — no variation in his date 1 consumption (i.e., $MRS = 1$) — the other must be as well.
2. More generally, if the MRS are to differ from 1, given that they must be equal between them, the low consumption-high MU state must be the same for both agents and similarly for the high consumption-low MU state. But this is impossible if there is no aggregate risk and total endowment is constant. Thus, as asserted, in the absence of aggregate risk, a PO allocation features perfectly insured individuals and MRS identically equal to 1.
3. If there is aggregate risk, however, the above reasoning also implies that, at a Pareto optimum, it is shared “proportionately.” This is literally true if agents preferences are homogeneous. Refer to the competitive equilibrium of Section 8.3 for an example.
4. Finally, if agents are differentially risk averse, in a Pareto optimal allocation the less risk averse will typically provide some insurance services to the more risk averse. This is most easily illustrated by assuming that one of the two agents, say agent 1, is risk neutral. By risk neutrality, agent one's marginal utility is constant. But then the marginal utility of agent 2 should also

be constant across states. For this to be the case, however, agent two's income uncertainty must be fully absorbed by agent 1, the risk-neutral agent.

5. More generally, optimal risk sharing dictates that the agent most tolerant of risk bears a disproportionate share of it.

8.5 Implementing Pareto Optimal Allocations: On the Possibility of Market Failure

Although to achieve the desired allocations, the agents of our previous section could just effect a handshake trade, real economic agents typically interact only through impersonal security markets or through deals involving financial intermediaries. One reason is that, in an organized security market, the contracts implied by the purchase or sale of a security are enforceable. This is important: Without an enforceable contract, if state 1 occurs, agent 2 might retreat from his ex-ante commitment and refuse to give up the promised consumption to agent 1, and vice versa if state 2 occurs. Accordingly, we now address the following question: What securities could empower these agents to achieve the optimal allocation for themselves?

Consider the Arrow-Debreu security with payoff in state 1 and call it security Q to clarify the notation below. Denote its price by q_Q , and let us compute the demand by each agent for this security denoted z_Q^i , $i = 1, 2$. The price is expressed in terms of period 0 consumption. We otherwise maintain the setup of the preceding section. Thus,

$$\begin{aligned} \text{Agent 1 solves: } & \max (4 - q_Q z_Q^1) + [1/2 \ln(1 + z_Q^1) + 1/2 \ln(5)] \\ & \text{s.t. } q_Q z_Q^1 \leq 4 \\ \text{Agent 2 solves: } & \max (4 - q_Q z_Q^2) + [1/2 \ln(5 + z_Q^2) + 1/2 \ln(1)] \\ & \text{s.t. } q_Q z_Q^2 \leq 4 \end{aligned}$$

Assuming an interior solution, the FOCs are, respectively,

$$-q_Q + \frac{1}{2} \left(\frac{1}{1+z_Q^1} \right) = 0; \quad -q_Q + \frac{1}{2} \left(\frac{1}{5+z_Q^2} \right) = 0 \Rightarrow \frac{1}{1+z_Q^1} = \frac{1}{5+z_Q^2};$$

also $z_Q^1 + z_Q^2 = 0$ in equilibrium, hence, $z_Q^1 = 2; z_Q^2 = -2$; these represent the holdings of each agent and $q_Q = (1/2)(1/3) = 1/6$. In effect, agent 1 gives up $q_Q z_Q^1 = (1/6)(2) = 1/3$ unit of consumption at date 0 to agent 2 in exchange for 2 units of consumption at date 1 if state 1 occurs. Both agents are better off as revealed by the computation of their expected utilities post-trade:

$$\begin{aligned} \text{Agent 1 expected utility} & : 4 - 1/3 + 1/2 \ln 3 + 1/2 \ln 5 = 5.013 \\ \text{Agent 2 expected utility} & : 4 + 1/3 + 1/2 \ln 3 + 1/2 \ln 1 = 4.879, \end{aligned}$$

though agent 2 only slightly so. Clearly agent 1 is made proportionately better off because security Q pays off in the state where his MU is highest. We may view

agent 2 as the issuer of this security as it entails, for him, a future obligation.⁶

Let us denote R the other conceivable Arrow-Debreu security, one paying in state 2. By symmetry, it would also have a price of $1/6$, and the demand at this price would be $z_R^1 = -2, z_R^2 = +2$, respectively. Agent 2 would give up $1/3$ unit of period 1 consumption to agent 1 in exchange for 2 units of consumption in state 2.

Thus, if both security Q and R are traded, the market allocation will replicate the optimal allocation of risks, as seen in Table 8.9.

Table 8.9: Market Allocation When Both Securities Are Traded

	$t = 0$	$t = 1$	
		θ_1	θ_2
Agent 1	4	3	3
Agent 2	4	3	3

In general, it will be possible to achieve the optimal allocation of risks provided the number of linearly independent securities equals the number of states of nature. By linearly independent we mean, again, that there is no security whose payoff pattern across states and time periods can be duplicated by a portfolio of other securities. This important topic will be discussed at length in Chapter 10. Here let us simply take stock of the fact that our securities Q, R are the simplest pair of securities with this property.

Although a complete set of Arrow-Debreu securities is sufficient for optimal risk sharing, it is not necessary in the sense that it is possible, by coincidence, for the desirable trades to be effected with a simplified asset structure. For our simple example, one security would allow the agents to achieve that goal because of the essential symmetry of the problem. Consider security Z with payoffs:

	θ_1	θ_2
Z	2	-2

Clearly, if agent 1 purchases 1 unit of this security ($z_Z^1 = 1$) and agent 2 sells one unit of this security ($z_Z^2 = -1$), optimal risk sharing is achieved. (At what price would this security sell?)

So far we have implicitly assumed that the creation of these securities is costless. In reality, the creation of a new security is an expensive proposition: Disclosure documents, promotional materials, etc., must be created, and the agents most likely to be interested in the security contacted. In this example, issuance will occur only if the cost of issuing Q and R does not exceed the (ex-

⁶In a noncompetitive situation, it is likely that agent 2 could extract a larger portion of the rent. Remember, however, that we maintain, throughout, the assumption of price-taking behavior for our two agents who are representatives of larger classes of similar individuals.

pected) utility gained from purchasing them. In this margin lies the investment banker's fee.

In the previous discussion we imagined each agent as issuing securities to the other simultaneously. More realistically, perhaps, we could think of the securities Q and R as being issued in sequence, one after the other (but both before period 1 uncertainty is resolved). Is there an advantage or disadvantage of going first, that is, of issuing the *first* security? Alternatively, we might be preoccupied with the fact that, although both agents benefit from the issuance of new securities, only the individual issuer pays the cost of establishing a new market. In this perspective it is interesting to measure the net gains to trade for each agent. These quantities are summarized in Table 8.10.

Table 8.10: The Net Gains From Trade: Expected Utility Levels and Net Trading Gains (Gain to issuer in bold)

	No Trade	Trade Only Q		Trade Both Q and R	
	EU	EU	$\Delta EU^{(i)}$	EU	$\Delta EU^{(ii)}$
Agent 1	4.8047	5.0206	0.2159	5.0986	0.0726
Agent 2	4.8047	4.883	0.0783	5.0986	0.2156
Total			0.2942		0.2882

(i) Difference in EU when trading Q only, relative to no trade.

(ii) Difference in EU when trading both Q and R , relative to trading Q only.

This computation tells us that, in our example, the issuer of the security gains less than the other party in the future trade. If agent 2 goes first and issues security Q , his net expected utility gain is 0.0783, which also represents the most he would be willing to pay his investment bank in terms of period 0 consumption to manage the sale for him. By analogy, the marginal benefit to agent 1 of *then* issuing security R is 0.0726. The reverse assignments would have occurred if agent 1 had gone first, due to symmetry in the agent endowments. That these quantities represent the upper bounds on possible fees comes from the fact that period 0 utility of consumption is the level of consumption itself.

The impact of all this is that each investment bank will, out of desire to maximize its fee potential, advise its client to issue his security second. No one will want to go first. Alternatively, if the effective cost of setting up the market for security Q is anywhere between 0.0783 and 0.288, there is a possibility of *market failure*, unless agent 2 finds a way to have agent 1 share in the cost of establishing the market. We speak of market failure because the social benefit of setting up the market would be positive 0.288 minus the cost itself — while the market might not go ahead if the private cost to agent 2 exceeds his own private benefit, measured at 0.0783 units of date 0 consumption. Of course, it might also be the case that the cost exceeds the total benefit. This is another reason for the market not to exist and, in general, for markets to be incomplete. But in this case, one would not talk of market failure. Whether the privately motivated decisions of individual agents lead to the socially optimal outcome —

in this case the socially optimal set of securities – is a fundamental question in financial economics.

There is no guarantee that private incentives will suffice to create the social optimal set of markets. We have identified a problem of sequencing – (the issuer of a security may not be the greatest beneficiary of the creation of the market) and as a result there may be a waiting game with suboptimal results. There is also a problem linked with the sharing of the cost of setting up a market. The benefits of a new market often are widely spread among a large number of potential participants and it may be difficult to find an appropriate mechanism to have them share the initial setup cost, for example, because of free rider or coordination problems. Note that in both these cases, as well as in the situation where the cost of establishing a market exceeds the total benefit for individual agents, we anticipate that technical innovations leading to decreases in the cost of establishing markets will help alleviate the problem and foster a convergence toward a more complete set of markets.

8.6 Conclusions

The asset pricing theory presented in this chapter is in some sense the father of all asset pricing relationships. It is fully general and constitutes an extremely valuable reference. Conceptually its usefulness is unmatched and this justifies us investing more in its associated apparatus. At the same time, it is one of the most abstract theories and its usefulness in practice is impaired by the difficulty in identifying individual states of nature and by the fact that, even when a state (or a set of states) can be identified, its (their) realization cannot always be verified. As a result it is difficult to write the appropriate conditional contracts. These problems go a long way in explaining why we do not see Arrow-Debreu securities being traded, a fact that does not strengthen the immediate applicability of the theory. In addition, as already mentioned, the static setting of the Arrow-Debreu theory is unrealistic for most applications. For all these reasons we cannot stop here and we will explore a set of alternative, sometimes closely related, avenues for pricing assets in the following chapters.

References

- Arrow, K. (1952), “An Extension of the Basic Theorems of Classical Welfare Economics,” in J. Neyman (ed.), *Proceedings of the Second Berkeley Symposium on Mathematical Statistics and Probability*, University of California Press, Berkeley, 507-532.
- Debreu, G. (1952), *Theory of Value*, John Wiley & Sons, New York.

Chapter 9 : The Consumption Capital Asset Pricing Model (CCAPM)

9.1 Introduction

So far, our asset pricing models have been either one-period models such as the CAPM or multi-period but static, such as the Arrow-Debreu model. In the latter case, even if a large number of future periods is assumed, all decisions, including security trades, take place at date zero. It is in that sense that the Arrow-Debreu model is static. Reality is different, however. Assets are traded every period, as new information becomes available, and decisions are made sequentially, one period at a time, all the while keeping in mind the fact that today's decisions impact tomorrow's opportunities. Our objective in this chapter is to capture these dynamic features and to price assets in such an environment.

Besides adding an important dimension of realism, another advantage of a dynamic setup is to make it possible to draw the link between the financial markets and the real side of the economy. Again, strictly speaking, this can be accomplished within an Arrow-Debreu economy. The main issues, however, require a richer dynamic context where real production decisions are not made once and for all at the beginning of time, but progressively, as time evolves. Building a model in which we can completely understand how real events impact the financial side of the economy in the spirit of fundamental financial analysis is beyond the scope of the present chapter. The model discussed here, however, opens up interesting possibilities in this regard, which current research is attempting to exploit. We will point out possible directions as we go along.

9.2 The Representative Agent Hypothesis and Its Notion of Equilibrium

9.2.1 An Infinitely-Lived Representative Agent

To accomplish these goals in a model of complete generality (in other words, with many different agents and firms) and in a way that asset prices can be tractably computed is beyond the present capability of economic science. As an alternative, we will make life simpler by postulating many *identical* infinitely lived consumers. This allows us to examine the decisions of a *representative*, stand-in consumer and explore their implications for asset pricing. In particular, we will assume that agents act to maximize the expected present value of discounted utility of consumption over their entire, infinite, lifetimes:

$$\max E \left(\sum_{t=0}^{\infty} \delta^t U(\tilde{c}_t) \right),$$

where δ is the discount factor and $U(\cdot)$ the period utility function with $U_1(\cdot) > 0, U_2(\cdot) < 0$. This construct is the natural generalization to the case of infinite

lifetimes of the preferences considered in our earlier two-period example. Its use can be justified by the following considerations.

First, if we model the economy as ending at some terminal date T (as opposed to assuming an infinite horizon), then the agent's investment behavior will reflect this fact. In the last period of his life, in particular, he will stop saving, liquidate his portfolio, and consume its entire value. There is no real-world counterpart for this action as the real economy continues forever. Assuming an infinite horizon eliminates these terminal date complications. Second, it can be shown, under fairly general conditions, that an infinitely lived agent setup is formally equivalent to one in which agents live only a finite number of periods themselves, provided they derive utility from the well-being of their descendants (a bequest motive). This argument is detailed in Barro (1974).

Restrictive as it may seem, the identical agents assumption can be justified by the fact that, in a competitive equilibrium with complete securities markets there is an especially intuitive sense of a representative agent: one whose utility function is a weighted average of the utilities of the various agents in the economy. In Box 9.1 we detail the precise way in which one can construct such a representative individual and we discuss some of the issues at stake.

Box 9.1 Constructing a Representative Agent

In order to illustrate the issue, let us return to the two-period ($t = 0, 1$) Arrow-Debreu economy considered earlier. In that economy, each agent k , $k = 1, 2, \dots, K$, solves:

$$\begin{aligned} \max \quad & U^k(c_0^k) + \delta^k \sum_{\theta=1}^N \pi_{\theta} U^k(c_{\theta}^k) \\ \text{s.t.} \quad & c_0^k + \sum_{\theta=1}^N q_{\theta} c_{\theta}^k \leq e_0^k + \sum_{\theta=1}^N q_{\theta} e_{\theta}^k \end{aligned}$$

where the price of period 0 endowment is normalized to 1, and the endowments of a typical agent k are described by the vector $\begin{pmatrix} e_0^k \\ e_1^k \\ \vdots \\ e_N^k \end{pmatrix}$.

In equilibrium, not only are the allocations optimal, but at the prevailing prices, supply equals demand in every market:

$$\begin{aligned} \sum_{k=1}^K c_0^k &= \sum_{k=1}^K e_0^k, \text{ and} \\ \sum_{k=1}^K c_{\theta}^k &= \sum_{k=1}^K e_{\theta}^k, \text{ for every state } \theta. \end{aligned}$$

We know this competitive equilibrium allocation is Pareto optimal: No one can be better off without making someone else worse off. One important implication of this property for our problem is that there exists some set of weights

$(\lambda_1, \dots, \lambda_K)$, which in general will depend on the initial endowments, such that the solution to the following problem gives an allocation that is identical to the equilibrium allocation:

$$\begin{aligned} & \max \sum_{k=1}^K \lambda_k \left\{ U^k(c_0^k) + \delta^k \sum_{\theta=1}^N \pi_\theta U^k(c_\theta^k) \right\}, \\ \text{s.t. } & \sum_{k=1}^K c_0^k = \sum_{k=1}^K e_0^k, \\ & \sum_{k=1}^K c_\theta^k = \sum_{k=1}^K e_\theta^k, \quad \forall \theta \\ & \sum_{k=1}^K \lambda_k = 1; \lambda_k > 0, \quad \forall k. \end{aligned}$$

This maximization is meant to represent the problem of a benevolent central planner attempting to allocate the aggregate resources of the economy so as to maximize the weighted sum of the utilities of the individual agents. We stated a similar problem in Chapter 8 in order to identify the conditions characterizing a Pareto optimal allocation of resources. Here we see this problem as suggestive of the form the representative agent's preference ordering, defined over aggregate consumption, can take (the representative agent is denoted by the superscript A):

$$\begin{aligned} U^A(c_0^A, c_\theta^A) &= U_0^A(c_0^A) + \sum_{\theta=1}^N \pi_\theta U^A(c_\theta^A), \text{ where} \\ U_0^A(c_0^A) &= \sum_{k=1}^K \lambda_k U_0^k(c_0^k) \text{ with } \sum_{k=1}^K c_0^k = \sum_{k=1}^K e_0^k \equiv c_0^A \\ U^A(c_\theta^A) &= \sum_{k=1}^K \delta^k \lambda_k U^k(c_\theta^k) \text{ with } \sum_{k=1}^K c_\theta^k = \sum_{k=1}^K e_\theta^k \equiv c_\theta^A, \text{ for each state } \theta. \end{aligned}$$

In this case, the aggregate utility function directly takes into account the distribution of consumption across agents. This setup generalizes to as many periods as we like and, with certain modifications, to an infinite horizon. It is an intuitive sense of a representative agent as one who constitutes a weighted average of all the economy's participants.

A conceptual problem with this discussion resides in the fact that, in general, the weights $\{\lambda_1, \lambda_2, \dots, \lambda_K\}$ will depend on the initial endowments: Loosely speaking, the agent with more wealth gets a bigger λ weight. It can be shown, however, that the utility function, constructed as previously shown, will, in addition, be independent of the initial endowment distribution if two further conditions are satisfied:

1. the discount factor of every agent is the same (i.e., all agents δ 's are the same).
2. agents' period preferences are of either of the two following forms:

$$U^k(c) = \frac{\gamma}{\gamma - 1} (\alpha^k + \gamma c)^{1 - \frac{1}{\gamma}} \quad \text{or} \quad U^k(c) = -e^{-\alpha^k c}.$$

If either of these two conditions are satisfied, that is, by and large, if the agents' preferences can be represented by either a CRRA or a CARA utility function, then there exists a representative agent economy for which the equilibrium Arrow-Debreu prices are the same as they are for the K agent economy, and for which

$$U^A(c) = g(c)H(\lambda_1, \dots, \lambda_K), \text{ where } g_1(c) > 0, g_{11}(c) < 0.$$

In this case, the weights do not affect preferences because they appear in the form of a multiplicative scalar.

Let us repeat that, even if the individual agent preferences do not take either of the two forms previously listed, there will still be a representative agent whose preferences are the weighted average of the individual agent preferences. Unlike the prior case, however, this ordering will then depend on the initial endowments [see Constantinides (1982)]. \square

9.2.2 On the Concept of a “No-Trade” Equilibrium

In a representative agent economy we must, of necessity, use a somewhat specialized notion of equilibrium — a *no-trade equilibrium*. If, indeed, for a particular model specification, some security is in positive net supply, the equilibrium price will be the price at which the representative agent is willing to hold that amount — the total supply — of the security. In other specifications we will price securities that do not appear explicitly — securities that are said to be in zero net supply. The prototype of the latter is an IOU type of contract: In a one agent economy, the total net supply of IOUs must, of course, be zero. In this case, if at some price the representative agent wants to supply (sell) the security, since there is no one to demand it, supply exceeds demand. Conversely, if at some price the representative agent wants to buy the security (and thus no one wants to supply it), demand exceeds supply. Financial markets are thus in equilibrium, if and only if, at the prevailing price, supply equals demand *and* both are simultaneously zero. In all cases, the equilibrium price is that price at which the representative agent wishes to hold exactly the amount of the security present in the economy. Therefore, the essential question being asked is: What prices must securities assume so that the amount the representative agent *must* hold (for all markets to clear) exactly equals what he *wants* to hold. At these prices, further trade is not utility enhancing. In a more conventional multi-agent economy, an identical state of affairs is verified post-trade. The representative agent class of models is not appropriate, of course, for the analysis of some issues in finance; for example, issues linked with the volume of trade cannot be studied since, in a representative agent model, trading volume is, by construction, equal to zero.

9.3 An Exchange (Endowment) Economy

9.3.1 The Model

This economy will be directly analogous to the Arrow-Debreu exchange economies considered earlier: production decisions are in the background and abstracted away. It is, however, an economy that admits recursive trading, resulting from investment decisions made over time, period after period (as opposed to being made once and for all at date 0).

There is one, perfectly divisible *share* which we can think of as representing the market portfolio of the CAPM (later we shall relax this assumption). Ownership of this share entitles the owner to all the economy's output (in this economy, all firms are publicly traded). Output is viewed as arising exogenously, and as being stochastically variable through time, although in a stationary fashion. This is the promised, although still remote, link with the real side of the economy. And we will indeed use macroeconomic data to calibrate the model in the forthcoming sections. At this point, we can think of the output process as being governed by a large-number-of-states version of the three-state probability transition matrix found in Table 9.1.

Table 9.1: Three-State Probability Transition Matrix

		Output in Period $t + 1$			
		Y^1	Y^2	Y^3	
Output in Period t	Y^1	$\begin{bmatrix} \pi_{11} & \pi_{12} & \pi_{13} \\ \pi_{21} & \pi_{22} & \pi_{23} \\ \pi_{31} & \pi_{32} & \pi_{33} \end{bmatrix}$			$= \mathbb{T}$
	Y^2				
	Y^3				

where $\pi_{ij} = \text{Prob}(Y_{t+1} = Y^j, | Y_t = Y^i)$ for any t .

That is, there are a given number of output states, levels of output that can be achieved at any given date, and the probabilities of transiting from one output state to another are constant and represented by entries in the matrix \mathbb{T} . The stationarity hypothesis embedded in this formulation may, at first sight, appear extraordinarily restrictive. The output levels defining the states may, however, be normalized variables, for instance to allow for a constant rate of growth. Alternatively, the states could themselves be defined in terms of growth rates of output rather than output levels. See Appendix 9.1 for an application.

If we adopt a continuous-state version of this perspective, the output process can be similarly described by a probability transition *function*

$$G(Y_{t+1} | Y_t) = \text{Prob}(Y_{t+1} \leq Y^j, | Y_t = Y^i).$$

We can imagine the security as representing ownership of a fruit tree where the (perishable) output (the quantity of fruit produced by the tree - the dividend)

varies from year to year. This interpretation is often referred to as the *Lucas fruit tree* economy in tribute to 1996 Nobel prize winner, R. E. Lucas Jr., who, in his 1978 article, first developed the CCAPM. The power of the approach, however, resides in the fact that any mechanism delivering a stochastic process on aggregate output, such as a full macroeconomic equilibrium model, can be grafted on the CCAPM. This opens up the way to an in-depth analysis of the rich relationships between the real and the financial sides of an economy.

This will be a *rational expectations economy*. By this expression we mean that the representative agent's expectations will be on average correct, and in particular will exhibit no systematic bias. In effect we are assuming, in line with a very large literature (and with most of what we have done implicitly so far), that the representative agent knows both the general structure of the economy and the exact output distribution as summarized by the matrix \mathbb{T} . One possible justification is that this economy has been functioning for a long enough time to allow the agent to learn the probability process governing output and to understand the environment in which he operates. Accumulating such knowledge is clearly in his own interest if he wishes to maximize his expected utility.

The agent buys and sells securities (fractions of the single, perfectly divisible share) and consumes dividends. His security purchases solve:

$$\begin{aligned} \max_{\{z_{t+1}\}} & E \left(\sum_{t=0}^{\infty} \delta^t U(\tilde{c}_t) \right) \\ \text{s.t. } & c_t + p_t z_{t+1} \leq z_t Y_t + p_t z_t \\ & z_t \leq 1, \forall t \end{aligned}$$

where p_t is the period t real price of the security in terms of consumption¹ (the price of consumption is 1) and z_t is the agent's beginning-of-period t holdings of the security. Holding a fraction z_t of the security entitles the agent to the corresponding fraction of the distributed dividend, which in an exchange economy without investment, equals total available output. The expectations operator applies across all possible values of Y feasible at each date t with the probabilities provided by the matrix \mathbb{T} .

Let us assume the representative agent's period utility function is strictly concave with $\lim_{c_t \rightarrow 0} U_1(c_t) = \infty$. Making this latter assumption insures that it is never optimal for the agent to select a zero consumption level. It thus normally insures an interior solution to the relevant maximization problem. The necessary and sufficient condition for the solution to this problem is then given by: For all t , z_{t+1} solves:

$$U_1(c_t)p_t = \delta E_t \left\{ U_1(\tilde{c}_{t+1}) \left(\tilde{p}_{t+1} + \tilde{Y}_{t+1} \right) \right\} \quad (9.1)$$

where $c_t = (p_t z_t + z_t Y_t - p_t z_{t+1})$. Note that the expectations operator applies across possible output state levels; if we make explicit the functional dependence

¹In the notation of the previous chapter: $p = q^e$.

on the output state variables, Equation (9.1) can be written (assuming Y_i is the current state):

$$U_1(c_t(Y^i))p_t(Y^i) = \delta \sum_j U_1(c_{t+1}(Y^j))(p_{t+1}(Y^j) + Y^j)\pi_{ij}$$

In Equation (9.1), $U_1(c_t)p_t$ is the utility loss in period t associated with the purchase of an additional unit of the security, while $\delta U_1(c_{t+1})$ is the marginal utility of an additional unit of consumption in period $t + 1$ and $(p_{t+1} + Y_{t+1})$ is the extra consumption (income) in period $t + 1$ from selling the additional unit of the security after collecting the dividend entitlement. The RHS is thus the expected discounted gain in utility associated with buying the extra unit of the security. The agent is in equilibrium (utility maximizing) at the prevailing price p_t if the loss in utility today, which he would incur by buying one more unit of the security ($U_1(c_t)p_t$), is exactly offset by (equals) the expected gain in utility tomorrow ($\delta E_t U_1(\tilde{c}_{t+1}) [\tilde{p}_{t+1} + \tilde{Y}_{t+1}]$), which the ownership of that additional security will provide. If this equality is not satisfied, the agent will try either to increase or to decrease his holdings of securities.²

For the entire economy to be in equilibrium, it must, therefore, be true that:

- (i) $z_t = z_{t+1} = z_{t+2} = \dots \equiv 1$, in other words, the representative agent owns the entire security;
- (ii) $c_t = Y_t$, that is, ownership of the entire security entitles the agent to all the economy's output and,
- (iii) $U_1(c_t)p_t = \delta E_t \left\{ U_1(\tilde{c}_{t+1}) (\tilde{p}_{t+1} + \tilde{Y}_{t+1}) \right\}$, or, the agents' holdings of the security are optimal given the prevailing prices. Substituting (ii) into (iii) informs us that the equilibrium price must satisfy:

$$U_1(Y_t)p_t = \delta E_t \left\{ U_1(\tilde{Y}_{t+1})(\tilde{p}_{t+1} + \tilde{Y}_{t+1}) \right\} \quad (9.2)$$

If there were many firms in this economy – say H firms, with firm h producing the (exogenous) output $\tilde{Y}_{h,t}$, then the same equation would be satisfied for each firm's stock price, $p_{h,t}$, that is,

$$p_{h,t} U_1(c_t) = \delta E_t \left\{ \left[U_1(\tilde{c}_{t+1}) (\tilde{p}_{h,t+1} + \tilde{Y}_{h,t+1}) \right] \right\} \quad (9.3)$$

where $c_t = \sum_{h=1}^H Y_{h,t}$ in equilibrium.

Equations (9.2) and (9.3) are the fundamental equations of the consumption-based capital asset pricing model.³

²In equilibrium, however, this is not possible and the price will have to adjust until the equality in Equation (9.1) is satisfied.

³The fact that the representative agent's consumption stream — via his MRS — is critical for asset pricing is true for all versions of this model, including ones with nontrivial production settings. More general versions of this model may not, however, display an identity between consumption and dividends. This will be the case, for example, if there is wage income to the agent.

A recursive substitution of Equation (9.2) into itself yields⁴

$$p_t = E_t \sum_{\tau=1}^{\infty} \delta^{\tau} \left[\frac{U_1(\tilde{Y}_{t+\tau})}{U_1(Y_t)} \tilde{Y}_{t+\tau} \right], \quad (9.4)$$

establishing the stock price as the sum of all expected discounted future dividends. Equation (9.4) resembles the standard discounting formula of elementary finance, but for the important observation that discounting takes place using the inter-temporal marginal rates of substitution defined on the consumption sequence of the representative agent. If the utility function displays risk neutrality and the marginal utility is constant ($U_{11} = 0$), Equation (9.4) reduces to

$$p_t = E_t \sum_{\tau=1}^{\infty} \delta^{\tau} [\tilde{Y}_{t+\tau}] = E_t \sum_{\tau=1}^{\infty} \left[\frac{\tilde{Y}_{t+\tau}}{(1+r_f)^{\tau}} \right], \quad (9.5)$$

which states that the stock price is the sum of expected future dividends discounted at the (constant) risk-free rate. The intuitive link between the discount factor and the risk-free rate leading to the second inequality in Equation (9.5) will be formally established in Equation (9.7). The difference between Equations (9.4) and (9.5) is the necessity, in a world of risk aversion, of discounting the flow of expected dividends at a rate higher than the risk-free rate, so as to include a risk premium. The question as to the appropriate risk premium constitutes the central issue in financial theory. Equation (9.4) proposes a definite, if not fully operational (due to the difficulty in measuring marginal rates of substitution), answer.

Box 9.2 Calculating the Equilibrium Price Function

Equation (9.2) implicitly defines the equilibrium price series. Can it be solved directly to produce the actual equilibrium prices $\{p(Y^j) : j = 1, 2, \dots, N\}$? The answer is positive. First, we must specify parameter values and functional forms. In particular, we need to select values for δ and for the various output levels Y^j , to specify the probability transition matrix \mathbb{T} and the form of the representative agent's period utility function (a CRRA function of the form $U(c) = \frac{c^{1-\gamma}}{1-\gamma}$ is a natural choice). We may then proceed as follows. Solve for the $\{p(Y^j) : j = 1, 2, \dots, N\}$ as the solution to a system of linear equations. Notice that Equation 9.2 can be written as the following system of linear equations (one for each of the N possible current states Y^j):

$$U_1(Y^1)p(Y^1) = \delta \sum_{j=1}^N \pi_{1j} U_1(Y^j) Y^j + \delta \sum_{j=1}^N \pi_{1j} U_1(Y^j) p(Y^j)$$

⁴That is, update Equation (9.2) with p_{t+1} on the left-hand side and p_{t+2} in the RHS and substitute the resulting RHS (which now contains a term in p_{t+2}) into the original Equation (9.2) ; repeat for p_{t+2} , p_{t+3} , and so on, regroup terms and extrapolate.

$$\begin{array}{ccc} \vdots & & \vdots \\ \vdots & & \vdots \end{array}$$

$$U_1(Y^N)p(Y^N) = \delta \sum_{j=1}^N \pi_{N1} U_1(Y^j) Y^j + \delta \sum_{j=1}^N \pi_{Nj} U_1(Y^j) p(Y^j)$$

with unknowns $p(Y^1), p(Y^2), \dots, p(Y^N)$. Notice that for each of these equations, the first term on the right-hand side is simply a number while the second term is a linear combination of the $p(Y^j)$ s. Barring a very unusual output process, this system will have a solution: one price for each Y^j , that is, the equilibrium price function.

Let us illustrate: Suppose $U(c) = \ln(c)$, $\delta = .96$ and $(Y^1, Y^2, Y^3) = (1.5, 1, .5)$ – an exaggeration of *boom, normal, and depression* times. The transition matrix is taken to be as found in Table 9.2.

Table 9.2 Transition Matrix

$$\begin{array}{ccc} & 1.5 & 1 & .5 \\ \begin{array}{c} 1.5 \\ 1 \\ .5 \end{array} & \left[\begin{array}{ccc} .5 & .25 & .25 \\ .25 & .5 & .25 \\ .25 & .25 & .5 \end{array} \right] & \end{array}.$$

The equilibrium conditions implicit in Equation (9.2) then reduce to :

$$\begin{aligned} Y^1 : 2/3p(1.5) &= .96 + .96 \left\{ \frac{1}{3}p(1.5) + \frac{1}{4}p(1) + \frac{1}{2}p(.5) \right\} \\ Y^2 : p(1) &= .96 + .96 \left\{ \frac{1}{6}p(1.5) + \frac{1}{2}p(1) + \frac{1}{2}p(.5) \right\} \\ Y^3 : 2p(.5) &= .96 + .96 \left\{ \frac{1}{6}p(1.5) + \frac{1}{4}p(1) + 1p(.5) \right\} \\ Y^1 : 0 &= .96 - .347p(1.5) + .24p(1) + .48p(.5) & (i) \\ \text{or, } Y^2 : 0 &= .96 + .16p(1.5) - .52p(1) + .48p(.5) & (ii) \\ Y^3 : 0 &= .96 + .16p(1.5) + .24p(1) - 1.04p(.5) & (iii) \end{aligned}$$

$$\begin{aligned} (i)-(ii) \text{ yields: } p(1.5) &= \frac{.76}{.507}p(1) = 1.5p(1) & (iv) \\ (ii)-(iii) \text{ gives: } p(.5) &= \frac{.76}{1.52}p(1) = \frac{1}{2}p(1) & (v) \end{aligned}$$

substituting (iv) and (v) into Equation (i) to solve for $p(1)$ yields $p(1) = 24$; $p(1.5) = 36$ and $p(.5) = 12$ follow. \square

9.3.2 Interpreting the Exchange Equilibrium

To bring about a closer correspondence with traditional asset pricing formulae we must first relate the asset prices derived previously to rates of return. In particular, we will want to understand, in this model context, what determines

the amount by which the risky asset's expected return exceeds that of a risk-free asset. This basic question is also the one for which the standard CAPM provides such a simple, elegant answer ($E\tilde{r}_j - r_f = \beta_j(E\tilde{r}_M - r_f)$). Define the period t to $t+1$ return for security j as

$$1 + r_{j,t+1} = \frac{p_{j,t+1} + Y_{j,t+1}}{p_{j,t}}$$

Then Equation 9.3 may be rewritten as:

$$1 = \delta E_t \left\{ \frac{U_1(\tilde{c}_{t+1})}{U_1(c_t)} (1 + \tilde{r}_{j,t+1}) \right\} \quad (9.6)$$

Let q_t^b denote the price in period t of a one-period riskless discount bond in zero net supply, which pays 1 unit of consumption (income) in every state in the next period. By reasoning analogous to that presented previously,

$$q_t^b U_1(c_t) = \delta E_t \{U_1(\tilde{c}_{t+1})1\}$$

The price q_t^b is the equilibrium price at which the agent desires to hold zero units of the security, and thus supply equals demand. This is so because if he were to buy one unit of this security at a price q_t^b , the loss in utility today would exactly offset the gain in expected utility tomorrow. The representative agent is, therefore, content to hold zero units of the security.

Since the risk-free rate over the period from date t to $t+1$, denoted $r_{f,t+1}$, is defined by $q_t^b(1 + r_{f,t+1}) = 1$, we have

$$\frac{1}{1 + r_{f,t+1}} = q_t^b = \delta E_t \left\{ \frac{U_1(\tilde{c}_{t+1})}{U_1(c_t)} \right\}, \quad (9.7)$$

which formally establishes the link between the discount rate and the risk-free rate of return we have used in Equation (9.5) under the risk neutrality hypothesis. Note that in the latter case ($U_{11} = 0$), Equation (9.7) implies that the risk-free rate must be a constant.

Now we will combine Equations (9.6) and (9.7). Since, for any two random variables \tilde{x}, \tilde{y} , $E(\tilde{x} \cdot \tilde{y}) = E(\tilde{x}) \cdot E(\tilde{y}) + cov(\tilde{x}, \tilde{y})$, we can rewrite equation (9.6) in the form

$$1 = \delta E_t \left\{ \frac{U_1(\tilde{c}_{t+1})}{U_1(c_t)} \right\} E_t \{1 + \tilde{r}_{j,t+1}\} + \delta cov_t \left\{ \frac{U_1(\tilde{c}_{t+1})}{U_1(c_t)}, \tilde{r}_{j,t+1} \right\} \quad (9.8)$$

Let us denote $E_t \{1 + \tilde{r}_{j,t+1}\} = 1 + \bar{r}_{j,t+1}$. Then substituting Equation (9.7) into Equation (9.8) gives

$$\begin{aligned} 1 &= \frac{1 + \bar{r}_{j,t+1}}{1 + r_{f,t+1}} + \delta cov_t \left(\frac{U_1(\tilde{c}_{t+1})}{U_1(c_t)}, \tilde{r}_{j,t+1} \right), \text{ or, rearranging,} \\ \frac{1 + \bar{r}_{j,t+1}}{1 + r_{f,t+1}} &= 1 - \delta cov_t \left(\frac{U_1(\tilde{c}_{t+1})}{U_1(c_t)}, \tilde{r}_{j,t+1} \right), \text{ or} \\ \bar{r}_{j,t+1} - r_{f,t+1} &= -\delta (1 + r_{f,t+1}) cov_t \left(\frac{U_1(\tilde{c}_{t+1})}{U_1(c_t)}, \tilde{r}_{j,t+1} \right). \end{aligned} \quad (9.9)$$

Equation (9.9) is the central relationship of the consumption CAPM and we must consider its implications. The LHS of Equation (9.9) is the risk-premium on security j . Equation (9.9) tells us that the risk premium will be large when $cov_t \left(\frac{U_1(\tilde{c}_{t+1})}{U_1(c_t)}, \tilde{r}_{j,t+1} \right)$ is large and negative, that is, for those securities paying high returns when consumption is high (and thus when $U_1(c_{t+1})$ is low), and low returns when consumption is low (and $U_1(c_{t+1})$ is high). These securities are not very desirable for consumption risk reduction (consumption smoothing): They pay high returns when investors don't need them (consumption is high anyway) and low returns when they are most needed (consumption is low). Since they are not desirable, they have a low price and high expected returns relative to the risk-free security.

The CAPM tells us that a security is relatively undesirable and thus commands a high return when it covaries positively with the market portfolio, that is, when its return is high precisely in those circumstances when the return on the market portfolio is also high, and conversely. The consumption CAPM is not in contradiction with this basic idea but it adds some further degree of precision. From the viewpoint of smoothing *consumption* and risk diversification, an asset is desirable if it has a high return when consumption is low and vice versa.

When the portfolio and asset pricing problem is placed in its proper multi-period context, the notion of utility of end of period wealth (our paradigm of Chapters 5 to 7) is no longer relevant and we have to go back to the more fundamental formulation in terms of the utility derived from consumption: $U(c_t)$. But then it becomes clear that the possibility of expressing the objective as maximizing the utility of end-of-period wealth in the 2 dates setting has, in some sense, lured us down a false trail: In a fundamental sense, the key to an asset's value is its covariation with the marginal utility of consumption, not with the marginal utility of wealth.

Equation (9.9) has the unappealing feature that the risk premium is defined, in part, in terms of the marginal utility of consumption, which is not observable. To eliminate this feature, we shall make the following approximation.

Let $U(c_t) = ac_t - \frac{b}{2}c_t^2$ (i.e., a quadratic utility function or a truncated Taylor series expansion of a general $U(\cdot)$) where $a > 0, b > 0$, and the usual restrictions apply on the range of consumption. It follows that $U_1(c_t) = a - bc_t$; substituting this into Equation (9.9) gives

$$\begin{aligned} \bar{r}_{j,t+1} - r_{f,t+1} &= -\delta(1 + r_{f,t+1}) cov_t \left(\tilde{r}_{j,t+1}, \frac{a - b\tilde{c}_{t+1}}{a - bc_t} \right) \\ &= -\delta(1 + r_{f,t+1}) \frac{1}{a - bc_t} cov_t(\tilde{r}_{j,t+1}, \tilde{c}_{t+1})(-b), \text{ or} \\ \bar{r}_{j,t+1} - r_{f,t+1} &= \frac{\delta b(1 + r_{f,t+1})}{a - bc_t} cov_t(\tilde{r}_{j,t+1}, \tilde{c}_{t+1}). \end{aligned} \quad (9.10)$$

Equation (9.10) makes this point easier to grasp: since the term in front of the covariance expression is necessarily positive, if next-period consumption covaries in a large positive way with $r_{j,t+1}$, then the risk premium on j will be high.

9.3.3 The Formal Consumption CAPM

As a final step in our construction, let us denote the portfolio most highly correlated with consumption by the index $j = c$, and its expected rate of return for the period from t to $t + 1$ by $\bar{r}_{c,t+1}$

Equation (9.10) applies as well to this security so we have

$$\bar{r}_{c,t+1} - r_{f,t+1} = \left[\frac{\delta b (1 + r_{f,t+1})}{a - bc_t} \right] cov_t(\tilde{r}_{c,t+1}, \tilde{c}_{t+1}). \quad (9.11)$$

Dividing Equation (9.10) by (9.11) and thus eliminating the term $\left[\frac{\delta(1+r_{f,t+1})b}{a-bc_t} \right]$, one obtains

$$\begin{aligned} \frac{\bar{r}_{j,t+1} - r_{f,t+1}}{\bar{r}_{c,t+1} - r_{f,t+1}} &= \frac{cov_t(\tilde{r}_{j,t+1}, \tilde{c}_{t+1})}{cov_t(\tilde{r}_{c,t+1}, \tilde{c}_{t+1})}, \text{ or} \\ \frac{\bar{r}_{j,t+1} - r_{f,t+1}}{\bar{r}_{c,t+1} - r_{f,t+1}} &= \frac{\frac{cov_t(\tilde{r}_{j,t+1}, \tilde{c}_{t+1})}{var(\tilde{c}_{t+1})}}{\frac{cov_t(\tilde{r}_{c,t+1}, \tilde{c}_{t+1})}{var(\tilde{c}_{t+1})}}, \text{ or} \\ \bar{r}_{j,t+1} - r_{f,t+1} &= \frac{\beta_{j,c_t}}{\beta_{c,c_t}} [\bar{r}_{c,t+1} - r_{f,t+1}] \end{aligned} \quad (9.12)$$

for $\beta_{j,c_t} = \frac{cov_t(\tilde{r}_{j,t+1}, \tilde{c}_{t+1})}{var(\tilde{c}_{t+1})}$, the consumption- β of asset j , and $\frac{cov_t(\tilde{r}_{c,t+1}, \tilde{c}_{t+1})}{var(\tilde{c}_{t+1})}$, the consumption- β of portfolio c . This equation defines the consumption CAPM.

If it is possible to construct a portfolio c such that $\beta_{c,c_t} = 1$ one gets the direct analogue to the CAPM, with $\bar{r}_{c,t+1}$ replacing the expected return on the market and β_{j,c_t} the relevant beta:

$$\bar{r}_{j,t+1} - r_{f,t+1} = \beta_{j,c_t} (\bar{r}_{c,t+1} - r_{f,t+1}). \quad (9.13)$$

9.4 Pricing Arrow-Debreu State-Contingent Claims with the CCAPM

Chapter 8 dwelled on the notion of an Arrow-Debreu state claim as the basic building block for all asset pricing and it is interesting to understand what form these securities and their prices assume in the consumption CAPM setting. Our treatment will be very general and will accommodate more complex settings where the state is characterized by more than one variable.

Whatever model we happen to use, let s_t denote the state in period t . In the prior sections s_t coincided with the period t output, Y_t .

Given that we are in state s in period t , what is the price of an Arrow-Debreu security that pays 1 unit of consumption if and only if state s' occurs in period $t+1$? We consider two cases.

1. Let the number of possible states be finite; denote the Arrow-Debreu price as

$$q(s_{t+1} = s'; s_t = s)$$

with the prime superscript referring to the value taken by the random state variable in the next period. Since this security is assumed to be in zero net supply,⁵ it must satisfy, in equilibrium,

$$U_1(c(s)) q(s_{t+1} = s'; s_t = s) = \delta U_1(c(s')) \text{prob}(s_{t+1} = s'; s_t = s),$$

or

$$q(s_{t+1} = s'; s_t = s) = \delta \frac{U_1(c(s'))}{U_1(c(s))} \text{prob}(s_{t+1} = s'; s_t = s).$$

As a consequence of our maintained stationarity hypothesis, the same price occurs when the economy is in state s and the claim pays 1 unit of consumption in the next period if and only if state s' occurs, whatever the current time period t . We may thus drop the time subscript and write

$$q(s'; s) = \delta \frac{U_1(c(s'))}{U_1(c(s))} \text{prob}(s'; s) = \delta \frac{U_1(c(s'))}{U_1(c(s))} \pi_{ss'},$$

in the notation of our transition matrix representation. This is equation (8.1) of our previous chapter.

2. For a continuum of possible states, the analogous expression is

$$q(s'; s) = \delta \frac{U_1(c(s'))}{U_1(c(s))} f(s'; s)$$

where $f(s'; s)$ is the conditional density function on s_{t+1} given s , evaluated at s' .

Note that *under risk neutrality*, we have a reconfirmation of our earlier identification of Arrow-Debreu prices as being proportional to the relevant state probabilities, with the proportionality factor corresponding to the time discount coefficient:

$$q(s'; s) = \delta f(s'; s) = \delta \pi_{ss}.$$

These prices are for one-period state-contingent claims; what about N -period claims? They would be priced exactly analogously:

$$q^N(s_{t+N} = s'; s_t = s) = \delta^N \frac{U_1(c(s'))}{U_1(c(s))} \text{prob}(s_{t+N} = s'; s_t = s).$$

The price of an N -period risk-free discount bond q_t^{bN} , given state s , is thus given by

$$q_t^{bN}(s) = \delta^N \sum_{s'} \frac{U_1(c(s'))}{U_1(c(s))} \text{prob}(s_{t+N} = s'; s_t = s) \quad (9.14)$$

⁵And thus its introduction does not alter the structure of the economy described previously.

or, in the continuum of states notation,

$$q_t^{bN}(s) = \delta^N \int_{s'} \frac{U_1(c(s'))}{U_1(c(s))} f_N(s'; s) ds' = E_s \left\{ \delta^N \frac{U_1(c_{t+N}(s'))}{U_1(c(s))} \right\},$$

where the expectation is taken over all possible states s' conditional on the current state being s .⁶

Now let us review Equation (9.4) in the light of the expressions we have just derived.

$$\begin{aligned} p_t &= E_t \sum_{\tau=1}^{\infty} \delta^\tau \left[\frac{U_1(c_{t+\tau})}{U_1(c_t)} Y_{t+\tau} \right] \\ &= \sum_{\tau=1}^{\infty} \sum_{s'} \delta^\tau \left[\frac{U_1(c_{t+\tau}(s'))}{U_1(c_t)} Y_{t+\tau}(s') \right] \text{prob}(s_{t+\tau} = s'; s_t = s) \\ &= \sum_{\tau} \sum_{s'} q^\tau(s', s) Y_{t+\tau}(s'), \end{aligned} \tag{9.15}$$

What this development tells us is that taking the appropriately discounted (at the inter-temporal MRS) sum of expected future dividends is simply valuing the stream of future dividends at the appropriate Arrow-Debreu prices! The fact that there are no restrictions in the present context in extracting the prices of Arrow-Debreu contingent claims is indicative of the fact that this economy is one of complete markets.⁷

Applying the same substitution to Equation (9.4) as employed to obtain Equation (9.8) yields:

$$\begin{aligned} p_t &= \sum_{\tau=1}^{\infty} \delta^\tau \left\{ E_t \left[\frac{U_1(\tilde{c}_{t+\tau})}{U_1(c_t)} \right] E_t [\tilde{Y}_{t+\tau}] + \text{cov} \left(\frac{U_1(\tilde{c}_{t+\tau})}{U_1(c_t)}, \tilde{Y}_{t+\tau} \right) \right\} \\ &= \sum_{\tau=1}^{\infty} \delta^\tau \left\{ E_t \left[\frac{U_1(\tilde{c}_{t+\tau})}{U_1(c_t)} \right] E_t [\tilde{Y}_{t+\tau}] \left\langle 1 + \frac{\text{cov} \left(\frac{U_1(\tilde{c}_{t+\tau})}{U_1(c_t)}, \tilde{Y}_{t+\tau} \right)}{E_t \left[\frac{U_1(\tilde{c}_{t+\tau})}{U_1(c_t)} \right] E_t [\tilde{Y}_{t+\tau}]} \right\rangle \right\}, \end{aligned}$$

where the expectations operator applies across all possible values of the state output variable, with probabilities given on the line corresponding to the current state s_t in the matrix \mathbb{T} raised to the relevant power (the number of periods to the date of availability of the relevant cash flow).

Using the expression for the price of a risk-free discount bond of τ periods to maturity derived earlier and the fact that $(1 + r_{f,t+\tau})^\tau q_t^{b\tau} = 1$, we can rewrite

⁶The corresponding state probabilities are given by the N th power of the matrix \mathbb{T} .

⁷This result, which is not trivial (we have an infinity of states of nature and only one asset - the equity), is the outcome of the twin assumptions of rational expectations and agents' homogeneity.

this expression as:

$$p_t = \sum_{\tau=1}^{\infty} \frac{\left\{ E_t[Y_{t+\tau}] \left\{ 1 + \frac{\text{cov}(U_1(\tilde{c}_{t+\tau}), \tilde{Y}_{t+\tau})}{E_t[U_1(\tilde{c}_{t+\tau})]E_t[\tilde{Y}_{t+\tau}]} \right\} \right\}}{(1 + r_{f,t+\tau})^\tau}. \quad (9.16)$$

The quantity being discounted (at the risk-free rate applicable to the relevant period) in the present value term is the equilibrium certainty equivalent of the real cash flow generated by the asset. This is the analogue for the CCAPM of the CAPM expression derived in Section 7.3.

If the cash flows exhibit no stochastic variation (i.e., they are risk free), then Equation (9.16) reduces to

$$p_t = \sum_{\tau=1}^{\infty} \frac{Y_{t+\tau}}{(1 + r_{f,t+\tau})^\tau}.$$

This relationship will be derived again in Chapter 10 where we discounted risk-free cash flows at the term structure of interest rates. If, on the other hand, the cash flows are risky, yet investors are risk neutral (constant marginal utility of consumption), Equation (9.16) becomes

$$p_t = \sum_{\tau=1}^{\infty} \frac{E[\tilde{Y}_{t+\tau}]}{(1 + r_{f,t+\tau})^\tau} \quad (9.17)$$

which is identical to Equation (9.5) once we recall, from Equation (9.7), that the risk-free rate must be constant under risk neutrality.

Equation (9.16) is fully in harmony with the intuition of Section 9.3: if the representative agent's consumption is highly positively correlated with the security's real cash flows, the certainty equivalent values of these cash flows will be smaller than their expected values (viz., $\text{cov}(U_1(c_{t+\tau}), Y_{t+\tau}) < 0$). This is so because such a security is not very useful for hedging the agent's future consumption risk. As a result it will have a low price and a high expected return. In fact, its price will be less than what it would be in an economy of risk-neutral agents [Equation (9.17)]. The opposite is true if the security's cash flows are negatively correlated with the agent's consumption.

9.5 Testing the Consumption CAPM: The Equity Premium Puzzle

In the rest of this chapter we discuss the empirical validity of the CCAPM. We do this here (and not with the CAPM and other pricing models seen so far) because a set of simple and robust empirical observations has been put forward that falsifies this model in an unusually strong way. This forces us to question its underlying hypotheses and, a fortiori, those underlying some of the less-sophisticated models seen before. Thus, in this instance, the recourse to sophisticated econometrics for drawing significant lessons about our approach to modeling financial markets is superfluous.

A few key empirical observations regarding financial returns in U.S. markets are summarized in Table 9.3, which shows that over a long period of observation the average ex-post return on a diversified portfolio of U.S. stocks (the market portfolio, as approximated in the U.S. by the S&P 500) has been close to 7 percent (in real terms, net of inflation) while the return on *one-year* T-bills (taken to represent the return on the risk-free asset) has averaged less than 1 percent. These twin observations make up for an equity risk premium of 6.2 percent. This observation is robust in the sense that it has applied in the U.S. for a very long period, and in several other important countries as well. Its meaning is not totally undisputed, however. Goetzmann and Jorion (1999), in particular, argue that the high return premium obtained for holding U.S. equities is the exception rather than the rule.⁸

Here we will take the 6 percent equity premium at face value, as has the huge literature that followed the uncovering of the *equity premium puzzle* by Mehra and Prescott (1985). The puzzle is this: Mehra and Prescott argue that the CCAPM is completely unable, once reasonable parameter values are inserted in the model, to replicate such a high observed equity premium.

Table 9.3 : Properties of U.S. Asset Returns

	U.S. Economy	
	(a)	(b)
r	6.98	16.54
r_f	.80	5.67
$r - r_f$	6.18	16.67

(a) Annualized mean values in percent;

(b) Annualized standard deviation in percent.

Source: Data from Mehra and Prescott (1985).

Let us illustrate their reasoning. According to the consumption CAPM, the only factors determining the characteristics of security returns are the representative agent's utility function, his subjective discount factor, and the process on consumption (which equals output or dividends in the exchange economy equilibrium). Consider the utility function first. It is natural in light of Chapter 4 to assume the agent's period utility function displays CRRA; thus let us set

$$U(c) = \frac{c^{1-\gamma}}{1-\gamma}.$$

⁸Using shorter, mostly postwar, data, premia close or even higher than the U.S. equity premium are obtained for France, Germany, the Netherlands, Sweden, Switzerland, and the United Kingdom [see, e.g., Campbell (1998)]. Goetzmann and Jorion however argue that such data samples do not correct for crashes and period of market interruptions, often associated with WWII, and thus are not immune from a survivorship bias. To correct for such a bias, they assemble long data series for all markets that existed during the twentieth century. They find that the United States has had "by far the highest uninterrupted real rate of appreciation of all countries, at about 5 percent annually. For other countries, the median appreciation rate is about 1.5 percent."

Empirical studies associated with this model have placed γ in the range of (1, 2). A convenient consequence of this utility specification is that the inter-temporal marginal rate of substitution can be written as

$$\frac{U_1(c_{t+1})}{U_1(c_t)} = \left(\frac{c_{t+1}}{c_t} \right)^{-\gamma}. \quad (9.18)$$

The second major ingredient is the consumption process. In our version of the model, consumption is a stationary process: It does not grow through time. In reality, however, consumption *is* growing through time. In a growing economy, the analogous notion to the variability of consumption is variability in the growth rate of consumption.

Let $x_{t+1} = \frac{c_{t+1}}{c_t}$, denote per capita consumption growth, and assume, for illustration that x_t is independently and identically lognormally distributed through time. For the period 1889 through 1978, the U.S. economy aggregate consumption has been growing at an average rate of 1.83 percent annually with a standard deviation of 3.57 percent, and a slightly negative measure of autocorrelation (-.14) [cf. Mehra and Prescott (1985)].

The remaining item is the agent's subjective discount factor δ : What value should it assume? Time impatience requires, of course, that $\delta < 1$, but this is insufficiently precise. One logical route to its estimation is as follows: Roughly speaking, the equity in the CCAPM economy represents a claim to the aggregate income from the underlying economy's entire capital stock. We have just seen that, in the United States, equity claims to private capital flows average a 7 percent annual real return, while debt claims average 1 percent.⁹ Furthermore, the economy-wide debt-to-equity rates are not very different from 1. These facts together suggest an overall average real annual return to capital of about 4 percent.

If there were no uncertainty in the model, and if the constant growth rate of consumption were to equal its long-run historical average (1.0183), the asset pricing Equation (9.6) would reduce to

$$1 = \delta E_t \left\{ \left(\frac{\tilde{c}_{t+1}}{c_t} \right)^{-\gamma} R_{t+1} \right\} = \delta(\bar{x})^{-\gamma} \bar{R}, \quad (9.19)$$

where R_{t+1} is the gross rate of return on capital and the upper bars denote historical averages.¹⁰ For $\gamma = 1$, $\bar{x} = 1.0183$, and $\bar{R} = 1.04$, we can solve for the implied δ to obtain $\delta \cong 0.97$. Since we have used an annual estimate for \bar{x} , the resulting δ must be viewed as an annual or yearly subjective discount factor; on a quarterly basis it corresponds to $\delta \cong 0.99$. If, on the other hand, we want to assume $\gamma = 2$, Equation (9.19) solves for $\delta = .99$ on an annual basis, yielding a quarterly δ even closer to 1. This reasoning demonstrates that assuming higher

⁹Strictly speaking, these are the returns to publicly traded debt and equity claims. If private capital earns substantially different returns, however, capital is being inefficiently allocated; we assume this is not the case.

¹⁰Time average and expected values should coincide in a stationary model, provided the time series is of sufficient length.

rates of risk aversion would be incompatible with maintaining the hypothesis of a time discount factor less than 1. While technically, in the case of positive consumption growth, we could entertain the possibility of a negative rate of time preference, and thus of a discount factor larger than 1, we rule it out on grounds of plausibility.

At the root of this difficulty is the low return on the risk-free asset (1 percent), which will haunt us in other ways. As we know, highly risk-averse individuals want to smooth consumption over time, meaning they want to transfer consumption from good times to bad times. When consumption is growing predictably, the good times lie in the future. Agents want to borrow now against their future income. In a representative agent model, this is hard to reconcile with a low rate on borrowing: everyone is on the same side of the market, a fact that inevitably forces a higher rate. This problem calls for an independent explanation for the abnormally low average risk-free rate [e.g., in terms of the liquidity advantage of short-term government debt as in Bansal and Coleman (1996)] or the acceptance of the possibility of a negative rate of time preference so that future consumption is given more weight than present consumption. We will not follow either of these routes here, but rather will, in the course of the present exercise, limit the coefficient of relative risk aversion to a maximal value of 2.

With these added assumptions we can manipulate the fundamental asset pricing Equation (9.2) to yield two equations that can be used indirectly to test the model. The key step in the reasoning is to demonstrate that, in the context of these assumptions, the equity price formula takes the form

$$p_t = vY_t$$

where v is a constant coefficient. That is, the stock price at date t is proportional to the dividend paid at date t .¹¹ To confirm this statement, we use a standard trick consisting of guessing that this is the form taken by the equilibrium pricing function and then verifying that this guess is indeed borne out by the structure of the model. Under the $p_t = vY_t$ hypothesis, Equation (9.1) becomes:

$$vY_t = \delta E_t \left\{ \left(v\tilde{Y}_{t+1} + \tilde{Y}_{t+1} \right) \frac{U_1(\tilde{c}_{t+1})}{U_1(c_t)} \right\}.$$

Using Equation (9.18) and dropping the conditional expectations operator, since x is independently and identically distributed through time (its mean is independent of time), this equation can be rewritten as

$$v = \delta E \left\{ (v+1) \frac{\tilde{Y}_{t+1}}{Y_t} \tilde{x}_{t+1}^{-\gamma} \right\}.$$

¹¹Note that this property holds true as well for the example developed in Box 9.2 as Equations (iv) and (v) attest.

The market clearing condition implies that $\frac{Y_{t+1}}{Y_t} = x_{t+1}$, thus

$$\begin{aligned} v &= \delta E \{ (v+1) \tilde{x}^{1-\gamma} \} \\ &= \frac{\delta E \{ \tilde{x}^{1-\gamma} \}}{1 - \delta E \{ \tilde{x}^{1-\gamma} \}}. \end{aligned}$$

This is indeed a constant and our initial guess is thus confirmed!

Taking advantage of the validated pricing hypothesis, the equity return can be written as:

$$R_{t+1} \equiv 1 + r_{t+1} = \frac{p_{t+1} + Y_{t+1}}{p_t} = \frac{v+1}{v} \frac{Y_{t+1}}{Y_t} = \frac{v+1}{v} x_{t+1}.$$

Taking expectations we obtain:

$$E_t \left(\tilde{R}_{t+1} \right) = E \left(\tilde{R}_{t+1} \right) = \frac{v+1}{v} E \left(\tilde{x}_{t+1} \right) = \frac{E(\tilde{x})}{\delta E \{ \tilde{x}^{1-\gamma} \}}.$$

The risk-free rate is [Equation (9.7)]:

$$R_{f,t+1} \equiv \frac{1}{q_t^b} = \left[\delta E_t \left\{ \frac{U_1(\tilde{c}_{t+1})}{U_1(c_t)} \right\} \right]^{-1} = \frac{1}{\delta} \frac{1}{E \{ \tilde{x}^{-\gamma} \}}, \quad (9.20)$$

which is seen to be constant under our current hypotheses.

Taking advantage of the lognormality hypothesis, the ratio of the two preceding equations can be expressed as (see Appendix 9.2 for details)

$$\frac{E \left(\tilde{R}_{t+1} \right)}{R_f} = \frac{E \{ \tilde{x} \} E \{ \tilde{x}^{-\gamma} \}}{E \{ \tilde{x}^{1-\gamma} \}} = \exp \left[\gamma \sigma_x^2 \right], \quad (9.21)$$

where σ_x^2 is the variance of $\ln x$. Taking logs, we finally obtain:

$$\ln(ER) - \ln(R_f) = \gamma \sigma_x^2. \quad (9.22)$$

Now, we are in a position to confront the model with the data. Let us start with Equation (9.22). Feeding in the return characteristics of the U.S. economy and solving for γ , we obtain (see Appendix 9.2 for the computation of σ_x^2),

$$\frac{\ln(ER) - \ln(E_{r_f})}{\sigma_x^2} = \frac{1.0698 - 1.008}{(.0357)^2} = 50.24 = \gamma.$$

Alternatively, if we assume $\gamma = 2$ and multiply by σ_x^2 as per Equation (9.22), one obtains an equity premium of

$$2(.00123) = .002 = (\ln(ER) - \ln(E_{r_f})) \cong ER - E_{r_f} \quad (9.23)$$

In either case, this reasoning identifies a major discrepancy between model prediction and reality. The observed equity premium can only be explained

by assuming an extremely high coefficient of relative risk aversion ($\simeq 50$), one that is completely at variance with independent estimates. An agent with risk aversion of this level would be too fearful to take a bath (many accidents involve falling in a bathtub!), or to cross the street. On the other hand, insisting on a more reasonable coefficient of risk aversion of 2 leads to predicting a minuscule premium of 0.2 percent, much below the 6.2 percent that has been historically observed over long periods.

Similarly, it is shown in Appendix 9.2 that $E\{x_t^{-\gamma}\} = .97$ for $\gamma = 2$; Equation (9.20) and the observed value for R_f (1.008) then implies that δ should be larger than 1 (1.02). This problem was to be anticipated from our discussion of the calibration of δ , which was based on reasoning similar to that underlying Equation (9.20). Here the problem is compounded by the fact that we are using an even lower risk-free rate (.8 percent) rather than the steady-state rate of return on capital of 4 percent used in the prior reasoning. In the present context, this difficulty in calibrating δ or, equivalently, in explaining the low rate of return on the risk-free asset has been dubbed the *risk-free rate puzzle* by Weil (1989). As said previously, we read this result as calling for a specific explanation for the observed low return on the risk-free asset, one that the CCAPM is not designed to provide.

9.6 Testing the Consumption CAPM: Hansen-Jagannathan Bounds

Another, parallel perspective on the puzzle is provided by the Hansen-Jagannathan (1991) bound. The idea is very similar to our prior test and the end result is the same. The underlying reasoning, however, postpones as long as possible making specific modeling assumptions. It is thus more general than a test of a specific version of the CCAPM. The bound proposed by Hansen and Jagannathan potentially applies to other asset pricing formulations. It similarly leads to a falsification of the standard CCAPM.

The reasoning goes as follows: For all homogeneous agent economies, the fundamental equilibrium asset pricing Equation (9.2) can be expressed as

$$p(s_t) = E_t[m_{t+1}(\tilde{s}_{t+1})X_{t+1}(\tilde{s}_{t+1}); s_t], \quad (9.24)$$

where s_t is the state today (it may be today's output in the context of a simple exchange economy or it may be something more elaborate as in the case of a production economy), $X_{t+1}(\tilde{s}_{t+1})$ is the total return in the next period (e.g., in the case of an exchange economy this equals $(\tilde{p}_{t+1} + \tilde{Y}_{t+1})$ and $m_{t+1}(\tilde{s}_{t+1})$ is the *equilibrium pricing kernel*, also known as the *stochastic discount factor*:

$$m_{t+1}(\tilde{s}_{t+1}) = \frac{\delta U_1(c_{t+1}(\tilde{s}_{t+1}))}{U_1(c_t)}.$$

As before $U_1(\cdot)$ is the marginal utility of the representative agent and c_t is his equilibrium consumption. Equation (9.24) is thus the general statement that the price of an asset today must equal the expectation of its total payout tomorrow

multiplied by the appropriate pricing kernel. For notational simplicity, let us suppress the state dependence, leaving it as understood, and write Equation (9.24) as

$$p_t = E_t[\tilde{m}_{t+1}\tilde{X}_{t+1}]. \quad (9.25)$$

This is equivalent to

$$1 = E_t[\tilde{m}_{t+1}\tilde{R}_{t+1}],$$

where \tilde{R}_{t+1} is the gross return on ownership of the asset. Since Equation (9.25) holds for each state s_t , it also holds unconditionally; we thus can also write

$$1 = E[\tilde{m}\tilde{R}]$$

where E denotes the unconditional expectation. For any two assets i and j (to be viewed shortly as the return on the market portfolio and the risk-free return, respectively) it must, therefore, be the case that

$$E[\tilde{m}(\tilde{R}_i - \tilde{R}_j)] = 0,$$

or

$$E[\tilde{m}\tilde{R}_{i-j}] = 0,$$

where, again for notational convenience, we substitute \tilde{R}_{i-j} for $\tilde{R}_i - \tilde{R}_j$. This latter expression furthermore implies the following series of relationships:

$$E\tilde{m}E\tilde{R}_{i-j} + cov(\tilde{m}, \tilde{R}_{i-j}) = 0,$$

or

$$E\tilde{m}E\tilde{R}_{i-j} + \rho(\tilde{m}, \tilde{R}_{i-j})\sigma_m\sigma_{R_{i-j}} = 0,$$

or

$$\frac{E\tilde{R}_{i-j}}{\sigma_{R_{i-j}}} + \rho(\tilde{m}, \tilde{R}_{i-j})\frac{\sigma_m}{E\tilde{m}} = 0,$$

or

$$\frac{E\tilde{R}_{i-j}}{\sigma_{R_{i-j}}} = -\rho(\tilde{m}, \tilde{R}_{i-j})\frac{\sigma_m}{E\tilde{m}}. \quad (9.26)$$

It follows from Equation (9.26) and the fact that a correlation is never larger than 1 that

$$\frac{\sigma_m}{E\tilde{m}} > \frac{|E\tilde{R}_{i-j}|}{\sigma_{R_{i-j}}}. \quad (9.27)$$

The inequality in expression (9.27) is referred to as the Hansen-Jagannathan lower bound on the pricing kernel. If, as noted earlier, we designate asset i as the market portfolio and asset j as the risk-free return, then the data from Table 9.3 and Equation (9.27) together imply (for the U.S. economy):

$$\frac{\sigma_m}{E\tilde{m}} > \frac{|E(\tilde{r}_M - r_f)|}{\sigma_{r_M - r_f}} = \frac{.062}{.167} = .37.$$

Let us check whether this bound is satisfied for our model. From Equation (9.18), $\tilde{m}(\tilde{c}_{t+1}, c_t) = \delta(x_t)^{-\gamma}$, the expectation of which we can be computed (See Appendix 9.2) to be

$$E\tilde{m} = \delta \exp(-\gamma\mu_x + \frac{1}{2}\gamma^2\sigma_x^2) = .99(.967945) = .96 \text{ for } \gamma = 2.$$

In fact, Equation (9.20) reminds us that Em is simply the expected value of the price of a one-period risk-free discount bound, which cannot be very far away from 1. This implies that for the Hansen-Jagannathan bound to be satisfied, the standard deviation of the pricing kernel cannot be much lower than .3; given the information we have on x_t , it is a short step to estimate this parameter numerically under the assumption of lognormality. When we do this (see Appendix 9.2 again), we obtain an estimate for $\sigma(m) = .002$, which is an order of magnitude lower than what is required for Equation (9.27) to be satisfied. The message is that it is very difficult to get the equilibrium to be anywhere near the required level. In a homogeneous agent, complete market model with standard preferences, where the variation in equilibrium consumption matches the data, consumption is just too smooth and the marginal utility of consumption does not vary sufficiently to satisfy the bound implied by the data (unless the curvature of the utility function — the degree of risk aversion — is assumed to be astronomically high, an assumption which, as we have seen, raises problems of its own).

9.7 Some Extensions

9.7.1 Reviewing the Diagnosis

Our first dynamic general equilibrium model thus fails when confronted with actual data. Let us review the source of this failure. Recall our original pricing Equation (9.9), specialized for a single asset, the market portfolio:

$$\begin{aligned} \bar{r}_{M,t+1} - r_{f,t+1} &= -\delta(1 + r_{f,t+1}) \text{cov}_t \left(\frac{U_1(\tilde{c}_{t+1})}{U_1(c_t)}, \tilde{r}_{M,t+1} \right) \\ &= -\delta(1 + r_{f,t+1}) \rho \left(\frac{U_1(\tilde{c}_{t+1})}{U_1(c_t)}, \tilde{r}_{M,t+1} \right) \sigma \left(\frac{U_1(\tilde{c}_{t+1})}{U_1(c_t)} \right) \sigma(\tilde{r}_{M,t+1}) \\ &= -(1 + r_{f,t+1}) \rho(\tilde{m}_t, \tilde{r}_{M,t+1}) \sigma(\tilde{m}_t) \sigma(\tilde{r}_{M,t+1}) \end{aligned}$$

Written in this way, it is clear that the equity premium depends upon the standard deviation of the MRS (or, equivalently, the stochastic discount factor), the standard deviation of the return on the market portfolio, and the correlation between these quantities. For the United States, and most other industrial countries, the problem with a model in which pricing and return relationships depend so much on consumption (and thus MRS) variation is that average per capita consumption does not vary much at all. If this model is to have any hope of matching the data, we must modify it in a way that will increase the standard deviation of the relevant MRS, or the variability of the dividend being

priced (and thus the $\sigma(r_{M,t+1})$). We do not have complete freedom over this latter quantity, however, as it must be matched to the data as well.

9.7.2 The CCAPM with Epstein-Zin Utility

At this stage it is interesting to inquire whether, in addition to its intellectual appeal on grounds of generality, Epstein and Zin's (1989) separation of time and risk preferences might contribute a solution to the equity premium puzzle, and more generally, alter our vision of the CCAPM and its message.

Let us start by looking specifically at the equity premium puzzle. It will facilitate our discussion to repeat Equations (5.10) and (5.12) defining the Epstein-Zin preference representation (refer to Chapter 5 for a discussion and for the log case):

$$\begin{aligned}
U(c_t, CE_{t+1}) &= \left[(1 - \delta)c_t^{\frac{1-\gamma}{\theta}} + \delta CE_{t+1}^{\frac{1-\gamma}{\theta}} \right]^{\frac{\theta}{1-\gamma}} \\
\text{where } [CE(\tilde{U}_{t+1})]^{1-\gamma} &= E_t(\tilde{U}_{t+1})^{1-\gamma}, \\
\text{and } \theta &= \frac{1-\gamma}{1-\frac{1}{\rho}}, \quad 0 < \delta < 1, \quad 1 \neq \gamma > 0, \quad \rho > 0;
\end{aligned}$$

Weil (1989) uses these preferences in a setting otherwise identical to that of Mehra and Prescott (1985). Asset prices and returns are computed similarly. What he finds, however, is that this greater generality does not resolve the *risk premium puzzle*, but rather tends to underscore what we have already introduced as the *risk-free rate puzzle*.

The Epstein-Zin (1989, 1991) preference representation does not innovate along the risk dimension, with the parameter γ alone capturing risk aversion in a manner very similar to the standard case. It is, therefore, not surprising that Weil (1989) finds that only if this parameter is fixed at implausibly high levels ($\gamma \approx 45$) can a properly calibrated model replicate the premium — the Mehra and Prescott (1985) result revisited. With respect to time preferences, if ρ is calibrated to respect empirical studies, then the model also predicts a risk-free rate that is much too high. The reason for this is the same as the one outlined at the end of Section 9.5: Separately calibrating the intertemporal substitution parameter ρ tends to strengthen the assumption that the representative agent is highly desirous of a smooth inter-temporal consumption stream. With consumption growing on average at 1.8 percent per year, the agent must be offered a very high risk-free rate in order to be induced to save more and thus making his consumption tomorrow even more in excess of what it is today (less smoothing).

While Epstein and Zin preferences do not help in solving the equity premium puzzle, it is interesting to study a version of the CCAPM with these generalized preferences. The idea is that the incorporation of separate time and

risk preferences may enhance the ability of that class of models to explain the general pattern of security returns beyond the equity premium itself. The setting is once again a Lucas (1978) style economy with N assets, with the return on the equilibrium portfolio of all assets representing the return on the market portfolio. Using an elaborate dynamic programming argument, Epstein and Zin (1989, 1991) derive an asset pricing equation of the form

$$E_t \left\{ \left[\delta \left(\frac{\tilde{c}_{t+1}}{c_t} \right)^{-\frac{1}{\rho}} \right]^\theta \left[\frac{1}{1 + \tilde{r}_{M,t+1}} \right]^{1-\theta} (1 + \tilde{r}_{j,t+1}) \right\} \equiv 1, \quad (9.28)$$

where $\tilde{r}_{M,t}$ denotes the period t return on the market portfolio, and r_t^j the period t return on some asset in it. Note that when time and risk preferences coincide ($\gamma = \frac{1}{\rho}, \theta = 1$), Equation (9.28) reduces to the pricing equation of the standard time-separable CCAPM case.

The pricing kernel itself is of the form

$$\left[\delta \left(\frac{\tilde{c}_{t+1}}{c_t} \right)^{-\frac{1}{\rho}} \right]^\theta \left[\frac{1}{1 + \tilde{r}_{M,t+1}} \right]^{1-\theta}, \quad (9.29)$$

which is a geometric average (with weights θ and $1 - \theta$, respectively) of the pricing kernel of the standard CCAPM, $\left[\delta \left(\frac{\tilde{c}_{t+1}}{c_t} \right)^{-\frac{1}{\rho}} \right]$, and the pricing kernel for the $\log(\rho = 0)$ case, $\left[\frac{1}{1 + \tilde{r}_{M,t+1}} \right]$.

Epstein and Zin (1991) next consider a linear approximation to the geometric average in Equation (9.29),

$$\theta \left[\delta \left(\frac{\tilde{c}_{t+1}}{c_t} \right)^{-\frac{1}{\rho}} \right] + (1 - \theta) \left[\frac{1}{1 + \tilde{r}_{M,t+1}} \right]. \quad (9.30)$$

Substituting Equation (9.30) into Equation (9.28) gives

$$\begin{aligned} E_t \left\{ \left\{ \theta \left[\delta \left(\frac{\tilde{c}_{t+1}}{c_t} \right)^{-\frac{1}{\rho}} \right] + (1 - \theta) \left[\frac{1}{1 + \tilde{r}_{M,t+1}} \right] \right\} (1 + \tilde{r}_{j,t+1}) \right\} &\approx 1, \text{ or} \\ E_t \left\{ \theta \left[\delta \left(\frac{\tilde{c}_{t+1}}{c_t} \right)^{-\frac{1}{\rho}} \right] (1 + \tilde{r}_{j,t+1}) + (1 - \theta) \left[\frac{1}{1 + \tilde{r}_{M,t+1}} \right] (1 + \tilde{r}_{j,t+1}) \right\} &\approx 1. \end{aligned} \quad (9.31)$$

Equation (9.31) is revealing. As we noted earlier, the standard CAPM relates the (essential, non-diversifiable) risk of an asset to the covariance of its returns with M , while the CCAPM relates its riskiness to the covariance of its returns with the growth rate of consumption (via the IMRS). With separate time and risk preferences, Equation (9.31) suggests that both covariances matter for an asset's return pattern¹². But why are these effects both present separately and

¹²To see this recall that for two random variables \tilde{x} and \tilde{y} , $E(\tilde{x}\tilde{y}) = E(\tilde{x})E(\tilde{y}) + cov(\tilde{x}, \tilde{y})$, and employ this substitution in both terms on the left-hand side of Equation (9.31).

individually? The covariance of an asset's return with M captures its atemporal, non-diversifiable risk (as in the static model). The covariance of its returns with the growth rate of consumptions fundamentally captures its risk across successive time periods. When risk and time preferences are separated, it is not entirely surprising that both sources of risk should be individually present. This relationship is more strikingly apparent if we assume joint lognormality and heteroskedasticity in consumption and asset returns; Campbell et al. (1997) then express Equation (9.31) in a form whereby the risk premium on asset i satisfies:

$$E_t(\tilde{r}_{i,t+1}) - r_{f,t+1} + \frac{\sigma_i^2}{2} = \delta \frac{\sigma_{ic}}{\psi} + (1 - \delta)\sigma_{iM}, \quad (9.32)$$

where $\sigma_{ic} = \text{cov}(\tilde{r}_{it}, \frac{\tilde{c}_t}{c_{t-1}})$, and $\sigma_{iM} = \text{cov}(\tilde{r}_{it}, \tilde{r}_{M,t})$. Both sources of risk are clearly present.

9.7.3 Habit Formation

In the narrower perspective of solving the equity premium puzzle, probably the most successful modification of the standard setup has been to admit utility functions that exhibit higher rates of risk aversion at the margin, and thus can translate small variations in consumption into a large variability of the MRS. One way to achieve this objective without being confronted with the risk-free rate puzzle — which is exacerbated if we simply decide to postulate a higher γ — is to admit some form of *habit formation*. This is the notion that the agent's utility today is determined not by her absolute consumption level, but rather by the relative position of her current consumption vis-à-vis what can be viewed as a *stock of habit*, summarizing either her past consumption history (with more or less weight placed on distant or old consumption levels) or the history of aggregate consumption (summarizing in a sense the consumption habits of her neighbors; a “keeping up with the Joneses” effect). This modeling perspective thus takes the view that utility of consumption is primarily dependent on (affected by) departures from prior consumption history, either one's own or that of a social reference group; departures from what we have been accustomed to consuming; or what we may have been led to consider “fair” consumption. This concept is open to a variety of different specifications, with diverse implications for behavior and asset pricing. The interested reader is invited to consult Campbell and Cochrane (1999) for a review. Here we will be content to illustrate briefly the underlying working principle. To that end, we specify the representative agent's period preference ordering to be of the form

$$U(c_t, c_{t-1}) \equiv \frac{(c_t - \chi c_{t-1})^{1-\gamma}}{1-\gamma},$$

where $\chi \leq 1$ is a parameter. In an extreme case, $\chi = 1$, the period utility depends only upon the deviation of current period t consumption from the prior period's consumption. As we noted earlier, actual data indicates that per capita consumption for the United States and most other developed countries is

very smooth. This implies that $(c_t - c_{t-1})$ is likely to be very small most of the time. For this specification, the agent's effective (marginal) relative risk aversion reduces to $R_R(c_t) = \frac{\gamma}{1 - (c_{t-1}/c_t)}$; with $c_t \approx c_{t-1}$, the effective $R_R(c)$ will thus be very high, even with a low γ , and the representative agent will appear as though he is very risk averse. This opens the possibility for a very high return on the risky asset. With a careful choice of the habit specification, the risk-free asset pricing equation will not be materially affected and the risk-free rate puzzle will be avoided [see Constantinides (1990) and Campbell and Cochrane (1999)].

We find this development interesting not only because of its implications for pricing assets in an exchange economy. It also suggests a more general reevaluation of the standard utility framework discussed in Chapter 2. It may, however, lead to questioning some of the basic tenets of our financial knowledge: It would hardly be satisfactory to solve a puzzle by assuming habit formation and high effective rates of risk aversion and ignore this behavioral assumption when attending to other problems tackled by financial theory. In fact, a confirmed habit formation utility specification would, for the same reasons, have significant implications for macroeconomics as well (as modern macroeconomics builds on the same theoretical principles as the CCAPM of this section). It is not clear, however, that high effective rates of risk aversion are consistent with our current understanding of short-run macroeconomic fluctuations. This discussion suggests that it is worthwhile to explore alternative potential solutions to the puzzle, all the while attempting to understand better the connections between the real and the financial sides of the economy.

9.7.4 Distinguishing Stockholders from Non-Stockholders

In this spirit, another approach to addressing the outstanding financial puzzles starts by recognizing that only a small fraction of the population holds substantial financial assets, stocks in particular. This fact implies that only the variability of the consumption stream of the stockholding class should matter for pricing risky assets. There are reasons to believe that the consumption patterns of this class of the population are both more variable and more highly correlated with stock returns than average per capita consumption.¹³ Observing, furthermore, that wages are very stable and that the aggregate wage share is countercyclical (that is, proportionately larger in bad times when aggregate income is relatively low), it is not unreasonable to assume that firms, and thus their owners, the shareholders, insure workers against income fluctuations associated with the business cycle. If this is a significant feature of the real world, it should have implications for asset pricing as we presently demonstrate.

Before trying to incorporate such a feature into a CCAPM-type model, it is useful first to recall the notion of risk sharing. Consider the problem of allocating an uncertain income (consumption) stream between two agents so as to maximize overall utility. Assume, furthermore, that these income shares are

¹³Mankiw and Zeldes (1991) attempt, successfully, to confirm this conjecture. They indeed find that shareholder consumption is 2.5 times as variable as non-shareholder consumption. Data problems, however, preclude taking their results as more than indicative.

not fixed across all states, but can be allocated on a state-by-state basis. This task can be summarized by the allocation problem

$$\begin{aligned} \max_{c_1(\tilde{\theta}), c_2(\tilde{\theta})} \quad & U(c_1(\tilde{\theta})) + \mu V(c_2(\tilde{\theta})), \text{ s.t.} \\ & c_1(\tilde{\theta}) + c_2(\tilde{\theta}) \leq Y(\tilde{\theta}), \end{aligned}$$

where $U(\cdot)$, $V(\cdot)$ are, respectively, the two agents' utility functions, $c_1(\tilde{\theta})$ and $c_2(\tilde{\theta})$ their respective income assignments, $Y(\tilde{\theta})$ the economy-wide state-dependent aggregate income stream, and μ their relative weight.

The necessary and sufficient first-order condition for this problem is

$$U_1(c_1(\tilde{\theta})) = \mu V_1(c_2(\tilde{\theta})). \quad (9.33)$$

Equation (9.33) states that the ratio of the marginal utilities of the two agents should be constant. We have seen it before as Equation (8.3) of Chapter 8. As we saw there, it can be interpreted as an optimal risk sharing condition in the sense that it implicitly assigns more of the income risk to the less risk-averse agent. To see this, take the extreme case where one of the agents, say the one with utility function $V(\cdot)$ is risk neutral — indifferent to risk. According to Equation (9.33) it will then be optimal for the other agent's income stream to be constant across all states: He will be perfectly insured. Agent $V(\cdot)$ will thus absorb all the risk (in exchange for a higher average income share).

To understand the potential place of these ideas in the consumption CAPM setting, let $V(\cdot)$ now denote the period utility function of the representative shareholder, and $U(\cdot)$ the period utility function of the representative worker who is assumed not to hold any financial assets and who consequently consumes his wage w_t . As before, let Y_t be the uncertain (exogenously given) output. The investment problem of the shareholders — the maximization problem with which we started this chapter — now becomes

$$\begin{aligned} \max_{\{z_t\}} \quad & E\left(\sum_{t=0}^{\infty} \delta^t V(\tilde{c}_t)\right) \\ \text{s.t.} \quad & c_t + p_t z_{t+1} \leq z_t d_t + p_t z_t \\ & d_t = Y_t - w_t \\ & U_1(w_t) = \mu V_1(d_t), \\ & z_t \leq 1, \forall t \end{aligned}$$

Here we simply introduce a distinction between the output of the tree, Y_t , and the dividends paid to its owners, d_t , on the plausible grounds that people (workers) need to be paid to take care of the trees and collect the fruits. This payment is w_t . Moreover, we introduce the idea that the wage bill may incorporate a risk insurance component, which we formalize by assuming that the variability of wage payments is determined by an optimal risk sharing rule equivalent to Equation (9.33). One key parameter is the income share, μ , which may be interpreted as reflecting the relative bargaining strengths of the two groups. Indeed, a larger μ gives more income to the worker.

Assets in this economy are priced as before with Equation (9.1) becoming

$$V_1(c_t)p_t = \delta E_t \left\{ V_1(\tilde{c}_{t+1}) \left(\tilde{p}_{t+1} + \tilde{d}_{t+1} \right) \right\}. \quad (9.34)$$

While the differences between Equations (9.1) and (9.34) may appear purely notational, their importance cannot be overstated. First, the pricing kernel derived from Equation (9.34) will build on the firm owners' MRS, defined over shareholder consumption (dividend) growth rather than the growth in average per capita consumption. Moreover, the definition of dividends as output minus a stabilized stream of wage payments opens up the possibility that the flow of payments to which firm owners are entitled is effectively much more variable, not only than consumption but than output as well. Therein lies a concept of leverage, one that has been dubbed *operating leverage*, similar to the familiar notion of financial leverage. In the same way that bondholders come first, and are entitled to a fixed, noncontingent interest payment, workers also have priority claims to the income stream of the firm and macroeconomic data on the cyclical behavior of the wage share confirm that wage payments are more stable than aggregate income.

We have explored the potential of these ideas in recent research work (Danthine and Donaldson, 2002) to which the reader is referred for details, and find that this class of models can generate significantly increased equity premia. When we add an extra notion of distributional risk associated with the possibility that μ varies stochastically, in a way that permits better accounting of the observed behavior of the wage share over the medium run, the premium approaches 6 percent.

9.8 Conclusions

The two modifications discussed in the previous section are far from representing the breadth and depth of the research that has been stimulated by the provocative result presented in Mehra and Prescott (1985). For a broader and more synthetic perspective, we refer the reader to the excellent recent survey of Kocherlakota (1996).

The material covered in this chapter contains recent developments illustrating some of the most important directions taken by modern financial theory. Much work remains to be done, as the latest sections indicate, and this is indeed a fertile area for current research. At this juncture, one may be led to the view that structural asset pricing theory, based on rigorous dynamic general equilibrium models, provides limited operational support in our quest for understanding financial market phenomena. While the search goes on, this state of affairs nevertheless explains the popularity of less encompassing approaches based on the concept of arbitrage reviewed in Chapters 10 to 13.

References

Bansal, R., Coleman, W.J. (1996), "A Monetary Explanation of the Equity

- Premium, Term Premium and Risk-Free Rates Puzzles,” *Journal of Political Economy* 104, 1135–1171.
- Barro, R. J. (1974), “Are Government Bonds Net Wealth?” *Journal of Political Economy* 82, 1095–1117.
- Campbell, J. Y. (1998), “Asset Prices, Consumption, and the Business Cycle,” NBER Working paper 6485, March 1998, forthcoming in the *Handbook of Macroeconomics*, Amsterdam: North Holland.
- Campbell, J. Y., Cochrane, J.H. (1999), “By Force of Habit: A Consumption-Based Explanation of Aggregate Stock Market Behavior,” *Journal of Political Economy*, 107, 205–251.
- Campbell, J., Lo, A., MacKinlay, A.C. (1997), *The Econometrics of Financial Markets*, Princeton University Press, Princeton, N.J.
- Constantinides, G. M. (1982), “Intertemporal Asset Pricing with Heterogeneous Consumers and without Demand Aggregation,” *Journal of Business* 55, 253–267.
- Constantinides, G. M. (1990), “Habit Formation: A Resolution of the Equity Premium Puzzle,” *Journal of Political Economy* 98, 519–543.
- Danthine, J. P., Donaldson, J.B. (2002), “Labor Relations and Asset Returns,” *Review of Economic Studies*, 69, 41–64.
- Epstein, L., Zin, S. (1989), “Substitution, Risk Aversion, and the Temporal Behavior of Consumption and Asset Returns: A Theoretical Framework,” *Econometrica* 57, 937–969.
- Epstein, L., Zin, S. (1991), “Substitution, Risk Aversion, and the Temporal Behavior of Consumption and Asset Returns: An Empirical Analysis,” *Journal of Political Economy* 99, 263–286.
- Goetzman, W., Jorion, P. (1999), “A Century of Global Stock Markets,” *Journal of Finance*, 55, 953–980.
- Hansen, L., Jagannathan, R. (1991), “Implications of Security Market Data for Models of Dynamic Economies,” *Journal of Political Economy* 99, 225–262.
- Kocherlakota, N. (1996), “The Equity Premium: It’s Still a Puzzle,” *Journal of Economic Literature* 34, 42–71.
- Lucas, R. E. (1978), “Asset Pricing in an Exchange Economy,” *Econometrica* 46, 1429–1445.
- Mankiw, G., Zeldes, S. (1991), “The Consumption of Stockholders and Non-Stockholders,” *Journal of Financial Economics* 29, 97–112.

Mehra, R., Prescott, E.C. (1985), “The Equity Premium: A Puzzle,” *Journal of Monetary Economics*, 15, 145–161.

Weil, Ph. (1989), “The Equity Premium Puzzle and the Risk-Free Rate Puzzle,” *Journal of Monetary Economics* 24, 401–421.

Appendix 9.1: Solving the CCAPM with Growth

Assume that there is a finite set of possible growth rates $\{x_1, \dots, x_N\}$ whose realizations are governed by a Markov process with transition matrix \mathbb{T} and entries π_{ij} . Then, for whatever x_i is realized in period $t + 1$,

$$d_{t+1} = x_{t+1}Y_t = x_{t+1}c_t = x_i c_t.$$

Under the usual utility specification, $U(c) = \frac{c^{1-\gamma}}{1-\gamma}$, the basic asset pricing equation reduces to

$$\begin{aligned} c_t^{-\gamma} p(Y_t, x_i) &= \delta \sum_{j=1}^N \pi_{ij} (x_j c_t)^{-\gamma} [c_t x_j + p(x_j Y_t, x_j)], \text{ or} \\ p(Y_t, x_i) &= \delta \sum_{j=1}^N \pi_{ij} \left(\frac{x_j c_t}{c_t} \right)^{-\gamma} [c_t x_j + p(x_j Y_t, x_j)]. \end{aligned}$$

So we see that the MRS is determined exclusively by the consumption growth rate.

The essential insight of Mehra and Prescott (1985) was to observe that a solution to this linear system has the form

$$p(Y_t, x_i) = p(c_t, x_i) = v_i c_t$$

for a set of constants $\{v_1, \dots, v_N\}$, each identified with the corresponding growth rate.

With this functional form, the asset pricing equation reduces to

$$\begin{aligned} v_i c_t &= \delta \sum_{j=1}^N \pi_{ij} (x_j)^{-\gamma} [x_j c_t + v_j x_j c_t], \text{ or} \\ v_i &= \delta \sum_{j=1}^N \pi_{ij} (x_j)^{1-\gamma} [1 + v_j]. \end{aligned} \tag{9.35}$$

This is again a system of linear equations in the N unknowns $\{v_1, \dots, v_N\}$. Provided the growth rates are not too large (so that the agent’s utility is not unbounded), a solution exists — a set of $\{v_1^*, \dots, v_N^*\}$ that solves the system of Equations (9.35).

Thus, for any state $(Y, x_j) = (c, x_j)$, the equilibrium equity asset price is

$$p(Y, x_j) = v_j^* Y.$$

If we suppose the current state is (Y, x_i) while next period it is $(x_j Y, x_j)$, then the one-period return earned by the equity security over this period is

$$\begin{aligned} r_{ij} &= \frac{p(x_j Y, x_j) + x_j Y - p(Y, x_i)}{p(Y, x_i)} \\ &= \frac{v_j^* x_j Y + x_j Y - v_i^* Y}{v_i^* Y} \\ &= \frac{x_j (v_j^* + 1)}{v_i^*} - 1, \end{aligned}$$

and the mean or expected return, conditional on state i , is

$$r_i = \sum_{j=1}^N \pi_{ij} r_j.$$

The unconditional equity return is thus given by

$$Er = \sum_{j=1}^N \hat{\pi}_j r_j$$

where $\hat{\pi}_j$ are the long-run stationary probabilities of each state.

The risk-free security is analogously priced as:

$$p^{rf}(c, x_i) = \delta \sum_{j=1}^N \pi_{ij} (x_j)^{-\gamma}, \text{ etc.}$$

Appendix 9.2: Some Properties of the Lognormal Distribution

Definition A9.1:

A variable x is said to follow a lognormal distribution if $\ln x$ is normally distributed. Let $\ln x \sim N(\mu_x, \sigma_x^2)$. If this is the case,

$$\begin{aligned} E(x) &= \exp \left\{ \mu_x + \frac{1}{2} \sigma_x^2 \right\} \\ E(x^a) &= \exp \left\{ a \mu_x + \frac{1}{2} a^2 \sigma_x^2 \right\} \\ \text{var}(x) &= \exp \{ 2 \mu_x + \sigma_x^2 \} (\exp \sigma_x^2 - 1) \end{aligned}$$

Suppose furthermore that x and y are two variables that are independently and identically lognormally distributed; then we also have

$$E(x^a y^b) = \exp \left\{ a \mu_x + b \mu_y + \frac{1}{2} (a^2 \sigma_x^2 + b^2 \sigma_y^2) + 2 \rho a b \sigma_x \sigma_y \right\}$$

where ρ is the correlation coefficient between $\ln x$ and $\ln y$.

Let us apply these relationships to consumption growth: x_t is lognormally distributed, that is, $\ln x_t \sim N(\mu_x, \sigma_x^2)$.

We know that $E(x_t) = 1.0183$ and $\text{var}(x_t) = (.0357)^2$. To identify (μ_x, σ_x^2) , we need to find the solutions of

$$\begin{aligned} 1.0183 &= \exp \left\{ \mu_x + \frac{1}{2} \sigma_x^2 \right\} \\ (.0357)^2 &= \exp \{ 2\mu_x + \sigma_x^2 \} (\exp \sigma_x^2 - 1) \end{aligned}$$

Substituting the first equation squared into the second [by virtue of the fact that $[\exp(y)]^2 = \exp(2y)$] and solving for σ_x^2 , one obtains

$$\sigma_x^2 = .00123.$$

Substituting this value in the equation for μ_x , one solves for $\mu_x = .01752$

We can directly use these values to solve Equation (9.20):

$$E \{ x_t^{-\gamma} \} = \exp \left\{ -\gamma \mu_x + \frac{1}{2} \gamma^2 \sigma_x^2 \right\} = \exp \{ -.03258 \} = .967945,$$

thus $\delta = 1.024$.

Focusing now on the numerator of Equation (9.21), one has:

$$\exp \left\{ \mu_x + \frac{1}{2} \sigma_x^2 \right\} \exp \left\{ -\gamma \mu_x + \frac{1}{2} \gamma^2 \sigma_x^2 \right\},$$

while the denominator is

$$\exp \left\{ (1 - \gamma) \mu_x + \frac{1}{2} (1 - \gamma)^2 \sigma_x^2 \right\}.$$

It remains to recall that $\frac{\exp(a) \exp(b)}{\exp(c)} = \exp(a + b - c)$ to obtain Equation (9.22).

Another application: The standard deviation of the pricing kernel $m_t = x_t^{-\gamma}$ where consumption growth x_t is lognormally distributed. Given that $E m_t$ is as derived in Section 9.6, one estimates

$$\sigma^2(m_t) \cong \frac{1}{k} \left\{ \sum_{i=1}^k [\delta(x_i)^{-\gamma} - E m_t]^2 \right\},$$

for $\ln x_i$ drawn from $N(.01752; .00123)$ and k sufficiently large (say $k = 10,000$). For $\gamma = 2$, one obtains $\sigma^2(m_t) = (.00234)^2$, which yields

$$\frac{\sigma_m}{E \tilde{m}} \cong \frac{.00234}{.9559} = .00245.$$

Part IV

Arbitrage Pricing

Chapter 10: Arrow-Debreu Pricing II: the Arbitrage Perspective

10.1 Introduction

Chapter 8 presented the Arrow-Debreu asset pricing theory from the equilibrium perspective. With the help of a number of modeling hypotheses and building on the concept of market equilibrium, we showed that the price of a future contingent dollar can appropriately be viewed as the product of three main components: a pure time discount factor, the probability of the relevant state of nature, and an intertemporal marginal rate of substitution reflecting the collective (market) assessment of the scarcity of consumption in the future relative to today. This important message is one that we confirmed with the CCAPM of Chapter 9. Here, however, we adopt the alternative arbitrage perspective and revisit the same Arrow-Debreu pricing theory. Doing so is productive precisely because, as we have stressed before, the design of an Arrow-Debreu security is such that once its price is available, whatever its origin and make-up, it provides the answer to the key valuation question: what is a unit of the future state contingent numeraire worth today. As a result, it constitutes the essential piece of information necessary to price arbitrary cash flows. Even if the equilibrium theory of Chapter 8 were all wrong, in the sense that the hypotheses made there turn out to be a very poor description of reality and that, as a consequence, the prices of Arrow-Debreu securities are not well described by equation (8.1), it remains true that if such securities are traded, their prices constitute the essential building blocks (in the sense of our Chapter 2 bicycle pricing analogy) for valuing any arbitrary risky cash-flow.

Section 10.2 develops this message and goes further, arguing that the detour via Arrow-Debreu securities is useful even if no such security is actually traded. In making this argument we extend the definition of the complete market concept. Section 10.3 illustrates the approach in the abstract context of a risk-free world where we argue that any *risk-free* cash flow can be easily and straightforwardly priced as an equivalent portfolio of *date-contingent* claims. These latter instruments are, in effect, discount bonds of various maturities.

Our main interest, of course, is to extend this approach to the evaluation of *risky* cash flows. To do so requires, by analogy, that for each future date-state the corresponding contingent cash flow be priced. This, in turn, requires that we know, for each future *date-state*, the price today of a security that pays off in that date-state and only in that date-state. This latter statement is equivalent to the assumption of market completeness.

In the rest of this chapter, we take on the issue of completeness in the context of securities known as options. Our goal is twofold. First, we want to give the reader an opportunity to review an important element of financial theory – the theory of options. A special appendix to this chapter, available on this text website, describes the essentials for the reader in need of a refresher. Second,

we want to provide a concrete illustration of the view that the recent expansion of derivative markets constitutes a major step in the quest for the “Holy Grail” of achieving a complete securities market structure. We will see, indeed, that options can, in principle, be used relatively straightforwardly to complete the markets. Furthermore, even in situations where this is not practically the case, we can use option pricing theory to value risky cash flows in a manner as though the financial markets were complete. Our discussion will follow the outline suggested by the following two questions.

1. How can options be used to complete the financial markets? We will first answer this question in a simple, highly abstract setting. Our discussion closely follows Ross (1976).

2. What is the link between the prices of market quoted options and the prices of Arrow-Debreu securities? We will see that it is indeed possible to infer Arrow-Debreu prices from option prices in a practical setting conducive to the valuation of an actual cash flow stream. Here our discussion follows Banz and Miller (1978) and Breeden and Litzenberger (1978).

10.2 Market Completeness and Complex Securities

In this section we pursue, more systematically, the important issue of market completeness first addressed when we discussed the optimality property of a general competitive equilibrium. Let us start with two definitions.

1. **Completeness.** Financial markets are said to be *complete* if, for each state of nature θ , there exists a market for contingent claim or Arrow-Debreu security θ ; in other words, for a claim promising delivery of one unit of the consumption good (or, more generally, the numeraire) if state θ is realized, and nothing otherwise. Note that this definition takes a form specifically appropriate to models where there is only one consumption good and several date states. This is the usual context in which financial issues are addressed.

2. **Complex security.** A complex security is one that pays off in more than one state of nature.

Suppose the number of states of nature $N = 4$; an example of a complex security is $S = (5, 2, 0, 6)$ with payoffs 5, 2, 0, and 6, respectively, in states of nature 1, 2, 3, and 4. If markets are complete, we can immediately price such a security since

$$(5, 2, 0, 6) = 5(1, 0, 0, 0) + 2(0, 1, 0, 0) + 0(0, 0, 1, 0) + 6(0, 0, 0, 1),$$

in other words, since the complex security can be replicated by a portfolio of Arrow-Debreu securities, the price of security S , p_S , must be

$$p_S = 5q_1 + 2q_2 + 6q_4.$$

We are appealing here to the law of one price¹ or, equivalently, to a condition of no arbitrage. This is the first instance of our using the second main approach

¹This is stating that the equilibrium prices of two separate units of what is essentially

to asset pricing, the arbitrage approach, that is our exclusive focus in Chapters 10-13. We are pricing the complex security on the basis of our knowledge of the prices of its components. The relevance of the Arrow-Debreu pricing theory resides in the fact that it provides the prices for what can be argued are the essential components of any asset or cash-flow.

Effectively, the argument can be stated in the following proposition.

Proposition 10.1: If markets are complete, any complex security or any cash flow stream can be replicated as a portfolio of Arrow-Debreu securities.

If markets are complete in the sense that prices exist for all the relevant Arrow-Debreu securities, then the "no arbitrage" condition implies that any complex security or cash flow can also be priced using Arrow-Debreu prices as fundamental elements. The portfolio, which is easily priced using the (Arrow-Debreu) prices of its individual components, is essentially the same good as the cash flow or the security it replicates: it pays the same amount of the consumption good in each and every state. Therefore it should bear the same price. This is a key result underlying much of what we do in the remainder of this chapter and our interest in Arrow-Debreu pricing.

If this equivalence is not observed, an arbitrage opportunity - the ability to make unlimited profits with no initial investment - will exist. By taking positions to benefit from the arbitrage opportunity, however, investors will expeditiously eliminate it, thereby forcing the price relationships implicitly asserted in Theorem 10.1. To illustrate how this would work, let us consider the prior example and postulate the following set of prices: $q_1 = \$0.86$, $q_2 = \$0.94$, $q_3 = \$0.93$, $q_4 = \$0.90$, and $q_{(5,2,0,6)} = \$9.80$. At these prices, the law of one price fails, since the price of the portfolio of state claims that exactly replicates the payoff to the complex security does not coincide with the complex's security's price:

$$q_{(5,2,0,6)} = \$9.80 < \$11.58 = 5q_1 + 2q_2 + 6q_4.$$

We see that the complex security is relatively undervalued vis-a-vis the state claim prices. This suggests acquiring a positive amount of the complex security while selling (short) the replicating portfolio of state claims. Table 10.1 illustrates a possible combination.

the same good should be identical. If this were not the case, a riskless and costless arbitrage opportunity would open up: Buy extremely large amounts at the low price and sell them at the high price, forcing the two prices to converge. When applied across two different geographical locations, (which is not the case here: our world is a point in space), the law of one price may not hold because of transport costs rendering the arbitrage costly.

Table 10.1: An arbitrage Portfolio

Security	$t = 0$	$t = 1$ payoffs			
	Cost	θ_1	θ_2	θ_3	θ_4
Buy 1 complex security	-\$9.80	5	2	0	6
Sell short 5 (1,0,0,0) securities	\$4.30	-5	0	0	0
Sell short 2 (0,1,0,0) securities	\$1.88	0	-2	0	0
Sell short 6 (0,0,0,1) securities	\$5.40	0	0	0	-6
Net	\$1.78	0	0	0	0

So the arbitrageur walks away with \$1.78 while (1) having made no investment of his own wealth and (2) without incurring any future obligation (perfectly hedged). She will thus replicate this portfolio as much as she can. But the added demand for the complex security will, *ceteris paribus*, tend to increase its price while the short sales of the state claims will depress their prices. This will continue (arbitrage opportunities exist) so long as the pricing relationships are not in perfect alignment.

Suppose now that only complex securities are traded and that there are M of them (N states). The following is true.

Proposition 10.2: If $M = N$, and all the M complex securities are linearly independent, then (i) it is possible to infer the prices of the Arrow-Debreu state-contingent claims from the complex securities' prices and (ii) markets are effectively complete.²

The hypothesis of linear independence can be interpreted as a requirement that there exist N truly different securities for completeness to be achieved. Thus it is easy to understand that if among the N complex securities available, one security, A , pays (1, 2, 3) in the three relevant states of nature, and the other, B , pays (2, 4, 6), only $N-1$ truly distinct securities are available: B does not permit any different redistribution of purchasing power across states than A permits. More generally, the linear independence hypothesis requires that no one complex security can be replicated as a portfolio of some of the other complex securities. You will remember that we made the same hypothesis at the beginning of Section 6.3.

Suppose the following securities are traded:

$$(3, 2, 0) \quad (1, 1, 1) \quad (2, 0, 2)$$

at equilibrium prices \$1.00, \$0.60, and \$0.80, respectively. It is easy to verify that these three securities are linearly independent. We can then construct the Arrow-Debreu prices as follows. Consider, for example, the security (1, 0, 0):

$$(1, 0, 0) = w_1(3, 2, 0) + w_2(1, 1, 1) + w_3(2, 0, 2)$$

²When we use the language "linearly dependent," we are implicitly regarding securities as N -vectors of payoffs.

$$\begin{aligned}\text{Thus, } 1 &= 3w_1 + w_2 + 2w_3 \\ 0 &= 2w_1 + w_2 \\ 0 &= w_2 + 2w_3\end{aligned}$$

Solve: $w_1 = 1/3, w_2 = -2/3, w_3 = 1/3$, and $q_{(1,0,0)} = 1/3(1.00) + (-2/3)(.60) + 1/3(.80) = .1966$

Similarly, we could replicate $(0, 1, 0)$ and $(0, 0, 1)$ with portfolios $(w_1 = 0, w_2 = 1, w_3 = -1/2)$ and $(w_1 = -1/3, w_2 = 2/3, w_3 = 1/6)$, respectively, and price them accordingly.

Expressed in a more general way, the reasoning just completed amounts to searching for a solution of the following system of equations:

$$\begin{pmatrix} \mathbf{3} & \mathbf{1} & \mathbf{2} \\ \mathbf{2} & \mathbf{1} & \mathbf{0} \\ \mathbf{0} & \mathbf{1} & \mathbf{2} \end{pmatrix} \begin{pmatrix} w_1^1 & w_1^2 & w_1^3 \\ w_2^1 & w_2^2 & w_2^3 \\ w_3^1 & w_3^2 & w_3^3 \end{pmatrix} = \begin{pmatrix} \mathbf{1} & \mathbf{0} & \mathbf{0} \\ \mathbf{0} & \mathbf{1} & \mathbf{0} \\ \mathbf{0} & \mathbf{0} & \mathbf{1} \end{pmatrix}$$

Of course, this system has solution $\begin{pmatrix} \mathbf{3} & \mathbf{1} & \mathbf{2} \\ \mathbf{2} & \mathbf{1} & \mathbf{0} \\ \mathbf{0} & \mathbf{1} & \mathbf{2} \end{pmatrix}^{-1} \begin{pmatrix} \mathbf{1} & \mathbf{0} & \mathbf{0} \\ \mathbf{0} & \mathbf{1} & \mathbf{0} \\ \mathbf{0} & \mathbf{0} & \mathbf{1} \end{pmatrix}$ only if the matrix of security payoffs can be inverted, which requires that it be of full rank, or that its determinant be nonzero, or that all its lines or columns be linearly independent.

Now suppose the number of linearly independent securities is strictly less than the number of states (such as in the final, no-trade, example of Section 8.3 where we assume only a risk-free asset is available). Then the securities markets are fundamentally incomplete: There may be some assets that cannot be accurately priced. Furthermore, risk sharing opportunities are less than if the securities markets were complete and, in general, social welfare is lower than what it would be under complete markets: some gains from exchange cannot be exploited due to the lack of instruments permitting these exchanges to take place.

We conclude this section by revisiting the project valuation problem. How should we, in the light of the Arrow-Debreu pricing approach, value an uncertain cash flow stream such as:

$$\begin{array}{ccccccccc} t = & 0 & 1 & 2 & 3 & \dots & T \\ & -I_0 & \tilde{C}F_1 & \tilde{C}F_2 & \tilde{C}F_3 & \dots & \tilde{C}F_T \end{array}$$

This cash flow stream is akin to a complex security since it pays in multiple states of the world. Let us specifically assume that there are N states at each date $t, t = 1, \dots, T$ and let us denote $q_{t,\theta}$ the price of the Arrow-Debreu security promising delivery of one unit of the numeraire if state θ is realized at date t . Similarly, let us identify as $CF_{t,\theta}$ the cash flow associated with the project

in the same occurrence. Then pricing the complex security à la Arrow-Debreu means valuing the project as in Equation (10.1).

$$NPV = -I_0 + \sum_{t=1}^T \sum_{\theta=1}^N q_{t,\theta} CF_{t,\theta}. \quad (10.1)$$

Although this is a demanding procedure, it is a pricing approach that is fully general and involves no approximation. For this reason it constitutes an extremely useful reference.

In a risk-free setting, the concept of state contingent claim has a very familiar real-world counterpart. In fact, the notion of the term structure is simply a reflection of “date-contingent” claims prices. We pursue this idea in the next section.

10.3 Constructing State Contingent Claims Prices in a Risk-Free World: Deriving the Term Structure

Suppose we are considering risk-free investments and risk-free securities exclusively. In this setting – where we ignore risk – the “states of nature” that we have been speaking of simply correspond to *future time periods*. This section shows that the process of computing the term structure from the prices of coupon bonds is akin to recovering Arrow-Debreu prices from the prices of complex securities.

Under this interpretation, the Arrow-Debreu state contingent claims correspond to risk-free discount bonds of various maturities, as seen in Table 10.2.

Table 10.2: Risk-Free Discount Bonds As Arrow-Debreu Securities

Current Bond Price					Future Cash Flows	
$t = 0$	1	2	3	4	...	T
$-q_1$	\$1,000					
$-q_2$		\$1,000				
...						
$-q_T$						\$1,000

where the cash flow of a “ j -period discount bond” is just

$t = 0$	1	...	j	$j + 1$...	T
$-q_j$	0	0	\$1,000	0	0	0

These are Arrow-Debreu securities because they pay off in one state (the period of maturity), and zero for all other time periods (states).

In the United States at least, securities of this type are not issued for maturities longer than one year. Rather, only interest bearing or coupon bonds are issued for longer maturities. These are complex securities by our definition: They pay off in many states of nature. But we know that if we have enough

distinct complex securities we can compute the prices of the Arrow-Debreu securities even if they are not explicitly traded. So we can also compute the prices of these zero coupon or discount bonds from the prices of the coupon or interest-bearing bonds, assuming no arbitrage opportunities in the bond market.

For example, suppose we wanted to price a 5-year discount bond coming due in November of 2009 (we view $t = 0$ as November 2004), and that we observe two coupon bonds being traded that mature at the same time:

- (i) $7\frac{7}{8}\%$ bond priced at $109\frac{25}{32}$, or $\$1097.8125/\$1,000$ of face value
- (ii) $5\frac{5}{8}\%$ bond priced at $100\frac{9}{32}$, or $\$1002.8125/\$1,000$ of face value

The coupons of these bonds are respectively,

$$.07875 * \$1,000 = \$78.75 / \text{year}$$

$$.05625 * \$1,000 = \$56.25/\text{year}^3$$

The cash flows of these two bonds are seen in Table 10.3.

Table 10.3: Present And Future Cash Flows For Two Coupon Bonds

Bond Type	Cash Flow at Time t					
	$t = 0$	1	2	3	4	5
$7\frac{7}{8}$ bond:	-1,097.8125	78.75	78.75	78.75	78.75	1,078.75
$5\frac{5}{8}$ bond:	-1,002.8125	56.25	56.25	56.25	56.25	1,056.25

Note that we want somehow to eliminate the interest payments (to get a discount bond) and that $\frac{78.75}{56.25} = 1.4$. So, consider the following strategy: Sell one $7\frac{7}{8}\%$ bond while simultaneously buying 1.4 unit of $5\frac{5}{8}\%$ bonds. The corresponding cash flows are found in Table 10.4.

Table 10.4 : Eliminating Intermediate Payments

Bond	Cash Flow at Time t					
	$t = 0$	1	2	3	4	5
-1x $7\frac{7}{8}$ bond:	+1,097.8125	-78.75	-78.75	-78.75	-78.75	-1,078.75
+1.4x $5\frac{5}{8}$ bond:	-1,403.9375	78.75	78.75	78.75	78.75	1,478.75
Difference:	-306.125	0	0	0	0	400.00

The net cash flow associated with this strategy thus indicates that the $t = 0$ price of a \$400 payment in 5 years is \$306.25. This price is implicit in the pricing of our two original coupon bonds. Consequently, the price of \$1,000 in 5 years must be

$$\$306.125 \times \frac{1000}{400} = \$765.3125$$

³In fact interest is paid every 6 months on this sort of bond, a refinement that would double the number of periods without altering the argument in any way.

Alternatively, the price today of \$1.00 in 5 years is \$.7653125. In the notation of our earlier discussion we have the following securities:

$$\begin{array}{c} \theta_1 \\ \theta_2 \\ \theta_3 \\ \theta_4 \\ \theta_5 \end{array} \begin{bmatrix} 78.75 \\ 78.75 \\ 78.75 \\ 78.75 \\ 1078.75 \end{bmatrix} \text{ and } \begin{bmatrix} 56.25 \\ 56.25 \\ 56.25 \\ 56.25 \\ 1056.25 \end{bmatrix} \text{ and we consider}$$

$$-\frac{1}{400} \begin{bmatrix} 78.75 \\ 78.75 \\ 78.75 \\ 78.75 \\ 1078.75 \end{bmatrix} + \frac{1.4}{400} \begin{bmatrix} 56.25 \\ 56.25 \\ 56.25 \\ 56.25 \\ 1056.25 \end{bmatrix} = \begin{bmatrix} 0 \\ 0 \\ 0 \\ 0 \\ 1 \end{bmatrix}.$$

This is an Arrow-Debreu security in the riskless context we are considering in this section.

If there are enough coupon bonds with different maturities with pairs coming due at the same time and with different coupons, we can thus construct a complete set of Arrow-Debreu securities and their implicit prices. Notice that the payoff patterns of the two bonds are fundamentally different: They are linearly independent of one another. This is a requirement, as per our earlier discussion, for being able to use them to construct a fundamentally new payoff pattern, in this case, the discount bond.

Implicit in every discount bond price is a well defined rate of return notion. In the case of the prior illustration, for example, the implied 5-year compound risk free rate is given by

$$\begin{aligned} \$765.3125(1 + r_5)^5 &= \$1000, \text{ or} \\ r_5 &= .078 \end{aligned}$$

This observation suggests an intimate relationship between discounting and Arrow-Debreu date pricing. Just as a full set of date claims prices should allow us to price any risk free cash flow, the rates of return implicit in the Arrow-Debreu prices must allow us to obtain the same price by discounting at the equivalent family of rates. This family of rates is referred to as term structure of interest rates.

Definition: The term structure of interest rates r_1, r_2, \dots is the family of interest rates corresponding to risk free discount bonds of successively greater maturity; i.e., r_i is the rate of return on a risk free discount bond maturing i periods from the present.

We can systematically recover the term structure from coupon bond prices provided we know the prices of coupon bonds of all different maturities. To

illustrate, suppose we observe risk-free government bonds of 1,2,3,4-year maturities all selling at par⁴ with coupons, respectively, of 6%, 6.5%, 7.2%, and 9.5%. We can construct the term structure as follows:

r_1 : Since the 1-year bond sells at par, we have $r_1 = 6\%$;

r_2 : By definition, we know that the two year bond is priced such that

$$1000 = \frac{65}{(1+r_1)} + \frac{1065}{(1+r_2)^2} \text{ which, given that } r_1 = 6\%, \text{ solves for } r_2 = 6.5113\%.$$

r_3 : is derived accordingly as the solution to

$$1000 = \frac{72}{(1+r_1)} + \frac{72}{(1+r_2)^2} + \frac{1072}{(1+r_3)^3}$$

With $r_1 = 6\%$ and $r_2 = 6.5113\%$, the solution is $r_3 = 7.2644\%$. Finally, given these values for r_1 to r_3 , r_4 solves:

$$1000 = \frac{95}{(1+r_1)} + \frac{95}{(1+r_2)^2} + \frac{95}{(1+r_3)^3} + \frac{1095}{(1+r_4)^4}, \text{ i.e., } r_4 = 9.935\%.$$

Note that these rates are the counterpart to the date contingent claim prices.

Table 10.5: Date Claim Prices vs. Discount Bond Prices

	Price of a N year claim	Analogous Discount Bond Price (\$1,000 Denomination)
N = 1	$q_1 = \$1/1.06 = \$.94339$	\$ 943.39
N = 2	$q_2 = \$1/(1.065113)^2 = \$.88147$	\$ 881.47
N = 3	$q_3 = \$1/(1.072644)^3 = \$.81027$	\$ 810.27
N = 4	$q_4 = \$1/1.09935)^4 = \$.68463$	\$ 684.63

Of course, once we have the discount bond prices (the prices of the Arrow-Debreu claims) we can clearly price all other risk-free securities; for example, suppose we wished to price a 4-year 8% bond:

$$\begin{array}{cccccc} t = 0 & 1 & 2 & 3 & 4 \\ -p_0(?) & 80 & 80 & 80 & 1080 \end{array}$$

and suppose also that we had available the discount bonds corresponding to Table 10.5 as in Table 10.6.

Then the portfolio of discount bonds (Arrow-Debreu claims) which replicates the 8% bond cash flow is (Table 10.7):

$$\{.08 \times 1\text{-yr bond}, .08 \times 2\text{-yr bond}, .08 \times 3\text{-yr bond}, 1.08 \times 4\text{-yr bond}\}.$$

⁴That is, selling at their issuing or face value, typically of \$1,000.

Table 10.6: Discount Bonds as Arrow-Debreu Claims

Bond	Price ($t = 0$)	CF Pattern			
		$t = 1$	2	3	4
1-yr discount	-\$943.39	\$1,000			
2-yr discount	-\$881.47		\$1,000		
3-yr discount	-\$810.27			\$1,000	
4-yr discount	-\$684.63				\$1,000

Table 10.7: Replicating the Discount Bond Cash Flow

Bond	Price ($t = 0$)	CF Pattern			
		$t = 1$	2	3	4
08 1-yr discount	$(.08)(-943.39) = -\$75.47$	\$80 (80 state 1 A-D claims)			
08 2-yr discount	$(.08)(-881.47) = -\$70.52$		\$80 (80 state 2 A-D claims)		
08 3-yr discount	$(.08)(-810.27) = -\$64.82$			\$80	
1.08 4-yr discount	$(1.08)(-684.63) = -\$739.40$				\$1,080

Thus:

$$p_{4yr.8\% \text{ bond}} = .08(\$943.39) + .08(\$881.47) + .08(\$810.27) + 1.08(\$684.63) = \$950.21.$$

Notice that we are emphasizing, in effect, the equivalence of the term structure of interest rates with the prices of date contingent claims. Each defines the other. This is especially apparent in Table 10.5.

Let us now extend the above discussion to consider the evaluation of arbitrary risk-free cash flows: any such cash flow can be evaluated as a portfolio of Arrow-Debreu securities; for example:

$t = 0$	1	2	3	4
	60	25	150	300

We want to price this cash flow today ($t = 0$) using the Arrow-Debreu prices we have calculated in Table 10.4.

$$\begin{aligned}
 p &= (\$60 \text{ at } t=1) \left(\frac{\$.94339 \text{ at } t=0}{\$1 \text{ at } t=1} \right) + (\$25 \text{ at } t=2) \left(\frac{\$.88147 \text{ at } t=0}{\$1 \text{ at } t=2} \right) + \dots \\
 &= (\$60) \frac{1.00}{1 + r_1} + (\$25) \frac{1.00}{(1 + r_2)^2} + \dots \\
 &= (\$60) \frac{1.00}{1.06} + (\$25) \frac{1.00}{(1.065113)^2} + \dots
 \end{aligned}$$

The second equality underlines the fact that *evaluating risk-free projects as portfolios of Arrow-Debreu state contingent securities is equivalent to discount-*

ing at the term structure:

$$= \frac{60}{(1+r_1)} + \frac{25}{(1+r_2)^2} + \frac{150}{(1+r_3)^3} + \dots \text{ etc}$$

In effect, we treat a risk-free project as a risk-free coupon bond with (potentially) differing coupons. There is an analogous notion of forward prices and its more familiar counterpart, the forward rate. We discuss this extension in Appendix 10.1.

10.4 The Value Additivity Theorem

In this section we present an important result illustrating the power of the Arrow-Debreu pricing apparatus to generate one of the main lessons of the CAPM. Let there be two assets (complex securities) a and b with date 1 payoffs \tilde{z}_a and \tilde{z}_b , respectively, and let their equilibrium prices be p_a and p_b . Suppose a third asset, c , turns out to be a linear combination of a and b . By that we mean that the payoff to c can be replicated by a portfolio of a and b . One can thus write

$$\tilde{z}_c = A\tilde{z}_a + B\tilde{z}_b, \text{ for some constant coefficients } A \text{ and } B \quad (10.2)$$

Then the proposition known as the **Value Additivity Theorem** asserts that the same linear relationship must hold for the date 0 prices of the three assets:

$$p_c = Ap_a + Bp_b.$$

Let us first prove this result and then discuss its implications. The proof easily follows from our discussion in Section 10.2 on the pricing of complex securities in a complete market Arrow-Debreu world. Indeed, for our 2 securities a, b , one must have:

$$p_i = \sum_s q_s z_{si}, \quad i = a, b \quad (10.3)$$

where q_s is the price of an Arrow-Debreu security that pays one unit of consumption in state s (and zero otherwise) and z_{si} is the payoff of asset i in state s

But then, the pricing of c must respect the following relationships:

$$p_c = \sum_s q_s z_{sc} = \sum_s q_s (Az_{sa} + Bz_{sb}) = \sum_s (Aq_s z_{sa} + Bq_s z_{sb}) = Ap_a + Bp_b$$

The first equality follows from the fact that c is itself a complex security and can thus be priced using Arrow-Debreu prices [i.e., an equation such as Equation (10.3) applies]; the second directly follows from Equation (10.2); the third is a pure algebraic expansion that is feasible because our pricing relationships are fundamentally linear; and the fourth follows from Equation (10.3) again.

Now this is easy enough. Why is it interesting? Think of a and b as being two stocks with negatively correlated returns; we know that c , a portfolio of these two stocks, is much less risky than either one of them. But p_c is a linear combination of p_a and p_b . Thus, the fact that they can be combined in a less risky portfolio has implications for the pricing of the two independently riskier securities and their equilibrium returns. Specifically, it cannot be the case that p_c would be *high* because it corresponds to a desirable, riskless, claim while the p_a and p_b would be *low* because they are risky.

To see this more clearly, let us take an extreme example. Suppose that a and b are *perfectly* negatively correlated. For an appropriate choice of A and B , say A^* and B^* , the resulting portfolio, call it d , will have zero risk; i.e., it will pay a constant amount in each and every state of nature. What should the price of this riskless portfolio be? Intuitively, its price must be such that purchasing d at p_d will earn the riskless rate of return. But how could the risk of a and b be remunerated while, simultaneously, d would earn the riskless rate and the value additivity theorem hold? The answer is that this is not possible. Therefore, there cannot be any remuneration for risk in the pricing of a and b . The prices p_a and p_b must be such that the expected return on a and b is the riskless rate. This is true despite the fact that a and b are two risky assets (they do not pay the same amount in each state of nature).

In formal terms, we have just asserted that the two terms of the Value Additivity Theorem $\tilde{z}_d = A^* \tilde{z}_a + B^* \tilde{z}_b$ and $p_d = A^* p_a + B^* p_b$, together with the fact that d is risk-free,

$$\begin{aligned} \frac{E \tilde{z}_d}{p_d} &= 1 + r_f, \text{ force} \\ \frac{E \tilde{z}_a}{p_a} &= \frac{E \tilde{z}_b}{p_b} = 1 + r_f. \end{aligned}$$

What we have obtained in this very general context is a confirmation of one of the main results of the CAPM: Diversifiable risk is not priced. If risky assets a and b can be combined in a riskless portfolio, that is, if their risk can be diversified away, their return cannot exceed the risk-free return. Note that we have made no assumption here on utility functions nor on the return expectations held by agents. On the other hand we have explicitly assumed that markets are complete and that consequently each and every complex security can be priced (by arbitrage) as a portfolio of Arrow-Debreu securities.

It thus behooves us to describe how Arrow-Debreu state claim prices might actually be obtained in practice. This is the subject of the remaining sections to Chapter 10.

10.5 Using Options to Complete the Market: An Abstract Setting

Let us assume a finite number of possible future date-states indexed $i = 1, 2, \dots, N$. Suppose, for a start, that three states of the world are possible in date $T = 1$,

yet only one security (a stock) is traded. The single security's payoffs are as follows:

State	Payoff
θ_1	$\begin{bmatrix} 1 \\ 2 \\ 3 \end{bmatrix}$.
θ_2	
θ_3	

Clearly this unique asset is not equivalent to a complete set of state-contingent claims. Note that we can identify the payoffs with the ex-post price of the security in each of the 3 states: the security pays 2 units of the numeraire commodity in state 2 and we decide that its price then is \$2.00. This amounts to normalizing the ex post, date 1, price of the commodity to \$1, much as we have done at date 0. On that basis, we can consider call options written on this asset with exercise prices \$1 and \$2, respectively. These securities are contracts giving the right (but not the obligation) to purchase the underlying security at prices \$1 and \$2, respectively, tomorrow. They are contingent securities in the sense that the right they entail is valuable only if the price of the underlying security exceeds the exercise price at expiration, and they are valueless otherwise. We think of the option expiring at $T = 1$, that is, when the state of nature is revealed.⁵ The *states of nature* structure enables us to be specific regarding what these contracts effectively promise to pay. Take the call option with exercise price \$1. If state 1 is realized, that option is a right to buy at \$1 the underlying security whose value is exactly \$1. The option is said to be *at the money* and, in this case, the right in question is valueless. If state 2 is realized, however, the stock is worth \$2. The right to buy, at a price of \$1, something one can immediately resell for \$2 naturally has a market value of \$1. In this case, the option is said to be *in the money*. In other words, at $T = 1$, when the state of nature is revealed, an option is worth the difference between the value of the underlying asset and its exercise price, if this difference is positive, and zero otherwise. The complete payoff vectors of these options at expiration are as follows:

$$C_T([1, 2, 3]; 1) = \begin{bmatrix} 0 \\ 1 \\ 2 \end{bmatrix} \begin{matrix} \theta_1 \\ \theta_2 \\ \theta_3 \end{matrix} \left\{ \begin{array}{l} \text{at the money} \\ \text{in the money} \\ \text{in the money} \end{array} \right\}.$$

Similarly,

$$C_T([1, 2, 3]; 2) = \begin{bmatrix} 0 \\ 0 \\ 1 \end{bmatrix} \begin{matrix} \theta_1 \\ \theta_2 \\ \theta_3 \end{matrix}$$

⁵In our simple two-date world there is no difference between an American option, which can be exercised at any date before the expiration date, and a European option, which can be exercised only at expiration.

In our notation, $C_T(S; K)$ is the payoff to a call option written on security S with exercise price K at expiration date T . We use $C_t(S; K)$ to denote the option's market price at time $t \leq T$. We frequently drop the time subscript to simplify notation when there is no ambiguity.

It remains now to convince ourselves that the three traded assets (the underlying stock and the two call options, each denoted by its payoff vector at T)

$$\begin{matrix} \theta_1 \\ \theta_2 \\ \theta_3 \end{matrix} \begin{bmatrix} 1 \\ 2 \\ 3 \end{bmatrix}, \begin{bmatrix} 0 \\ 1 \\ 2 \end{bmatrix}, \begin{bmatrix} 0 \\ 0 \\ 1 \end{bmatrix}$$

constitute a complete set of securities markets for states $(\theta_1, \theta_2, \theta_3)$. This is so because we can use them to create all the state claims. Clearly $\begin{bmatrix} 0 \\ 0 \\ 1 \end{bmatrix}$ is present.

To create $\begin{bmatrix} 0 \\ 1 \\ 0 \end{bmatrix}$, observe that

$$\begin{bmatrix} 0 \\ 1 \\ 0 \end{bmatrix} = w_1 \begin{bmatrix} 1 \\ 2 \\ 3 \end{bmatrix} + w_2 \begin{bmatrix} 0 \\ 1 \\ 2 \end{bmatrix} + w_3 \begin{bmatrix} 0 \\ 0 \\ 1 \end{bmatrix},$$

where $w_1 = 0$, $w_2 = 1$, and $w_3 = -2$.

The vector $\begin{bmatrix} 1 \\ 0 \\ 0 \end{bmatrix}$ can be similarly created.

We have thus illustrated one of the main ideas of this chapter, and we need to discuss how general and applicable it is in more realistic settings. A preliminary issue is why trading call option securities $C([1,2,3];1)$ and $C([1,2,3];2)$ might be the preferred approach to completing the market, relative to the alternative possibility of directly issuing the Arrow-Debreu securities $[1,0,0]$ and $[0,1,0]$? In the simplified world of our example, in the absence of transactions costs, there is, of course, no advantage to creating the options markets. In the real world, however, if a new security is to be issued, its issuance must be accompanied by costly disclosure as to its characteristics; in our parlance, the issuer must disclose as much as possible about the security's payoff in the various states. As there may be no agreement as to what the relevant future states are – let alone what the payoffs will be – this disclosure is difficult. And if there is no consensus as to its payoff pattern, (i.e., its basic structure of payoffs), investors will not want to hold it, and it will not trade. But the payoff pattern of an option on an already-traded asset is obvious and verifiable to everyone. For this reason, it is, in principle, a much less expensive new security to issue. Another way to describe the advantage of options is to observe that it is useful conceptually, but difficult in practice, to define and identify a single state of nature. It is more practical to define contracts contingent on a well-defined *range* of states. The fact that these states are themselves defined in terms of, or revealed via, market prices is another facet of the superiority of this type of contract.

Note that options are by definition in zero net supply, that is, in this context

$$\sum_k C_t^k([1, 2, 3]; K) = 0$$

where $C_t^k([1, 2, 3]; K)$ is the value of call options with exercise price K , held by agent k at time $t \leq T$. This means that there must exist a group of agents with negative positions serving as the counter-party to the subset of agents with positive holdings. We naturally interpret those agents as agents who have *written* the call options.

We have illustrated the property that markets can be completed using call options. Now let us explore the generality of this result. Can call options always be used to complete the market in this way? The answer is not necessarily. It depends on the payoff to the underlying fundamental assets. Consider the asset:

$$\begin{matrix} \theta_1 \\ \theta_2 \\ \theta_3 \end{matrix} \begin{bmatrix} 2 \\ 2 \\ 3 \end{bmatrix}.$$

For any exercise price K , all options written on this security must have payoffs of the form:

$$C([2, 2, 3]; K) = \begin{cases} \begin{bmatrix} 2 - K \\ 2 - K \\ 3 - K \end{bmatrix} & \text{if } K \leq 2 \\ \begin{bmatrix} 0 \\ 0 \\ 3 - K \end{bmatrix} & \text{if } 2 < K \leq 3 \end{cases}$$

Clearly, for any K ,

$$\begin{bmatrix} 2 \\ 2 \\ 3 \end{bmatrix} \text{ and } \begin{bmatrix} 2 - K \\ 2 - K \\ 3 - K \end{bmatrix}$$

have identical payoffs in state θ_1 and θ_2 , and, therefore, they cannot be used to generate Arrow-Debreu securities

$$\begin{bmatrix} 1 \\ 0 \\ 0 \end{bmatrix} \text{ and } \begin{bmatrix} 0 \\ 1 \\ 0 \end{bmatrix}.$$

There is no way to complete the markets with options in the case of this underlying asset. This illustrates the following truth: We cannot generally write options that distinguish between two states if the underlying assets pay identical returns in those states.

The problem just illustrated can sometimes be solved *if we permit options to be written on **portfolios** of the basic underlying assets*. Consider the case of

four possible states at $T = 1$, and suppose that the only assets currently traded are

$$\begin{matrix} \theta_1 \\ \theta_2 \\ \theta_3 \\ \theta_4 \end{matrix} \begin{bmatrix} 1 \\ 1 \\ 2 \\ 2 \end{bmatrix} \quad \text{and} \quad \begin{bmatrix} 1 \\ 2 \\ 1 \\ 2 \end{bmatrix}.$$

It can be shown that it is not possible, using call options, to generate a complete set of securities markets using only these underlying securities. Consider, however, the portfolio composed of 2 units of the first asset and 1 unit of the second:

$$2 \begin{bmatrix} 1 \\ 1 \\ 2 \\ 2 \end{bmatrix} + 1 \begin{bmatrix} 1 \\ 2 \\ 1 \\ 2 \end{bmatrix} = \begin{bmatrix} 3 \\ 4 \\ 5 \\ 6 \end{bmatrix}.$$

The portfolio pays a different return in each state of nature. Options written on the portfolio alone can thus be used to construct a complete set of traded Arrow-Debreu securities. The example illustrates a second general truth, which we will enumerate as Proposition 10.3.

Proposition 10.3: A necessary as well as sufficient condition for the creation of a complete set of Arrow-Debreu securities is that there exists a single portfolio with the property that options can be written on it and such that its payoff pattern distinguishes among all states of nature.

Going back to our last example, it is easy to see that the created portfolio and the three natural calls to be written on it:

$$\begin{bmatrix} 3 \\ 4 \\ 5 \\ 6 \end{bmatrix} \quad \text{plus} \quad \begin{bmatrix} 0 \\ 1 \\ 2 \\ 3 \end{bmatrix}_{(K=3)} \quad \text{and} \quad \begin{bmatrix} 0 \\ 0 \\ 1 \\ 2 \end{bmatrix}_{(K=4)} \quad \text{and} \quad \begin{bmatrix} 0 \\ 0 \\ 0 \\ 1 \end{bmatrix}_{(K=5)}$$

are sufficient, (i.e., constitute a complete set of markets in our four-state world). Combinations of the $(K = 5)$ and $(K = 4)$ vectors can create:

$$\begin{bmatrix} 0 \\ 0 \\ 1 \\ 0 \end{bmatrix}.$$

Combinations of this vector, and the $(K = 5)$ and $(K = 3)$ vectors can then create:

$$\begin{bmatrix} 0 \\ 1 \\ 0 \\ 0 \end{bmatrix}, \text{ etc.}$$

Probing further we may inquire if *the writing of calls on the underlying assets is always sufficient*, or whether there are circumstances under which other types of options may be necessary. Again, suppose there are four states of nature, and consider the following set of *primitive* securities:

$$\begin{matrix} \theta_1 \\ \theta_2 \\ \theta_3 \\ \theta_4 \end{matrix} \begin{bmatrix} 0 \\ 0 \\ 0 \\ 1 \end{bmatrix} \begin{bmatrix} 1 \\ 1 \\ 0 \\ 1 \end{bmatrix} \begin{bmatrix} 0 \\ 1 \\ 1 \\ 1 \end{bmatrix}.$$

Because these assets pay either one or zero in each state, calls written on them will either replicate the asset itself, or give the zero payoff vector. The writing of call options will not help because they cannot further discriminate among states. But suppose we write a put option on the first asset with exercise price 1. A put is a contract giving the right, but not the obligation, to *sell* an underlying security at a pre-specified exercise price on a given expiration date. The put option with exercise price 1 has positive value at $T = 1$ in those states where the underlying security has value less than 1. The put on the first asset with exercise price = \$1 thus has the following payoff:

$$\begin{bmatrix} 1 \\ 1 \\ 1 \\ 0 \end{bmatrix} = P_T([0, 0, 0, 1]; 1).$$

You can confirm that the securities plus the put are sufficient to allow us to construct (as portfolios of them) a complete set of Arrow-Debreu securities for the indicated four states. In general, one can prove Proposition 10.4.

Proposition 10.4:

If it is possible to create, using options, a complete set of traded securities, simple put and call options written on the underlying assets are sufficient to accomplish this goal.

That is, portfolios of options are not required.

10.6 Synthesizing State-Contingent Claims: A First Approximation

The abstract setting of the discussion above aimed at conveying the message that options are natural instruments for completing the markets. In this section, we show how we can directly create a set of state-contingent claims, *as well as their equilibrium prices*, using option prices or option pricing formulae in a more realistic setting. The interest in doing so is, of course, to exploit the possibility, inherent in Arrow-Debreu prices, of pricing any complex security. In

this section we first approach the problem under the hypothesis that the price of the underlying security or portfolio can take only discrete values.

Assume that a risky asset is traded with current price S and future price S_T . It is assumed that S_T discriminates across all states of nature so that Proposition 10.3 applies; without loss of generality, we may assume that S_T takes the following set of values:

$$S_1 < S_2 < \dots < S_{\hat{\theta}} < \dots < S_N,$$

where S_{θ} is the price of this complex security if state θ is realized at date T . Assume also that call options are written on this asset with all possible exercise prices, and that these options are traded. Let us also assume that $S_{\theta} = S_{\theta-1} + \delta$ for every state θ . (This is not so unreasonable as stocks, say, are traded at prices that can differ only in multiples of a minimum price change).⁶ Throughout the discussion we will fix the time to expiration and will not denote it notationally.

Consider, for any state $\hat{\theta}$, the following portfolio P :

$$\begin{aligned} &\text{Buy one call with } K = S_{\hat{\theta}-1} \\ &\text{Sell two calls with } K = S_{\hat{\theta}} \\ &\text{Buy one call with } K = S_{\hat{\theta}+1} \end{aligned}$$

At any point in time, the value of this portfolio, V_P , is

$$V_P = C(S, K = S_{\hat{\theta}-1}) - 2C(S, K = S_{\hat{\theta}}) + C(S, K = S_{\hat{\theta}+1}).$$

To see what this portfolio represents, let us examine its payoff *at expiration* (refer to Figure 10.1):

Insert Figure 10.1 about here

For $S_T \leq S_{\hat{\theta}-1}$, the value of our options portfolio, P , is zero. A similar situation exists for $S_T \geq S_{\hat{\theta}+1}$ since the loss on the 2 written calls with $K = S_{\hat{\theta}}$ exactly offsets the gains on the other two calls. In state $\hat{\theta}$, the value of the portfolio is δ corresponding to the value of $C_T(S_{\hat{\theta}}, K = S_{\hat{\theta}-1})$, the other two options being out of the money when the underlying security takes value $S_{\hat{\theta}}$. The payoff from such a portfolio thus equals:

$$\text{Payoff to } P = \begin{cases} 0 & \text{if } S_T < S_{\hat{\theta}} \\ \delta & \text{if } S_T = S_{\hat{\theta}} \\ 0 & \text{if } S_T > S_{\hat{\theta}} \end{cases}$$

in other words, it pays a positive amount δ in state $\hat{\theta}$, and nothing otherwise. That is, it replicates the payoff of the Arrow-Debreu security associated with state $\hat{\theta}$ up to a factor (in the sense that it pays δ instead of 1). Consequently,

⁶Until recently, the minimum price change was equal to \$1/16 on the NYSE. At the end of 2000, *decimal pricing* was introduced whereby the prices are quoted to the nearest \$1/100 (1 cent).

the current price of the state $\hat{\theta}$ contingent claim (i.e., one that pays \$1.00 if state $\hat{\theta}$ is realized and nothing otherwise) must be

$$q_{\hat{\theta}} = \frac{1}{\delta} [C(S, K = S_{\hat{\theta}-1}) + C(S, K = S_{\hat{\theta}+1}) - 2C(S, K = S_{\hat{\theta}})].$$

Even if these calls are not traded, if we identify our relevant states with the prices of some security – say the market portfolio – then we can use readily available option pricing formulas (such as the famous Black & Scholes formula) to obtain the necessary call prices and, from them, compute the price of the state-contingent claim. We explore this idea further in the next section.

10.7 Recovering Arrow-Debreu Prices From Options Prices: A Generalization

By the CAPM, the only relevant risk is systematic risk. We may interpret this to mean that the only states of nature that are economically or financially relevant are those that can be identified with different values of the market portfolio.⁷ The market portfolio thus may be selected to be the complex security on which we write options, portfolios of which will be used to replicate state-contingent payoffs. The conditions of Proposition 10.1 are satisfied, guaranteeing the possibility of completing the market structure.

In Section 10.6, we considered the case for which the underlying asset assumed a discrete set of values. If the underlying asset is the market portfolio M , however, this cannot be strictly valid: As an index it can essentially assume an infinite number of possible values. How is this added feature accommodated?

1. Suppose that S_T , the price of the underlying portfolio (we may think of it as a proxy for M), assumes a *continuum* of possible values. We want to price an Arrow-Debreu security that pays \$1.00 if $\tilde{S}_T \in \left[-\frac{\delta}{2} + \hat{S}_T, \hat{S}_T + \frac{\delta}{2}\right]$, in other words, if S_T assumes any value in a range of width δ , centered on \hat{S}_T . We are thus identifying our states of nature with **ranges of possible values** for the market portfolio. Here the subscript T refers to the future date at which the Arrow-Debreu security pays \$1.00 if the relevant state is realized.

2. Let us construct the following portfolio⁸ for some small positive number $\varepsilon > 0$,

$$\begin{aligned} &\text{Buy one call with } K = \hat{S}_T - \frac{\delta}{2} - \varepsilon \\ &\text{Sell one call with } K = \hat{S}_T - \frac{\delta}{2} \\ &\text{Sell one call with } K = \hat{S}_T + \frac{\delta}{2} \\ &\text{Buy one call with } K = \hat{S}_T + \frac{\delta}{2} + \varepsilon. \end{aligned}$$

⁷That is, diversifiable risks have zero market value (see Chapter 7 and Section 10.4). At an individual level, personal risks are, of course, also relevant. They can, however, be insured or diversified away. Insurance contracts are often the most appropriate to cover these risks. Recall our discussion of this issue in Chapter 1.

⁸The option position corresponding to this portfolio is known as a *butterfly spread* in the jargon.

Figure 10.2 depicts what this portfolio pays *at expiration*.

Insert Figure 10.2 about here

Observe that our portfolio pays ε on a range of states and 0 almost everywhere else. By purchasing $1/\varepsilon$ units of the portfolio, we will mimic the payoff of an Arrow-Debreu security, except for the two small diagonal sections of the payoff line where the portfolio pays something between 0 and ε . This undesirable feature (since our objective is to replicate an Arrow-Debreu security) will be taken care of by using a standard mathematical trick involving taking limits.

3. Let us thus consider buying $1/\varepsilon$ units of the portfolio. The total payment, when $\hat{S}_T - \frac{\delta}{2} \leq S_T \leq \hat{S}_T + \frac{\delta}{2}$, is $\varepsilon \cdot \frac{1}{\varepsilon} \equiv 1$, for any choice of ε . We want to let $\varepsilon \mapsto 0$, so as to eliminate payments in the ranges $S_T \in [\hat{S}_T - \frac{\delta}{2} - \varepsilon, \hat{S}_T - \frac{\delta}{2})$ and $S_T \in (\hat{S}_T + \frac{\delta}{2}, \hat{S}_T + \frac{\delta}{2} + \varepsilon]$. The value of $1/\varepsilon$ units of this portfolio is:

$$\frac{1}{\varepsilon} \left\{ C(S, K = \hat{S}_T - \frac{\delta}{2} - \varepsilon) - C(S, K = \hat{S}_T - \frac{\delta}{2}) \right. \\ \left. - [C(S, K = \hat{S}_T + \frac{\delta}{2}) - C(S, K = \hat{S}_T + \frac{\delta}{2} + \varepsilon)] \right\},$$

where a minus sign indicates that the call was sold (thereby reducing the cost of the portfolio by its sale price). On balance the portfolio will have a positive price as it represents a claim on a positive cash flow in certain states of nature. Let us assume that the pricing function for a call with respect to changes in the exercise price can be differentiated (this property is true, in particular, in the case of the Black & Scholes option pricing formula). We then have:

$$\begin{aligned} & \lim_{\varepsilon \mapsto 0} \frac{1}{\varepsilon} \left\{ C(S, K = \hat{S}_T - \frac{\delta}{2} - \varepsilon) - C(S, K = \hat{S}_T - \frac{\delta}{2}) \right. \\ & \quad \left. - [C(S, K = \hat{S}_T + \frac{\delta}{2}) - C(S, K = \hat{S}_T + \frac{\delta}{2} + \varepsilon)] \right\} \\ &= - \lim_{\varepsilon \mapsto 0} \left\{ \underbrace{C(S, K = \hat{S}_T - \frac{\delta}{2} - \varepsilon) - C(S, K = \hat{S}_T - \frac{\delta}{2})}_{\substack{-\varepsilon \\ \leq 0}} \right\} \\ & \quad + \lim_{\varepsilon \mapsto 0} \left\{ \underbrace{C(S, K = \hat{S}_T + \frac{\delta}{2} + \varepsilon) - C(S, K = \hat{S}_T + \frac{\delta}{2})}_{\substack{\varepsilon \\ \leq 0}} \right\} \\ &= C_2 \left(S, K = \hat{S}_T + \frac{\delta}{2} \right) - C_2 \left(S, K = \hat{S}_T - \frac{\delta}{2} \right). \end{aligned}$$

Here the subscript 2 indicates the partial derivative with respect to the second argument (K), evaluated at the indicated exercise prices. In summary, the limiting portfolio has a payoff at expiration as represented in Figure 10.3

Insert Figure 10.3 about here

and a (current) price $C_2 \left(S, K = \hat{S}_T + \frac{\delta}{2} \right) - C_2 \left(S, K = \hat{S}_T - \frac{\delta}{2} \right)$ that is positive since the payoff is positive. We have thus priced an Arrow-Debreu state-contingent claim one period ahead, given that we define states of the world as coincident with ranges of a proxy for the market portfolio.

4. Suppose, for example, we have an uncertain payment with the following payoff at time T :

$$CF_T = \begin{cases} 0 & \text{if } S_T \notin [\hat{S}_T - \frac{\delta}{2}, \hat{S}_T + \frac{\delta}{2}] \\ 50000 & \text{if } S_T \in [\hat{S}_T - \frac{\delta}{2}, \hat{S}_T + \frac{\delta}{2}] \end{cases}.$$

The value today of this cash flow is:

$$50,000 \cdot \left[C_2 \left(S, K = \hat{S}_T + \frac{\delta}{2} \right) - C_2 \left(S, K = \hat{S}_T - \frac{\delta}{2} \right) \right].$$

The formula we have developed is really very general. In particular, for any arbitrary values S_T^1 and S_T^2 , the price of an Arrow-Debreu contingent claim that pays off \$1.00 if the underlying market portfolio assumes a value $S_T \in [S_T^1, S_T^2]$, is given by

$$q(S_T^1, S_T^2) = C_2(S, K = S_T^2) - C_2(S, K = S_T^1). \quad (10.4)$$

We value this quantity in Box 10.1 for a particular set of parameters making explicit use of the Black-Scholes option pricing formula.

Box 10.1: Pricing A-D Securities with Black-Scholes

For calls priced according to the Black-Scholes option pricing formula, Breen and Litzenberger (1978) prove that

$$\begin{aligned} AD(S_T^1, S_T^2) &= C_2(S, K = S_T^2) - C_2(S, K = S_T^1) \\ &= e^{-rT} \{ N(d_2(S_T^1)) - N(d_2(S_T^2)) \} \end{aligned}$$

where

$$d_2(S_T^i) = \frac{\left[\ln \left(\frac{S_0}{S_T^i} \right) + \left(r_f - \delta - \frac{\sigma^2}{2} \right) T \right]}{\sigma \sqrt{T}}$$

In this expression, T is the time to expiration, r_f the annualized continuously compounded riskless rate over that period, δ the continuous annualized portfolio dividend yield, σ the standard deviation of the continuously compounded rate of return on the underlying index portfolio, $N(\cdot)$ the standard normal distribution, and S_0 the current value of the index.

Suppose the not-continuously-compounded risk-free rate is .06, the not-continuously compounded dividend yield is $\delta = .02$, $T = 5$ years, $S_0 = 1,500$, $S_T^2 = 1,700$, $S_T^1 = 1,600$, $\sigma = .20$; then

$$\begin{aligned}
d_2(S_T^1) &= \frac{\left\{ \ln\left(\frac{1500}{1600}\right) + \left[\ln(1.06) - \ln(1.02) - \frac{(.20)^2}{2} \right] (.5) \right\}}{.20\sqrt{.5}} \\
&= \frac{\{-.0645 + (.0583 - .0198 - .02)(.5)\}}{.1414} \\
&= -.391 \\
d_2(S_T^2) &= \frac{\left\{ \ln\left(\frac{1500}{1700}\right) + (.0583 - .0198 - .02)(.5) \right\}}{.1414} \\
&= \frac{\{-.1252 + .00925\}}{.1414} \\
&= -.820 \\
AD(S_T^1, S_T^2) &= e^{-\ln(1.06)(.5)} \{N(-.391) - N(-.820)\} \\
&= .9713 \{.2939 - .1517\} \\
&= .1381,
\end{aligned}$$

or about \$.14. \square

Suppose we wished to price an uncertain cash flow to be received in one period from now, where a period corresponds to a duration of time T . What do we do? Choose several ranges of the value of the market portfolio corresponding to the various states of nature that may occur – say three states: “recession,” “slow growth,” and “boom” and estimate the cash flow in each of these states (see Figure 10.4). It would be unusual to have a large number of states as the requirement of having to estimate the cash flows in each of those states is likely to exceed our forecasting abilities.

Insert Figure 10.4 about here

Suppose the cash flow estimates are, respectively, CF_B, CF_{SG}, CF_R , where the subscripts denote, respectively, “boom,” “slow growth,” and “recession.” Then,

$$\text{Value of the } CF = V_{CF} = q(S_T^3, S_T^4) CF_B + q(S_T^2, S_T^3) CF_{SG} + q(S_T^1, S_T^2) CF_R,$$

where $S_T^1 < S_T^2 < S_T^3 < S_T^4$, and the Arrow-Debreu prices are estimated from option prices or option pricing formulas according to Equation (10.4).

We can go one (final) step further if we assume for a moment that the cash flow we wish to value can be described by a continuous function of the value of the market portfolio.

In principle, for a very fine partition of the range of possible values of the market portfolio, say $\{S_1, \dots, S_N\}$, where $S_i < S_{i+1}$, $S_N = \max S_T$, and $S_1 = \min$

S_T , we could price the Arrow-Debreu securities that pay off in each of these $N-1$ states defined by the partition:

$$\begin{aligned} q(S_1, S_2) &= C_2(S, S_2) - C_2(S, S_1) \\ q(S_2, S_3) &= C_2(S, S_3) - C_2(S, S_2) \dots \text{etc.} \end{aligned}$$

Simultaneously, we could approximate a cash flow function $CF(S_T)$ by a function that is constant in each of these ranges of S_T (a so-called “step function”), in other words, $\hat{CF}(S_T) = CF_i$, for $S_{i-1} \leq S_T \leq S_i$. For example,

$$\hat{CF}(S_T) = CF_i = \frac{CF(S, S_T = S_i) + CF(S, S_T = S_{i-1})}{2} \text{ for } S_{i-1} \leq S_T \leq S_i$$

This particular approximation is represented in Figure 10.5. The value of the approximate cash flow would then be

$$\begin{aligned} V_{CF} &= \sum_{i=1}^N \hat{CF}_i \cdot q(S_{i-1}, S_i) \\ &= \sum_{i=1}^N \hat{CF}_i [C_2(S, S_T = S_i) - C_2(S, S_T = S_{i-1})] \end{aligned} \quad (10.5)$$

Insert Figure 10.5 about here

Our approach is now clear. The precise value of the uncertain cash flow will be the sum of the approximate cash flows evaluated at the Arrow-Debreu prices as the norm of the partition (the size of the interval $S_i - S_{i-1}$) tends to zero. It can be shown (and it is intuitively plausible) that the limit of Equation (10.5) as $\max_i |S_{i+1} - S_i| \mapsto 0$ is the integral of the cash flow function multiplied by the second derivative of the call's price with respect to the exercise price. The latter is the infinitesimal counterpart to the difference in the first derivatives of the call prices entering in Equation (10.4).

$$\begin{aligned} &\lim_{\max_i |S_{i+1} - S_i| \mapsto 0} \sum_{i=1}^N \hat{CF}_i [C_2(S, S_T = S_{i+1}) - C_2(S, S_T = S_i)] \quad (10.6) \\ &= \int CF(S_T) C_{22}(S, S_T) dS_T. \end{aligned}$$

As a particular case of a constant cash flow stream, a risk-free bond paying \$1.00 in every state is then priced as per

$$p^{rf} = \frac{1}{(1 + r_f)} = \int_0^\infty C_{22}(S, S_T) dS_T.$$

Box 10.2: Extracting Arrow-Debreu Prices from Option Prices: A Numerical Illustration

Let us now illustrate the power of the approach adopted in this and the previous section. For that purpose, Table 10.8 [adapted from Pirkner, Weigend, and Zimmermann (1999)] starts by recording call prices, obtained from the Black-Scholes formula for a call option, on an underlying index portfolio, currently valued at $S = 10$, for a range of strike prices going from $K = 7$ to $K = 13$ (columns 1 and 2). Column 3 computes the value of portfolio P of Section 10.6. Given that the difference between the exercise prices is always 1 (i.e., $\delta = 1$), holding exactly one unit of this portfolio replicates the \$1.00 payoff of the Arrow-Debreu security associated with $K = 10$. This is shown on the bottom line of column 7, which corresponds to $S = 10$. From column 3, we learn that the price of this Arrow-Debreu security, which must be equal to the value of the replicating portfolio, is \$0.184. Finally, the last two columns approximate the first and second derivatives of the call price with respect to the exercise price. In the current context this is naturally done by computing the first and second differences (the price increments and the increments of the increments as the exercise price varies) from the price data given in column 2. This is a literal application of Equation (10.4). One thus obtains the full series of Arrow-Debreu prices for states of nature identified with values of the underlying market portfolios ranging from 8 to 12, confirming that the \$0.184 price occurs when the state of nature is identified as $S = 10$ (or $9.5 < S < 10.5$).

Table 10.8: Pricing an Arrow-Debreu State Claim

K	$C(S, K)$	Cost of Position	Payoff if $S_T =$							ΔC	$\Delta(\Delta C) = q_\theta$
			7	8	9	10	11	12	13		
7	3.354										
										-0.895	
8	2.459									-0.789	0.106
9	1.670	+1.670	0	0	0	1	2	3	4	-0.625	0.164
10	1.045	-2.090	0	0	0	0	-2	-4	-6	-0.441	0.184
11	0.604	+0.604	0	0	0	0	0	1	2	-0.279	0.162
12	0.325									-0.161	0.118
13	0.164										
		0.184	0	0	0	1	0	0	0		

10.8 Arrow-Debreu Pricing in a Multiperiod Setting

The fact that the Arrow-Debreu pricing approach is static makes it most adequate for the pricing of one-period cash flows and it is, quite naturally, in this context that most of our discussion has been framed. But as we have emphasized previously, it is formally equally appropriate for pricing multiperiod cash flows. The estimation (for instance via option pricing formulas and the methodology introduced in the last two sections) of Arrow-Debreu prices for several periods ahead is inherently more difficult, however, and relies on more perilous assumptions than in the case of one period ahead prices. (This parallels the fact that the assumptions necessary to develop closed form option pricing formulae are more questionable when they are used in the context of pricing long-term options). Pricing long-term assets, whatever the approach adopted, requires making hypotheses to the effect that the recent past tells us something about the future, which, in ways to be defined and which vary from one model to the next, translates into hypotheses that some form of stationarity prevails. Completing the Arrow-Debreu pricing approach with an additional stationarity hypothesis provides an interesting perspective on the pricing of multiperiod cash flows. This is the purpose of the present section.

For notational simplicity, let us first assume that the same two states of nature (ranges of value of M) can be realized in each period, and that all future state-contingent cash flows have been estimated. The structure of the cash flow is found in Figure 10.6.

Insert Figure 10.6

Suppose also that we have estimated, using our formulae derived earlier, the values of the one-period state-contingent claims as follows:

$$\begin{array}{cc} & \begin{array}{cc} \text{Tomorrow} \\ 1 & 2 \end{array} \\ \begin{array}{c} \text{Today} \end{array} & \begin{array}{cc} 1 & 2 \\ \left[\begin{array}{cc} .54 & .42 \\ .46 & .53 \end{array} \right] \end{array} = \mathbf{q} \end{array}$$

where q_{11} ($= .54$) is the price today of an Arrow-Debreu claim paying \$1 if state 1 (a boom) occurs tomorrow, given that we are in state 1 (boom) today. Similarly, q_{12} is the price today of an Arrow-Debreu claim paying \$1 if state 2 (recession) occurs tomorrow given that we are in state 1 today. Note that these prices differ because the distribution of the value of M tomorrow differs depending on the state today.

Now let us introduce our stationarity hypothesis. Suppose that \mathbf{q} , the matrix of values, is invariant through time.⁹ That is, the same two states of nature

⁹If it were not, the approach in Figure 10.7 would carry on provided we would be able

describe the possible futures at all future dates and the contingent one-period prices remain the same. This allows us to interpret powers of the \mathbf{q} matrix, \mathbf{q}^2 , \mathbf{q}^3 , . . . in a particularly useful way. Consider \mathbf{q}^2 (see also Figure 10.7):

$$\mathbf{q}^2 = \begin{bmatrix} .54 & .42 \\ .46 & .53 \end{bmatrix} \cdot \begin{bmatrix} .54 & .42 \\ .46 & .53 \end{bmatrix} = \begin{bmatrix} (.54)(.54) + (.42)(.46) & (.54)(.42) + (.42)(.53) \\ (.46)(.54) + (.53)(.46) & (.46)(.42) + (.53)(.53) \end{bmatrix}$$

Note there are two ways to be in state 1 two periods from now, given we are in state 1 today. Therefore, the price today of \$1.00, if state 1 occurs in two periods, given we are in state 1 today is:

$$\underbrace{(.54)(.54)}_{\text{value of \$1 in 2 periods if state 1 occurs and the intermediate state is 1}} + \underbrace{(.42)(.46)}_{\text{value of \$1.00 in 2 periods if state 1 occurs and the intermediate state is 2}} .$$

Similarly, $\mathbf{q}_{22}^2 = (.46)(.42) + (.53)(.53)$ is the price today, if today's state is 2, of \$1.00 contingent on state 2 occurring in 2 periods. In general, for powers N of the matrix \mathbf{q} , we have the following interpretation for \mathbf{q}_{ij}^N : Given that we are in state i today, it gives the price today of \$1.00, contingent on state j occurring in N periods. Of course, if we hypothesized three states, then the Arrow-Debreu matrices would be 3×3 and so forth.

How can this information be used in a "capital budgeting" problem? First we must estimate the cash flows. Suppose they are as outlined in Table 10.9.

Table 10.9: State Contingent Cash Flows

$t = 0$	1	2	3
state 1	$\begin{bmatrix} 42 \end{bmatrix}$	$\begin{bmatrix} 48 \end{bmatrix}$	$\begin{bmatrix} 60 \end{bmatrix}$
state 2	$\begin{bmatrix} 65 \end{bmatrix}$	$\begin{bmatrix} 73 \end{bmatrix}$	$\begin{bmatrix} 58 \end{bmatrix}$

Then the present value (PV) of the cash flows, contingent on state 1 or state 2 to compute forward Arrow-Debreu prices; in other words, the Arrow-Debreu matrix would change from date to date and it would have to be time-indexed. Mathematically, the procedure described would carry over, but the information requirement would, of course, be substantially larger.

2 are given by:

$$\begin{aligned}
PV &= \begin{bmatrix} PV_1 \\ PV_2 \end{bmatrix} \\
&= \begin{bmatrix} .54 & .42 \\ .46 & .53 \end{bmatrix} \begin{bmatrix} 42 \\ 65 \end{bmatrix} + \begin{bmatrix} .54 & .42 \\ .46 & .53 \end{bmatrix}^2 \begin{bmatrix} 48 \\ 73 \end{bmatrix} + \begin{bmatrix} .54 & .42 \\ .46 & .53 \end{bmatrix}^3 \begin{bmatrix} 60 \\ 58 \end{bmatrix} \\
&= \begin{bmatrix} .54 & .42 \\ .46 & .53 \end{bmatrix} \begin{bmatrix} 42 \\ 65 \end{bmatrix} + \begin{bmatrix} .4848 & .4494 \\ .4922 & .4741 \end{bmatrix} \begin{bmatrix} 48 \\ 73 \end{bmatrix} + \begin{bmatrix} .4685 & .4418 \\ .4839 & .4580 \end{bmatrix} \begin{bmatrix} 60 \\ 54 \end{bmatrix} \\
&= \begin{bmatrix} 49.98 \\ 53.77 \end{bmatrix} + \begin{bmatrix} 56.07 \\ 58.23 \end{bmatrix} + \begin{bmatrix} 53.74 \\ 55.59 \end{bmatrix} = \begin{bmatrix} 159.79 \\ 167.59 \end{bmatrix}.
\end{aligned}$$

This procedure can be expanded to include as many states of nature as one may wish to define. This amounts to choosing as fine a partition of the range of possible values of M that one wishes to choose. It makes no sense to construct a finer partition, however, if we have no real basis for estimating different cash flows in those states. For most practical problems, three or four states are probably sufficient. But an advantage of this method is that it forces one to think carefully about what a project cash flow will be in each state, and what the relevant states, in fact, are.

One may wonder whether this methodology implicitly assumes that the states are equally probable. That is not the case. Although the probabilities, which would reflect the likelihood of the value of M lying in the various intervals, are not explicit, they are built into the prices of the state-contingent claims.

We close this chapter by suggesting a way to tie the approach proposed here with our previous work in this Chapter. Risk-free cash flows are special (degenerate) examples of risky cash flows. It is thus easy to use the method of this section to price risk-free flows. The comparison with the results obtained with the method of Section 10.3 then provides a useful check of the appropriateness of the assumptions made in the present context.

Consider our earlier example with Arrow-Debreu prices given by:

$$\begin{array}{cc}
& \begin{matrix} 1 & 2 \end{matrix} \\
\begin{matrix} \text{State 1} \\ \text{State 2} \end{matrix} & \begin{bmatrix} .54 & .42 \\ .46 & .53 \end{bmatrix}
\end{array}$$

If we are in state 1 today, the price of \$1.00 in each state tomorrow (i.e., a risk-free cash flow tomorrow of \$1.00) is $.54 + .42 = .96$. This implies a risk-free rate of:

$$(1 + r_f^1) = \frac{1.00}{.96} = 1.0416 \text{ or } 4.16\%.$$

To put it differently, $.54 + .42 = .96$ is the price of a one-period discount bond paying \$1.00 in one period, given that we are in state 1 today. More generally, we would evaluate the following risk-free cash flow as:

$$\begin{array}{cccc}
t = 0 & 1 & 2 & 3 \\
& 100 & 100 & 100
\end{array}$$

$$\begin{aligned}
PV &= \begin{bmatrix} PV_1 \\ PV_2 \end{bmatrix} \\
&= \begin{bmatrix} .54 & .42 \\ .46 & .53 \end{bmatrix} \begin{bmatrix} 100 \\ 100 \end{bmatrix} + \begin{bmatrix} .54 & .42 \\ .46 & .53 \end{bmatrix}^2 \begin{bmatrix} 100 \\ 100 \end{bmatrix} + \begin{bmatrix} .54 & .42 \\ .46 & .53 \end{bmatrix}^3 \begin{bmatrix} 100 \\ 100 \end{bmatrix} \\
&= \begin{bmatrix} .54 & .42 \\ .46 & .53 \end{bmatrix} \begin{bmatrix} 100 \\ 100 \end{bmatrix} + \begin{bmatrix} .4848 & .4494 \\ .4922 & .4741 \end{bmatrix} \begin{bmatrix} 100 \\ 100 \end{bmatrix} + \begin{bmatrix} .4685 & .4418 \\ .4839 & .4580 \end{bmatrix} \begin{bmatrix} 100 \\ 100 \end{bmatrix}
\end{aligned}$$

So

$$\begin{aligned}
PV_1 &= [.54 + .42]100 + [.4848 + .4494]100 + [.4685 + .4418] \\
&= [.96]100 + [.9342]100 + [.9103]100 \\
&= 280.45
\end{aligned}$$

where $[.96]$ = price of a one-period discount bond given state 1 today, $[.9342]$ = price of a two-period discount bond given state 1 today, $[.9103]$ = price of a three-period discount bond given state 1 today. The PV given state 2 is computed analogously. Now this provides us with a **verification test**: If the price of a discount bond using this method does not coincide with the prices using the approach developed in Section 10.3 (which relies on quoted coupon bond prices), then this must mean that our states are not well defined or numerous enough or that the assumptions of the option pricing formulae used to compute Arrow-Debreu prices are inadequate.

10.9 Conclusions

This chapter has served two main purposes. First, it has provided us with a platform to think more in depth about the all-important notion of market completeness. Our demonstration that, in principle, a portfolio of simple calls and puts written on the market portfolio might suffice to reach a complete market structure suggests the ‘Holy Grail’ may not be totally out of reach. Caution must be exercised, however, in interpreting the necessary assumptions. Can we indeed assume that the market portfolio – and what do we mean by the latter – is an adequate reflection of all the economically relevant states of nature? And the time dimension of market completeness should not be forgotten. The most relevant state of nature for a Swiss resident of 40 years of age may be the possibility of a period of prolonged depression with high unemployment in Switzerland 25 years from now (i.e., when he is nearing retirement¹⁰). Now extreme aggregate economic conditions would certainly be reflected in the Swiss Market Index (SMI), but options with 20-year maturities are not customarily

¹⁰The predominant pension regime in Switzerland is a defined benefit scheme with the benefits defined as a fraction of the last salary.

traded. Is it because of a lack of demand (possibly meaning that our assumption as to the most relevant state is not borne out), or because the structure of the financial industry is such that the supply of securities for long horizons is deficient?¹¹

The second part of the chapter discussed how Arrow-Debreu prices can be extracted from option prices (in the case where the relevant option is actively traded) or option pricing formulas (in the case where they are not). This discussion helps make Arrow-Debreu securities a less abstract concept. In fact, in specific cases the detailed procedure is fully operational and may indeed be the wiser route to evaluating risky cash flows. The key hypotheses are similar to those we have just discussed: The relevant states of nature are adequately distinguished by the market portfolio, a hypothesis that may be deemed appropriate if the context is limited to the valuation of risky cash flows. Moreover, in the case where options are not traded, the quality of the extracted Arrow-Debreu prices depends on the appropriateness of the various hypotheses imbedded in the option pricing formulas to which one has recourse. This issue has been abundantly discussed in the relevant literature.

References

- Banz, R., Miller, M. (1978), "Prices for State-Contingent Claims: Some Estimates and Applications," *Journal of Business* 51, 653-672.
- Breeden, D., Litzenberger, R. H. (1978), "Prices of State-contingent Claims Implicit in Option Prices," *Journal of Business* 51, 621-651.
- Pirkner, C.D., Weigend, A.S., Zimmermann, H. (1999), "Extracting Risk-Neutral Densities from Option Prices Using Mixture Binomial Trees," University of St-Gallen. Mimeographed.
- Ross, S. (1976), "Options and Efficiency," *Quarterly Journal of Economics* 90, 75-810.
- Shiller, R.J. (1993), *Macro Markets-Creating Institutions for Managing Society's Largest Economic Risks*, Clarendon Press, Oxford.
- Varian, H. (1987), "The Arbitrage Principle in Financial Economics," *Journal of Economic Perspectives* 1(2), 55-72

Appendix 10.1: Forward Prices and Forward Rates

Forward prices and forward rates correspond to the prices of (rates of return earned by) securities to be issued in the future.

¹¹A forceful statement in support of a similar claim is found in Shiller (1993) (see also the conclusions to Chapter 1). For the particular example discussed here, it may be argued that shorting the SMI (Swiss Market Index) would provide the appropriate hedge. Is it conceivable to take a short SMI position with a 20-year horizon?

Let ${}_kf_\tau$ denote the (compounded) rate of return on a risk-free discount bond to be issued at a future date k and maturing at date $k + \tau$. These forward rates are defined by the equations:

$$\begin{aligned}(1 + r_1)(1 + {}_1f_1) &= (1 + r_2)^2 \\ (1 + r_1)(1 + {}_1f_2)^2 &= (1 + r_3)^3 \\ (1 + r_2)^2(1 + {}_2f_1) &= (1 + r_3)^3, \text{etc.}\end{aligned}$$

We emphasize that the forward rates are *implied* forward rates, in the sense that the corresponding contracts are typically not traded. However, it is feasible to *lock in* these forward rates; that is, to guarantee their availability in the future. Suppose we wished to lock in the one-year forward rate one year from now. This amounts to creating a new security “synthetically” as a portfolio of existing securities, and is accomplished by simply undertaking a series of *long* and *short* transactions today. For example, take as given the implied discount bond prices of Table 10.5 and consider the transactions in Table 10.10.

Table 10.10: Locking in a Forward Rate

$t =$	0	1	2
Buy a 2-yr bond	- 1,000	65	1,065
Sell short a 1-yr bond	+ 1,000	-1,060	
	0	- 995	1,065

The portfolio we have constructed has a zero cash flow at date 0, requires an investment of \$995 at date 1, and pays \$1,065 at date 2. The gross return on the date 1 investment is

$$\frac{1065}{995} = 1.07035.$$

That this is exactly equal to the corresponding forward rate can be seen from the forward rate definition:

$$1 + {}_1f_1 = \frac{(1 + r_2)^2}{(1 + r_1)} = \frac{(1.065163)^2}{1.06} = 1.07035.$$

Let us scale back the previous transactions to create a \$1,000 payoff for the forward security. This amounts to multiplying all of the indicated transactions by $\frac{1000}{1065} = .939$.

Table 10.11: Creating a \$1,000 Payoff

$t =$	0	1	2
Buy .939 x 2-yr bonds	- 939	61.0	1,000
Sell short .939 x 1-yr bonds	+ 939	-995.34	
	0	-934.34	1,000

This price (\$934.34) is the no arbitrage price of this forward bond, no arbitrage in the sense that if there were any other contract calling for the delivery of such a bond at a price different from \$934.34, an arbitrage opportunity would exist.¹²

¹²The approach of this section can, of course, be generalized to more distant forward rates.

Chapter 11 : The Martingale Measure: Part I

11.1 Introduction

The theory of risk-neutral valuation reviewed in the present chapter proposes yet another way to tackle the valuation problem.¹ Rather than modify the denominator – the discount factor – to take account of the risky nature of a cash flow to be valued, or the numerator, by transforming the expected cash flows into their certainty equivalent, risk-neutral valuation simply corrects the *probabilities* with respect to which the expectation of the future cash flows is taken. This is done in such a way that discounting at the risk-free rate is legitimate. It is thus a procedure by which an asset valuation problem is transformed into one in which the asset's expected cash flow, computed now with respect to a new set of risk-neutral probabilities, can be discounted at the risk-free rate. The risk-neutral valuation methodology thus places an arbitrary valuation problem into a context in which all fairly priced assets earn the risk-free rate.

Importantly, the Martingale pricing theory or, equivalently, the theory of risk-neutral valuation is founded on preference-free pure arbitrage principles. That is, it is free of the structural assumptions on preferences, expectations, and endowments that make the CAPM and the CCAPM so restrictive. In this respect, the present chapter illustrates how far one can go in pricing financial assets while abstracting from the usual structural assumptions.

Risk-neutral probability distributions naturally assume a variety of forms, depending on the choice of setting. We first illustrate them in the context of a well-understood, finite time Arrow-Debreu complete markets economy. This is not the context in which the idea is most useful, but it is the one from which the basic intuition can be most easily understood. In addition, this strategy serves to clarify the very tight relationship between Arrow-Debreu pricing and Martingale pricing despite the apparent differences in terminology and perspectives.

11.2 The Setting and the Intuition

Our setting for these preliminary discussions is the particularly simple one with which we are now long familiar. There are two dates, $t = 0$ and $t = 1$. At date $t = 1$, any one of $j = 1, 2, \dots, J$ possible states of nature can be realized; denote the j th state by θ_j and its objective probability by π_j . We assume $\pi_j > 0$ for all θ_j .

Securities are competitively traded in this economy. There is a risk-free security that pays a fixed return r_f ; its period t price is denoted by $q^b(t)$. By convention, we customarily assume $q^b(0) = 1$, and its price at date 1 is $q^b(1) \equiv q^b(\theta_j, 1) = (1 + r_f)$, for all states θ_j . Since the date 1 price of the security is $(1 + r_f)$ in any state, we can as well drop the first argument in the

¹The theory of risk-neutral valuation was first developed by Harrison and Kreps (1979). Pliska (1997) provides an excellent review of the notion in discrete time. The present chapter is based on his presentation.

pricing function indicating the state in which the security is valued².

Also traded are N fundamental risky securities, indexed $i = 1, 2, \dots, N$, which we think of as stocks. The period $t = 0$ price of the i th such security is represented as $q_i^e(0)$. In period $t = 1$ its contingent payoff, given that state θ_j is realized, is given by $q_i^e(\theta_j, 1)$.³ It is also assumed that investors may hold any linear combination of the fundamental risk-free and risky securities. No assumption is made, however, regarding the number of securities that may be linearly independent vis-à-vis the number of states of nature: The securities market may or may not be complete. Neither is there any mention of agents' preferences. Otherwise the setting is standard Arrow-Debreu. Let \mathbf{S} denote the set of all fundamental securities, the stocks and the bond, and linear combinations thereof.

For this setting, the existence of a set of risk-neutral probabilities or, in more customary usage, a risk-neutral probability measure, effectively means the existence of a set of state probabilities, $\pi_j^{RN} > 0$, $j = 1, 2, \dots, J$ such that for each and every fundamental security $i = 1, 2, \dots, N$

$$q_i^e(0) = \frac{1}{(1 + r_f)} E_{\pi^{RN}}(q_i^e(\theta, 1)) = \frac{1}{(1 + r_f)} \sum_{j=1}^J \pi_j^{RN} q_i^e(\theta_j, 1) \quad (11.1)$$

(the analogous relationship automatically holds for the risk-free security).

To gain some intuition as to what might be necessary, at a minimum, to guarantee the existence of such probabilities, first observe that in our setting the π_j^{RN} represent strictly positive numbers that must satisfy a large system of equations of the form

$$q_i^e(0) = \pi_1^{RN} \left(\frac{q_i^e(\theta_1, 1)}{1 + r_f} \right) + \dots + \pi_J^{RN} \left(\frac{q_i^e(\theta_J, 1)}{1 + r_f} \right), i = 1, 2, \dots, N, \quad (11.2)$$

together with the requirement that $\pi_j^{RN} > 0$ for all j and $\sum_{j=1}^J \pi_j^{RN} = 1$.⁴

Such a system most certainly will not have a solution if there exist two fundamental securities, s and k , with the same $t = 0$ price, $q_s^e(0) = q_k^e(0)$, for which one of them, say k , pays as much as s in every state, and strictly more in at least one state; in other words,

$$q_k^e(\theta_j, 1) \geq q_s^e(\theta_j, 1) \text{ for all } j, \text{ and } q_k^e(\theta_{\hat{j}}, 1) > q_s^e(\theta_{\hat{j}}, 1) \quad (11.3)$$

for at least one $j = \hat{j}$. The Equations (11.2) corresponding to securities s and k would, for any set $\{\pi_j^{RN}: j = 1, 2, \dots, J\}$ have the same left-hand sides,

²In this chapter, it will be useful for the clarity of exposition to alter some of our previous notational conventions. One of the reasons is that we will want, symmetrically for all assets, to distinguish between their price at date 0 and their price at date 1 under any given state θ_j .

³In the parlance and notation of Chapter 8, $q_i^e(\theta_j, 1)$ is the cash flow associated with security i if state θ_j is realized, $CF^i(\theta_j)$.

⁴Compare this system of equations with those considered in Section 10.2 when extracting Arrow-Debreu prices from a complete set of prices for complex securities.

yet different right-hand sides, implying no solution to the system. But two such securities cannot themselves be consistently priced because, together, they constitute an *arbitrage opportunity*: Short one unit of security s , long one unit of security k and pocket the difference $q_k^e(\theta_j, 1) - q_s^e(\theta_j, 1) > 0$ if state j occurs; replicate the transaction many times over. These remarks suggest, therefore, that the existence of a risk-neutral measure is, in some intimate way, related to the absence of arbitrage opportunities in the financial markets. This is, in fact, the case, but first some notation, definitions, and examples are in order.

11.3 Notation, Definitions, and Basic Results

Consider a portfolio, P , composed of n_P^b risk-free bonds and n_P^i units of risky security i , $i = 1, 2, \dots, N$. No restrictions will be placed on n_P^b , n_P^i : Short sales are permitted; they can, therefore, take negative values, and fractional share holdings are acceptable. The value of this portfolio at $t = 0$, $V_P(0)$, is given by

$$V_P(0) = n_P^b q^b(0) + \sum_{i=1}^N n_P^i q_i^e(0), \quad (11.4)$$

while its value at $t = 1$, given that state θ_j , is realized is

$$V_P(\theta_j, 1) = n_P^b q^b(1) + \sum_{i=1}^N n_P^i q_i^e(\theta_j, 1). \quad (11.5)$$

With this notation we are now in a position to define our basic concepts.

Definition 11.1:

A portfolio P in \mathbf{S} constitutes an arbitrage opportunity provided the following conditions are satisfied:

$$\begin{aligned} (i) \quad V_P(0) &= 0, \\ (ii) \quad V_P(\theta_j, 1) &\geq 0, \text{ for all } j \in \{1, 2, \dots, J\}, \\ (iii) \quad V_P(\theta_j, 1) &> 0, \text{ for at least one } j \in \{1, 2, \dots, J\}. \end{aligned} \quad (11.6)$$

This is the standard sense of an arbitrage opportunity: With no initial investment and no possible losses (thus no risk), a profit can be made in at least one state. Our second crucial definition is Definition 11.2.

Definition 11.2:

A probability measure $\{\pi_j^{RN}\}_{j=1}^J$ defined on the set of states θ_j , $j = 1, 2, \dots, J$, is said to be a risk-neutral probability measure if

$$\begin{aligned} (i) \quad \pi_j^{RN} &> 0, \text{ for all } j = 1, 2, \dots, J, \text{ and} \\ (ii) \quad q_i^e(0) &= E_{\pi^{RN}} \left\{ \frac{q_i^e(\theta, 1)}{1 + r_f} \right\}, \end{aligned} \quad (11.7)$$

for all fundamental risky securities $i = 1, 2, \dots, N$ in \mathbf{S} .

Both elements of this definition are crucial. Not only must each individual security be priced equal to the present value of its expected payoff, the latter computed using the risk-neutral probabilities (and thus it must also be true of portfolios of them), but these probabilities must also be strictly positive. To find them, if they exist, it is necessary only to solve the system of equations implied by part (ii) of Equation (11.6) of the risk-neutral probability definition. Consider the Examples 11.1 through 11.4.

Example 11.1:

There are two periods and two fundamental securities, a stock and a bond, with prices and payoffs presented in Table 11.1.

Table 11.1: Fundamental Securities for Example 11.1

Period $t = 0$ Prices	Period $t = 1$ Payoffs	
		θ_1 θ_2
$q^b(0): 1$	$q^b(1):$	1.1 1.1
$q^e(0): 4$	$q^e(\theta_j, 1):$	3 7

By the definition of a risk-neutral probability measure, it must be the case that simultaneously

$$4 = \pi_1^{RN} \left(\frac{3}{1.1} \right) + \pi_2^{RN} \left(\frac{7}{1.1} \right)$$

$$1 = \pi_1^{RN} + \pi_2^{RN}$$

Solving this system of equations, we obtain $\pi_1^{RN} = .65$, $\pi_2^{RN} = .35$.

For future reference note that the fundamental securities in this example define a complete set of financial markets for this economy, and that there are clearly no arbitrage opportunities among them.

Example 11.2:

Consider next an analogous economy with three possible states of nature, and three securities, as found in Table 11.2.

Table 11.2: Fundamental Securities for Example 11.2

Period $t = 0$ Prices	Period $t = 1$ Payoffs			
		θ_1	θ_2	θ_3
$q^b(0): 1$	$q^b(1):$	1.1	1.1	1.1
$q_1^e(0): 2$	$q_1^e(\theta_j, 1):$	3	2	1
$q_2^e(0): 3$	$q_2^e(\theta_j, 1):$	1	4	6

The relevant system of equations is now

$$\begin{aligned} 2 &= \pi_1^{RN} \left(\frac{3}{1.1} \right) + \pi_2^{RN} \left(\frac{2}{1.1} \right) + \pi_3^{RN} \left(\frac{1}{1.1} \right) \\ 3 &= \pi_1^{RN} \left(\frac{1}{1.1} \right) + \pi_2^{RN} \left(\frac{4}{1.1} \right) + \pi_3^{RN} \left(\frac{6}{1.1} \right) \\ 1 &= \pi_1^{RN} + \pi_2^{RN} + \pi_3^{RN}. \end{aligned}$$

The solution to this set of equations,

$$\pi_1^{RN} = .3, \quad \pi_2^{RN} = .6, \quad \pi_3^{RN} = .1,$$

satisfies the requirements of a risk-neutral measure. By inspection we again observe that this financial market is complete, and that there are no arbitrage opportunities among the three securities.

Example 11.3:

To see what happens when the financial markets are incomplete, consider the securities in Table 11.3.

Table 11.3: Fundamental Securities for Example 11.3

Period $t = 0$ Prices	Period $t = 1$ Payoffs			
		θ_1	θ_2	θ_3
$q^b(0): 1$	$q^b(1):$	1.1	1.1	1.1
$q_1^e(0): 2$	$q_1^e(\theta_j, 1):$	1	2	3

For this example the relevant system is

$$\begin{aligned} 2 &= \pi_1^{RN} \left(\frac{1}{1.1} \right) + \pi_2^{RN} \left(\frac{2}{1.1} \right) + \pi_3^{RN} \left(\frac{3}{1.1} \right) \\ 1 &= \pi_1^{RN} + \pi_2^{RN} + \pi_3^{RN} \end{aligned}$$

Because this system is under-determined, there will be many solutions. Without loss of generality, first solve for π_2^{RN} and π_3^{RN} in terms of π_1^{RN} :

$$2.2 - \pi_1^{RN} = 2\pi_2^{RN} + 3\pi_3^{RN}$$

$$1 - \pi_1^{RN} = \pi_2^{RN} + \pi_3^{RN},$$

which yields the solution $\pi_3^{RN} = .2 + \pi_1^{RN}$, and $\pi_2^{RN} = .8 - 2\pi_1^{RN}$.

In order for a triple $(\pi_1^{RN}, \pi_2^{RN}, \pi_3^{RN})$ to simultaneously solve this system of equations, while also satisfying the strict positivity requirement of risk-neutral probabilities, the following inequalities must hold:

$$\begin{aligned}\pi_1^{RN} &> 0 \\ \pi_2^{RN} &= .8 - 2\pi_1^{RN} > 0 \\ \pi_3^{RN} &= .2 + \pi_1^{RN} > 0\end{aligned}$$

By the second inequality $\pi_1^{RN} < .4$, and by the third $\pi_1^{RN} > -.2$. In order that all probabilities be strictly positive, it must, therefore, be the case that

$$0 < \pi_1^{RN} < .4,$$

with π_2^{RN} and π_3^{RN} given by the indicated equalities.

In an incomplete market, therefore, there appear to be many risk-neutral probability sets: any triple $(\pi_1^{RN}, \pi_2^{RN}, \pi_3^{RN})$ where

$$(\pi_1^{RN}, \pi_2^{RN}, \pi_3^{RN}) \in \{(\lambda, .8 - 2\lambda, .2 + \lambda) : 0 < \lambda < .4\}$$

serves as a risk-neutral probability measure for this economy.

Example 11.4:

Lastly, we may as well see what happens if the set of fundamental securities contains an arbitrage opportunity (see Table 11.4).

Table 11.4: Fundamental Securities for Example 11.4

Period $t = 0$ Prices	Period $t = 1$ Payoffs			
		θ_1	θ_2	θ_3
$q^b(0): 1$	$q^b(1):$	1.1	1.1	1.1
$q_1^e(0): 2$	$q_1^e(\theta_j, 1):$	2	3	1
$q_2^e(0): 2.5$	$q_2^e(\theta_j, 1):$	4	5	3

Any attempt to solve the system of equations defining the risk-neutral probabilities fails in this case. There is no solution. Notice also the implicit arbitrage opportunity: risky security 2 dominates a portfolio of one unit of the risk-free security and one unit of risky security 1, yet it costs less.

It is also possible to have a solution in the presence of arbitrage. In this case, however, at least one of the solution probabilities will be zero, disqualifying the set for the risk-neutral designation.

Together with our original intuition, these examples suggest that arbitrage opportunities are incompatible with the existence of a risk-neutral probability measure. This is the substance of the first main result.

Proposition 11.1:

Consider the two-period setting described earlier in this chapter. Then there exists a risk-neutral probability measure on \mathbf{S} , if and only if there are no arbitrage opportunities among the fundamental securities.

Proposition 11.1 tells us that, provided the condition of the absence of arbitrage opportunities characterizes financial markets, our ambition to use distorted, risk-neutral probabilities to compute expected cash flows and discount at the risk-free rate has some legitimacy! Note, however, that the proposition admits the possibility that there may be many such measures, as in Example 11.3.

Proposition 11.1 also provides us, in principle, with a method for testing whether a set of fundamental securities contains an arbitrage opportunity: If the system of Equations (11.7.ii) has no solution probability vector where all the terms are strictly positive, an arbitrage opportunity is present. Unless we are highly confident of the actual states of nature and the payoffs to the various fundamental securities in those states, however, this observation is of limited use. But even for a very large number of securities it is easy to check computationally.

Although we have calculated the risk-neutral probabilities with respect to the prices and payoff of the fundamental securities only, the analogous relationship must hold for arbitrary portfolios in \mathbf{S} – all linear combinations of the fundamental securities – in the absence of arbitrage opportunities. This result is formalized in Proposition 11.2.

Proposition 11.2:

Suppose the set of securities \mathbf{S} is free of arbitrage opportunities. Then for any portfolio \hat{P} in \mathbf{S}

$$V_{\hat{P}}(0) = \frac{1}{(1+r_f)} E_{\pi^{RN}} \tilde{V}_{\hat{P}}(\theta, 1), \quad (11.8)$$

for any risk-neutral probability measure π^{RN} on \mathbf{S} .

Proof:

Let \hat{P} be an arbitrary portfolio in \mathbf{S} , and let it be composed of $n_{\hat{P}}^b$ bonds and $n_{\hat{P}}^i$ shares of fundamental risky asset i . In the absence of arbitrage, \hat{P} must be priced equal to the value of its constituent securities, in other words,

$$V_{\hat{P}}(0) = n_{\hat{P}}^b q^b(0) + \sum_{i=1}^N n_{\hat{P}}^i q_i^e(0) = n_{\hat{P}}^b E_{\pi^{RN}} \left(\frac{q^b(1)}{1+r_f} \right) + \sum_{i=1}^N n_{\hat{P}}^i E_{\pi^{RN}} \left(\frac{\tilde{q}_i^e(\theta, 1)}{1+r_f} \right),$$

for any risk neutral probability measure π^{RN} ,

$$= E_{\pi^{RN}} \left\{ \frac{n_{\hat{P}}^b q^b(1) + \sum_{i=1}^N n_{\hat{P}}^i \tilde{q}_i^e(\theta, 1)}{1+r_f} \right\} = \frac{1}{(1+r_f)} E_{\pi^{RN}} \left(\tilde{V}_{\hat{P}}(\theta, 1) \right).$$

Proposition 11.2 is merely a formalization of the obvious fact that if every security in the portfolio is priced equal to the present value, discounted at r_f , of its expected payoffs computed with respect to the risk-neutral probabilities, the same must be true of the portfolio itself. This follows from the linearity of the expectations operator and the fact that the portfolio is valued as the sum total of its constituent securities, which must be the case in the absence of arbitrage opportunities.

A multiplicity of risk-neutral measures on \mathbf{S} does not compromise this conclusion in any way, because each of them assigns the same value to the fundamental securities and thus to the portfolio itself via Equation (11.8). For completeness, we note that a form of a converse to Proposition 11.2 is also valid.

Proposition 11.3:

Consider an arbitrary period $t = 1$ payoff $\tilde{x}(\theta, 1)$ and let M represent the set of all risk-neutral probability measures on the set \mathbf{S} . Assume \mathbf{S} contains no arbitrage opportunities. If

$$\frac{1}{(1 + r_f)} E_{\pi^{RN}} \tilde{x}(\theta, 1) = \frac{1}{(1 + r_f)} E_{\hat{\pi}^{RN}} \tilde{x}(\theta, 1) \quad \text{for any } \pi^{RN}, \hat{\pi}^{RN} \in M,$$

then there exists a portfolio in \mathbf{S} with the same $t = 1$ payoff as $\tilde{x}(\theta, 1)$.

It would be good to be able to dispense with the complications attendant to multiple risk-neutral probability measures on \mathbf{S} . When this is possible is the subject of Section 11.4.

11.4 Uniqueness

Examples 11.1 and 11.2 both possessed unique risk-neutral probability measures. They were also complete markets models. This illustrates an important general proposition.

Proposition 11.4:

Consider a set of securities \mathbf{S} without arbitrage opportunities. Then \mathbf{S} is complete if and only if there exists exactly one risk-neutral probability measure.

Proof:

Let us prove one side of the proposition, as it is particularly revealing. Suppose \mathbf{S} is complete and there were two risk-neutral probability measures, $\{\pi_j^{RN} : j = 1, 2, \dots, J\}$ and $\{\tilde{\pi}_j^{RN} : j = 1, 2, \dots, J\}$. Then there must be at least one state \hat{j} for which $\pi_{\hat{j}}^{RN} \neq \tilde{\pi}_{\hat{j}}^{RN}$. Since the market is complete, one must be able to construct a portfolio P in \mathbf{S} such that

$$V_P(0) > 0, \text{ and } \begin{cases} V_P(\theta_j, 1) = 0 & j \neq \hat{j} \\ V_P(\theta_j, 1) = 1 & j = \hat{j} \end{cases}.$$

This is simply the statement of the existence of an Arrow-Debreu security associated with $\theta_{\hat{j}}$.

But then $\{\pi_j^{RN}: j = 1, 2, \dots, J\}$ and $\{\bar{\pi}_j^{RN}: j = 1, 2, \dots, J\}$ cannot both be risk-neutral measures as, by Proposition 11.2,

$$\begin{aligned} V_P(0) &= \frac{1}{(1+r_f)} E_{\pi^{RN}} \tilde{V}_P(\theta, 1) = \frac{\pi_j^{RN}}{(1+r_f)} \\ &\neq \frac{\bar{\pi}_j^{RN}}{(1+r_f)} = \frac{1}{(1+r_f)} E_{\bar{\pi}^{RN}} \tilde{V}_P(\theta, 1) \\ &= V_P(0), \text{ a contradiction.} \end{aligned}$$

Thus, there cannot be more than one risk-neutral probability measure in a complete market economy.

We omit a formal proof of the other side of the proposition. Informally, if the market is not complete, then the fundamental securities do not span the space. Hence, the system of Equations (11.6) contains more unknowns than equations, yet they are all linearly independent (no arbitrage). There must be a multiplicity of solutions and hence a multiplicity of risk-neutral probability measures.

Concealed in the proof of Proposition 11.4 is an important observation: The price of an Arrow-Debreu security that pays 1 unit of payoff if event θ_j is realized and nothing otherwise must be $\frac{\pi_j^{RN}}{(1+r_f)}$, the present value of the corresponding risk-neutral probability. In general,

$$q_j(0) = \frac{\pi_j^{RN}}{(1+r_f)}$$

where $q_j(0)$ is the $t = 0$ price of a state claim paying 1 if and only if state θ_j realized. Provided the financial market is complete, risk-neutral valuation is nothing more than valuing an uncertain payoff in terms of the value of a replicating portfolio of Arrow-Debreu claims. Notice, however, that we thus identify the all-important Arrow-Debreu prices without having to impose any of the economic structure of Chapter 8; in particular, knowledge of the agents' preferences is not required. This approach can be likened to describing the Arrow-Debreu pricing theory from the perspective of Proposition 10.2. It is possible, and less restrictive, to limit our inquiry to extracting Arrow-Debreu prices from the prices of a (complete) set of complex securities and proceed from there to price arbitrary cash flows. In the absence of further structure, nothing can be said, however, on the determinants of Arrow-Debreu prices (or risk-neutral probabilities).

Let us illustrate with the data of our second example. There we identified the unique risk-neutral measure to be:

$$\pi_1^{RN} = .3, \pi_2^{RN} = .6, \pi_3^{RN} = .1,$$

Together with $r_f = .1$, these values imply that the Arrow-Debreu security prices must be

$$q_1(0) = .3/1.1 = .27; q_2(0) = .6/1.1 = .55; q_3(0) = .1/1.1 = .09.$$

Conversely, given a set of Arrow-Debreu claims with strictly positive prices, we can generate the corresponding risk-neutral probabilities and the risk-free rate. As noted in earlier chapters, the period zero price of a risk-free security (one that pays one unit of the numeraire in every date $t = 1$ state) in this setting is given by

$$p_{r_f} = \sum_{j=1}^J q_j(0),$$

and thus

$$(1 + r_f) = \frac{1}{p_{r_f}} = \frac{1}{\sum_{j=1}^J q_j(0)}$$

We define the risk-neutral probabilities $\{\pi_j^{RN}(\theta)\}$ according to

$$\pi_j^{RN} = \frac{q_j(0)}{\sum_{j=1}^J q_j(0)} \quad (11.9)$$

Clearly $\pi_j^{RN} > 0$ for each state j (since $q_j(0) > 0$ for every state) and, by construction $\sum_{j=1}^J \pi_j^{RN} = 1$. As a result, the set $\{\pi_j^{RN}\}$ qualifies as a risk-neutral probability measure.

Referring now to the example developed in Section 8.3, let us recall that we had found a complete set of Arrow-Debreu prices to be $q_1(0) = .24$; $q_2(0) = .3$; this means, in turn, that the unique risk-neutral measure for the economy there described is

$$\pi_1^{RN} = .24/.54 = .444, \pi_2^{RN} = .3/.54 = .556.$$

For complete markets we see that the relationship between strictly positively priced state claims and the risk-neutral probability measure is indeed an intimate one: each implies the other. Since, in the absence of arbitrage possibilities, there can exist only one set of state claims prices, and thus only one risk-neutral probability measure, Proposition 11.4 is reconfirmed.

11.5 Incompleteness

What about the case in which \mathbf{S} is an incomplete set of securities? By Proposition 11.4 there will be a multiplicity of risk-neutral probabilities, but these will all give the same valuation to elements of \mathbf{S} (Proposition 11.2). Consider, however, a $t = 1$ bounded state-contingent payoff vector $\tilde{x}(\theta, 1)$ that does *not* coincide with the payoff to any portfolio in \mathbf{S} . By Proposition 11.4, different risk-neutral probability measures will assign different values to this payoff: essentially, its price is not well defined. It is possible, however, to establish *arbitrage bounds* on the value of this claim. For any risk-neutral probability π^{RN} ,

defined on \mathbf{S} , consider the following quantities:

$$\begin{aligned} H_x &= \inf \left\{ E_{\pi^{RN}} \left[\frac{\tilde{V}_P(\theta, 1)}{1 + r_f} \right] : V_P(\theta_j, 1) \geq x(\theta_j, 1), \forall j = 1, 2, \dots, J \text{ and } P \in \mathbf{S} \right\} \\ L_x &= \sup \left\{ E_{\pi^{RN}} \left[\frac{\tilde{V}_P(\theta, 1)}{1 + r_f} \right] : V_P(\theta_j, 1) \leq x(\theta_j, 1), \forall j = 1, 2, \dots, J \text{ and } P \in \mathbf{S} \right\} \end{aligned} \quad (11.10)$$

In these evaluations we don't care what risk-neutral measure is used because any one of them gives identical valuations for all portfolios in \mathbf{S} . Since, for some γ , $\gamma q^b(1) > x(\theta_j, 1)$, for all j , H_x is bounded above by $\gamma q^b(0)$, and hence is well defined (an analogous comment applies to L_x). The claim is that the no arbitrage price of x , $q^x(0)$ lies in the range

$$L_x \leq q^x(0) \leq H_x$$

To see why this must be so, suppose that $q^x(0) > H_x$ and let P^* be any portfolio in \mathbf{S} for which

$$\begin{aligned} q^x(0) &> V_{P^*}(0) > H_x, \text{ and} \\ V_{P^*}(\theta_j, 1) &\geq x(\theta_j, 1), \text{ for all } \theta_j, j = 1, 2, \dots, N. \end{aligned} \quad (11.11)$$

We know that such a P^* exists because the set

$$\mathbf{S}_x = \{P : P \in \mathbf{S}, V_P(\theta_j, 1) \geq x(\theta_j, 1), \text{ for all } j = 1, 2, \dots, J\}$$

is closed. Hence there is a \hat{P} in \mathbf{S}_x such that $E_{\pi^{RN}} \frac{\tilde{V}_{\hat{P}}(\theta, 1)}{(1+r_f)} = H_x$. By the continuity of the expectations operator, we can find a $\lambda > 1$ such that $\lambda \hat{P}$ in \mathbf{S}_x and ⁵

$$q^x(0) > \frac{1}{1 + r_f} E_{\pi^{RN}} \tilde{V}_{\lambda \hat{P}}(\theta, 1) = \lambda \frac{1}{1 + r_f} E_{\pi^{RN}} \tilde{V}_{\hat{P}}(\theta, 1) = \lambda H_x > H_x.$$

Since $\lambda > 1$, for all j , $V_{\lambda \hat{P}}(\theta_j, 1) > V_{\hat{P}}(\theta_j, 1) \geq x(\theta_j, 1)$; let $P^* = \lambda \hat{P}$. Now the arbitrage argument: Sell the security with title to the cash flow $x(\theta_j, 1)$, and buy the portfolio P^* . At time $t = 0$, you receive, $q^x(0) - V_{P^*}(0) > 0$, while at time $t = 1$ the cash flow from the portfolio, by Equation (11.11), fully covers the obligation under the short sale in every state; in other words, there is an arbitrage opportunity. An analogous argument demonstrates that $L_x \leq q^x(0)$.

In some cases it is readily possible to solve for these bounds.

Example 11.5:

Revisit, for example, our earlier Example 11.3, and consider the payoff

⁵By $\lambda \hat{P}$ we mean a portfolio with constituent bonds and stocks in the proportions $\lambda n_{\hat{P}}^b, \lambda n_{\hat{P}}^i$.

$$\frac{x(\hat{\theta}_j, 1) :}{\left| \begin{array}{c|c|c} \theta_1 & \theta_2 & \theta_3 \\ \hline 0 & 0 & 1 \end{array} \right|}$$

This security is most surely not in the span of the securities (1.1, 1.1, 1.1) and (1, 2, 3), a fact that can be confirmed by observing that the system of equations implied by equating

$$(0, 0, 1) = a(1.1, 1.1, 1.1) + b(1, 2, 3),$$

in other words, the system:

$$\begin{aligned} 0 &= 1.1a + b \\ 0 &= 1.1a + 2b \\ 1 &= 1.1a + 3b \end{aligned}$$

has no solution. But any portfolio in \mathbf{S} can be expressed as a linear combination of (1.1, 1.1, 1.1) and (1, 2, 3) and thus must be of the form

$$a(1.1, 1.1, 1.1) + b(1, 2, 3) = (a(1.1) + b, a(1.1) + 2b, a(1.1) + 3b)$$

for some a, b real numbers.

We also know that in computing H_x, L_x , any risk-neutral measure can be employed. Recall that we had identified the solution of Example 11.3 to be

$$(\pi_1^{RN}, \pi_2^{RN}, \pi_3^{RN}) \in \{(\lambda, .8 - 2\lambda, .2 + \lambda) : 0 < \lambda < .4\}$$

Without loss of generality, choose $\lambda = .2$; thus

$$(\pi_1^{RN}, \pi_2^{RN}, \pi_3^{RN}) = (.2, .4, .4).$$

For any choice of a, b (thereby defining a $\tilde{V}_P(\theta; 1)$)

$$\begin{aligned} E_{\pi^{RN}} \left[\frac{\tilde{V}_P(\theta; 1)}{(1 + r_f)} \right] &= \frac{.2 \{(1.1)a + b\} + .4 \{(1.1)a + 2b\} + .4 \{(1.1)a + 3b\}}{1.1} \\ &= \frac{(1.1)a + (2.2)b}{1.1} = a + 2b. \end{aligned}$$

Thus,

$$H_x = \inf_{a, b \in R} \{(a + 2b) : a(1.1) + b \geq 0, a(1.1) + 2b \geq 0, \text{ and } a(1.1) + 3b \geq 1\}$$

Similarly,

$$L_x = \sup_{a, b \in R} \{(a + 2b) : a(1.1) + b \leq 0, a(1.1) + 2b \leq 0, a(1.1) + 3b \leq 1\}$$

Table 11.5 Solutions for H_x and L_x

	H_x	L_x
a*	-.4545	-1.8182
b*	.5	1
H_x	.5455	
L_x		1818

Because the respective sets of admissible pairs are closed in R^2 , we can replace inf and sup by, respectively, min and max.

Solving for H_x, L_x thus amounts to solving small linear programs. The solutions, obtained via MATLAB are detailed in Table 11.5.

The value of the security (state claim), we may conclude, lies in the interval (.1818, .5455).

Before turning to the applications there is one additional point of clarification.

11.6 Equilibrium and No Arbitrage Opportunities

Thus far we have made no reference to financial equilibrium, in the sense discussed in earlier chapters. Clearly equilibrium implies no arbitrage opportunities: The presence of an arbitrage opportunity will induce investors to assume arbitrarily large short and long positions, which is inconsistent with the existence of equilibrium. The converse is also clearly not true. It could well be, in some specific market, that supply exceeds demand or conversely, without this situation opening up an arbitrage opportunity in the strict sense understood in this chapter. In what follows the attempt is made to convey the sense of risk-neutral valuation as an equilibrium phenomena.

Table 11.6 The Exchange Economy of Section 8.3 – Endowments and Preferences

	Endowments			Preferences
	$t = 0$	$t = 1$		
Agent 1	10	1	2	$U^1(c_0, c_1) = \frac{1}{2}c_0^1 + .9(\frac{1}{3}\ln(c_1^1) + \frac{2}{3}\ln(c_2^1))$
Agent 2	5	4	6	$U^2(c_0, c_1) = \frac{1}{2}c_0^2 + .9(\frac{1}{3}\ln(c_1^2) + \frac{2}{3}\ln(c_2^2))$

To illustrate, let us return to the first example in Chapter 8. The basic data of that Arrow-Debreu equilibrium is provided in Table 11.6. and the $t = 0$ corresponding equilibrium state prices are $q_1(0) = .24$ and $q_2(0) = .30$. In this case the risk-neutral probabilities are

$$\pi_1^{RN} = \frac{.24}{.54}, \text{ and } \pi_2^{RN} = \frac{.30}{.54}.$$

Suppose a stock were traded where $q^e(\theta_1, 1) = 1$, and $q^e(\theta_2, 1) = 3$. By risk-neutral valuation (or equivalently, using Arrow-Debreu prices), its period $t = 0$

price must be

$$q^e(0) = .54 \left[\frac{.24}{.54}(1) + \frac{.30}{.54}(3) \right] = 1.14;$$

the price of the risk-free security is $q^b(0) = .54$.

Verifying this calculation is a bit tricky because, in the original equilibrium, this stock was not traded. Introducing such assets requires us to decide what the original endowments must be, that is, who owns what in period 0. We cannot just add the stock arbitrarily, as the wealth levels of the agents would change as a result and, in general, this would alter the state prices, risk-neutral probabilities, and all subsequent valuations. The solution of this problem is to compute the equilibrium for a similar economy in which the two agents have the same preferences and in which the only traded assets are this stock and a bond. Furthermore, the initial endowments of these instruments must be such as to guarantee the same period $t = 0$ and $t = 1$ net endowment allocations as in the first equilibrium.

Let \hat{n}_e^i, \hat{n}_b^i denote, respectively, the initial endowments of the equity and debt securities of agent i , $i = 1, 2$. The equivalence noted previously is accomplished as outlined in Table 11.7 (see Appendix 11.1).

Table 11.7 Initial Holdings of Equity and Debt Achieving Equivalence with Arrow-Debreu Equilibrium Endowments

	$t = 0$		
	Consumption	\hat{n}_e^i	\hat{n}_b^i
Agent 1:	10	$\frac{1}{2}$	$\frac{1}{2}$
Agent 2:	5	1	3

A straightforward computation of the equilibrium prices yields the same $q^e(0) = 1.14$, and $q^b(0) = .54$ as predicted by risk-neutral valuation.

We conclude this section with one additional remark. Suppose one of the two agents were risk neutral; without loss of generality let this be agent 1. Under the original endowment scheme, his problem becomes:

$$\begin{aligned} & \max(10 + 1q_1(0) + 2q_2(0) - c_1^1 q_1(0) - c_2^1 q_2(0)) + .9(\frac{1}{3}c_1^1 + \frac{2}{3}c_2^1) \\ & \text{s.t. } c_1^1 q_1(0) + c_2^1 q_2(0) \leq 10 + q_1(0) + 2q_2(0) \end{aligned}$$

The first order conditions are

$$\begin{aligned} c_1^1 : & \quad q_1(0) = \frac{1}{3}.0.9 \\ c_2^1 : & \quad q_2(0) = \frac{2}{3}.0.9 \end{aligned}$$

from which it follows that $\pi_1^{RN} = \frac{\frac{1}{3}.0.9}{0.9} = \frac{1}{3}$ while $\pi_2^{RN} = \frac{\frac{2}{3}.0.9}{0.9} = \frac{2}{3}$, that is, in equilibrium, the risk-neutral probabilities coincide with the true probabilities. This is the source of the term *risk-neutral probabilities*: If at least one agent is risk neutral, the risk-neutral probabilities and the true probabilities coincide.

We conclude from this example that risk-neutral valuation holds in equilibrium, as it must because equilibrium implies no arbitrage. The risk-neutral probabilities thus obtained, however, are to be uniquely identified with that equilibrium, and it is meaningful to use them only for valuing securities that are elements of the participants' original endowments.

11.7 Application: Maximizing the Expected Utility of Terminal Wealth

11.7.1 Portfolio Investment and Risk-Neutral Probabilities

Risk-neutral probabilities are intimately related to the *basis* or the set of fundamental securities in an economy. Under no arbitrage, given the prices of fundamental securities, we obtain a risk-neutral probability measure, and vice versa. This raises the possibility that it may be possible to formulate any problem in wealth allocation, for example the classic consumption-savings problem, in the setting of risk-neutral valuation. In this section we consider a number of these connections.

The simplest portfolio allocation problem with which we have dealt involves an investor choosing a portfolio so as to maximize the expected utility of his period $t = 1$ (terminal) wealth (we retain, without loss of generality, the two-period framework). In our current notation, this problem takes the form: choose portfolio P , among all feasible portfolios, (i.e., P must be composed of securities in \mathbf{S} and the date-0 value of this portfolio (its acquisition price) cannot exceed initial wealth) so as to maximize expected utility of terminal wealth, which corresponds to the date-1 value of P :

$$\begin{aligned} \max_{\{n_P^b, n_P^i, i=1,2,\dots,N\}} & EU(\tilde{V}_P(\theta, 1)) \\ \text{s.t. } & V_P(0) = V_0, P \in \mathbf{S}, \end{aligned} \quad (11.12)$$

where V_0 is the investor's initial wealth, $U(\cdot)$ is her period utility function, assumed to have the standard properties, and n_P^b, n_P^i are the positions (not proportions, but units of indicated assets) in the risk-free asset and the risky asset $i = 1, 2, \dots, N$, respectively, defining portfolio P . It is not obvious that there should be a relationship between the solvability of this problem and the existence of a risk-neutral measure, but this is the case.

Proposition 11.5:

If Equation (11.12) has a solution, then there are no arbitrage opportunities in \mathbf{S} . Hence there exists a risk-neutral measure on \mathbf{S} .

Proof:

The idea is that an arbitrage opportunity is a costless way to endlessly improve upon the (presumed) optimum. So no optimum can exist. More formally, we prove the proposition by contradiction. Let $\hat{P} \in \mathbf{S}$ be a solution to Equation (11.12), and let \hat{P} have the structure $\{n_{\hat{P}}^b, n_{\hat{P}}^i : i = 1, 2, \dots, N\}$. Assume also

that there exists an arbitrage opportunity, in other words, a portfolio \tilde{P} , with structure $\{n_{\tilde{P}}^b, n_{\tilde{P}}^i : i = 1, 2, \dots, N\}$, such that $V_{\tilde{P}}(0) = 0$ and $E\tilde{V}_{\tilde{P}}(\theta, 1) > 0$. Consider the portfolio P^* with structure

$$\{n_{P^*}^b, n_{P^*}^i : i = 1, 2, \dots, N\}$$

$$n_{P^*}^b = n_{\tilde{P}}^b + n_{\tilde{P}}^b \text{ and } n_{P^*}^i = n_{\tilde{P}}^i + n_{\tilde{P}}^i, \quad i = 1, 2, \dots, N.$$

P^* is still feasible for the agent and it provides strictly more wealth in at least one state. Since $U(\cdot)$ is strictly increasing,

$$EU(\tilde{V}_{P^*}(\theta, 1)) > EU(\tilde{V}_{\tilde{P}}(\theta, 1)).$$

This contradicts \hat{P} as a solution to Equation (11.12). We conclude that there cannot exist any arbitrage opportunities and thus, by Proposition 11.1, a risk-neutral probability measure on \mathbf{S} must exist.

Proposition 11.5 informs us that arbitrage opportunities are incompatible with an optimal allocation – the allocation can always be improved upon by incorporating units of the arbitrage portfolio. More can be said. The solution to the agents' problem can, in fact, be used to identify the risk-neutral probabilities. To see this, let us first rewrite the objective function in Equation (11.12) as follows:

$$\begin{aligned} & \max_{\{n_P^i : i=1,2,\dots,N\}} EU \left((1+r_f) \left\{ V_0 - \sum_{i=1}^N n_P^i q_i^e(0) \right\} + \sum_{i=1}^N n_P^i q_i^e(\theta, 1) \right) \\ &= \max_{\{n_P^i : i=1,2,\dots,N\}} \sum_{j=1}^J \pi_j U \left((1+r_f) \left\{ V_0 + \sum_{i=1}^N n_P^i \frac{q_i^e(\theta_j, 1)}{1+r_f} - \sum_{i=1}^N n_P^i q_i^e(0) \right\} \right) \\ &= \max_{\{n_P^i : i=1,2,\dots,N\}} \sum_{j=1}^J \pi_j U \left((1+r_f) \left\{ V_0 + \sum_{i=1}^N n_P^i \left(\frac{q_i^e(\theta_j, 1)}{1+r_f} - q_i^e(0) \right) \right\} \right) \end{aligned} \quad (11.13)$$

The necessary and sufficient first-order conditions for this problem are of the form:

$$\begin{aligned} 0 &= \sum_{j=1}^J \pi_j U_1 \left((1+r_f) \left\{ V_0 + \sum_{i=1}^N n_P^i \left(\frac{q_i^e(\theta_j, 1)}{1+r_f} - q_i^e(0) \right) \right\} \right) \\ & \quad (1+r_f) \left[\frac{q_i^e(\theta_j, 1)}{1+r_f} - q_i^e(0) \right] \end{aligned} \quad (11.14)$$

Note that the quantity $\pi_j U_1(V_P^1(\theta_j, 1))(1+r_f)$ is strictly positive because $\pi_j > 0$ and $U(\cdot)$ is strictly increasing. If we normalize these quantities we can convert

them into probabilities. Let us define

$$\tilde{\pi}_j = \frac{\pi_j U_1(V_P(\theta_j, 1))(1 + r_f)}{\sum_{j=1}^J \pi_j U_1(V_P(\theta_j, 1))(1 + r_f)} = \frac{\pi_j U_1(V_P(\theta_j, 1))}{\sum_{j=1}^J \pi_j U_1(V_P(\theta_j, 1))}, \quad j = 1, 2, \dots, J.$$

Since $\tilde{\pi}_j > 0$, $j = 1, 2, \dots, J$, $\sum_{j=1}^J \tilde{\pi}_j = 1$, and, by (11.14)

$$q_i^e(0) = \sum_{j=1}^J \tilde{\pi}_j \frac{q_i^e(\theta_j, 1)}{1 + r_f};$$

these three properties establish the set $\{\tilde{\pi}_j : j = 1, 2, \dots, N\}$ as a set of risk-neutral probabilities.

We have just proved one half of the following proposition:

Proposition 11.6:

Let $\{n_{p^*}^b, n_{p^*}^i : i = 1, 2, \dots, N\}$ be the solution to the optimal portfolio problem (11.12). Then the set $\{\pi_j^* : j = 1, 2, \dots, J\}$, defined by

$$\pi_j^* = \frac{\pi_j U_1(V_{p^*}(\theta_j, 1))}{\sum_{j=1}^J \pi_j U_1(V_{p^*}(\theta_j, 1))}, \quad (11.15)$$

constitutes a risk-neutral probability measure on \mathbf{S} . Conversely, if there exists a risk-neutral probability measure $\{\pi_j^{RN} : j = 1, 2, \dots, J\}$ on \mathbf{S} , there must exist a concave, strictly increasing, differentiable utility function $U(\cdot)$ and an initial wealth V_0 for which Equation (11.12) has a solution.

Proof:

We have proved the first part. The proof of the less important converse proposition is relegated to Appendix 11.2.

11.7.2 Solving the Portfolio Problem

Now we can turn to solving Equation (11.12). Since there is as much information in the risk-neutral probabilities as in the security prices, it should be possible to fashion a solution to Equation (11.12) using that latter construct. Here we will choose to restrict our attention to the case in which the financial markets are complete.

In this case there exists exactly one risk-neutral measure, which we denote by $\{\pi_j^{RN} : j = 1, 2, \dots, N\}$. Since the solution to Equation (11.12) will be a portfolio in \mathbf{S} that maximizes the date $t = 1$ expected utility of wealth, the solution procedure can be decomposed into a two-step process:

Step 1: Solve

$$\begin{aligned} & \max EU(\tilde{x}(\theta, 1)) \\ \text{s.t. } & E_{\pi^{RN}} \left(\frac{\tilde{x}(\theta, 1)}{1 + r_f} \right) = V_0 \end{aligned} \quad (11.16)$$

The solution to this problem identifies the feasible uncertain payoff that maximizes the agent's expected utility. But why is the constraint a perfect summary of feasibility? The constraint makes sense first because, under complete markets, every uncertain payoff lies in \mathbf{S} . Furthermore, in the absence of arbitrage opportunities, every payoff is valued at the present value of its expected payoff computed using the unique risk-neutral probability measure. The essence of the budget constraint is that a feasible payoff be affordable: that its price equals V_0 , the agent's initial wealth.

Step 2: Find the portfolio P in \mathbf{S} such that

$$V_P(\theta_j, 1) = x(\theta_j, 1), j = 1, 2, \dots, J.$$

In step 2 we simply find the precise portfolio allocations of fundamental securities that give rise to the optimal uncertain payoff identified in step 1. The theory is all in step 1; in fact, we have used all of our major results thus far to write the constraint in the indicated form.

Now let us work out a problem, first abstractly and then by a numerical example. Equation (11.16) of step 1 can be written as

$$\max E_{\pi} U(\tilde{x}(\theta, 1)) - \lambda [E_{\pi^{RN}} \left(\frac{\tilde{x}(\theta, 1)}{1 + r_f} \right) - V_0], \quad (11.17)$$

where λ denotes the Lagrange multiplier and where we have made explicit the probability distributions with respect to which each of the expectations is being taken.

Equation (11.17) can be rewritten as

$$\max_x \sum_{j=1}^J \pi_j \left[U(x(\theta_j, 1)) - \lambda \frac{\pi_j^{RN}}{\pi_j} \frac{x(\theta_j, 1)}{(1 + r_f)} \right] - \lambda V_0. \quad (11.18)$$

The necessary first-order conditions, one equation for each state θ_j , are thus

$$U_1(x(\theta_j, 1)) = \frac{\lambda \pi_j^{RN}}{\pi_j(1 + r_f)}, j = 1, 2, \dots, J. \quad (11.19)$$

from which the optimal asset payoffs may be obtained as per

$$x(\theta_j, 1) = U_1^{-1} \left(\frac{\lambda \pi_j^{RN}}{\pi_j(1 + r_f)} \right), j = 1, 2, \dots, J \quad (11.20)$$

with U_1^{-1} representing the inverse of the MU function.

The Lagrange multiplier λ is the remaining unknown. It must satisfy the budget constraint when Equation (11.20) is substituted for the solution; that is, λ must satisfy

$$E_{\pi^{RN}} \left(\frac{1}{(1+r_f)} U_1^{-1} \left(\frac{\lambda \pi_j^{RN}}{\pi_j(1+r_f)} \right) \right) = V_0. \quad (11.21)$$

A value for λ that satisfies Equation (11.21) may not exist. For all the standard utility functions that we have dealt with, $U(x) = \ln x$ or $\frac{x^{1-\gamma}}{1-\gamma}$ or $e^{-\nu x}$, however, it can be shown that such a λ will exist. Let $\hat{\lambda}$ solve Equation (11.21); the optimal feasible contingent payoff is thus given by

$$x(\theta_j, 1) = U_1^{-1} \left(\frac{\hat{\lambda} \pi_j^{RN}}{\pi_j(1+r_f)} \right) \quad (11.22)$$

(from (11.21)). Given this payoff, step 2 involves finding the portfolio of fundamental securities that will give rise to it. This is accomplished by solving the customary system of linear equations.

11.7.3 A Numerical Example

Now, a numerical example: Let us choose a utility function from the familiar CRRA class, $U(x) = \frac{x^{1-\gamma}}{1-\gamma}$, and consider the market structure of Example 11.2: Markets are complete and the unique risk-neutral probability measure is as noted.

Since $U_1(x) = x^{-\gamma}$, $U_1^{-1}(y) = y^{-\frac{1}{\gamma}}$, Equation (11.20) reduces to

$$x(\theta_j, 1) = \left(\frac{\lambda \pi_j^{RN}}{\pi_j(1+r_f)} \right)^{-\frac{1}{\gamma}} \quad (11.23)$$

from which follows the counterpart to Equation (11.21):

$$\sum_{j=1}^J \pi_j^{RN} \left(\frac{1}{(1+r_f)} \left(\frac{\lambda \pi_j^{RN}}{\pi_j(1+r_f)} \right)^{-\frac{1}{\gamma}} \right) = V_0.$$

Isolating λ gives

$$\hat{\lambda} = \left\{ \sum_{j=1}^J \pi_j^{RN} \left(\frac{1}{(1+r_f)} \left(\frac{\pi_j^{RN}}{\pi_j(1+r_f)} \right)^{-\frac{1}{\gamma}} \right) \right\}^{\gamma} V_0^{-\gamma}. \quad (11.24)$$

Let us consider some numbers: Assume $\gamma = 3$, $V_0 = 10$, and that (π_1, π_2, π_3) , the true probability distribution, takes on the value $(1/3, 1/3, 1/3)$. Refer to

Example 11.2 where the risk-neutral probability distribution was found to be $(\pi_1^{RN}, \pi_2^{RN}, \pi_3^{RN}) = (.3, .6, .1)$. Accordingly, from (11.24)

$$\begin{aligned}\hat{\lambda} &= 10^{-3} \left\{ .3 \left(\frac{1}{(1.1)} \left(\frac{.3}{(1/3)(1.1)} \right)^{-1/3} \right) \right. \\ &\quad \left. + .6 \left(\frac{1}{(1.1)} \left(\frac{.6}{(1/3)(1.1)} \right)^{-1/3} \right) + .1 \left(\frac{1}{(1.1)} \left(\frac{.1}{(1/3)(1.1)} \right)^{-1/3} \right) \right\}^3 \\ \hat{\lambda} &= \left(\frac{1}{1000} \right) \{ .2916 + .4629 + .14018 \}^3 = .0007161.\end{aligned}$$

The distribution of the state-contingent payoffs follows from (11.23):

$$x(\theta_j, 1) = \left(\frac{.0007161 \pi_j^{RN}}{\pi_j(1 + r_f)} \right)^{-\frac{1}{\gamma}} = \begin{cases} 11.951 & j = 1 \\ 9.485 & j = 2 \\ 17.236 & j = 3. \end{cases} \quad (11.25)$$

The final step is to convert this payoff to a portfolio structure via the identification:

$$\begin{aligned}(11.951, 11.485, 17.236) &= n_P^b(1.1, 1.1, 1.1) + n_P^1(3, 2, 1) + n_P^2(1, 4, 6) \text{ or} \\ 11.951 &= 1.1n_P^b + 3n_P^1 + n_P^2 \\ 11.485 &= 1.1n_P^b + 2n_P^1 + 4n_P^2 \\ 17.236 &= 1.1n_P^b + n_P^1 + 6n_P^2\end{aligned}$$

The solution to this system of equations is

$$\begin{aligned}n_P^b &= 97.08 \text{ (invest a lot in the risk-free asset)} \\ n_P^1 &= -28.192 \text{ (short the first stock)} \\ n_P^2 &= -10.225 \text{ (also short the second stock)}\end{aligned}$$

Lastly, we confirm that this portfolio is feasible:

Cost of portfolio = $97.08 + 2(-28.192) + 3(-10.225) = 10 = V_0$, the agent's initial wealth, as required.

Note the computational simplicity of this method: We need only solve a linear system of equations. Using more standard methods would result in a system of three nonlinear equations to solve. Analogous methods are also available to provide bounds in the case of market incompleteness.

11.8 Conclusions

Under the procedure of risk-neutral valuation, we construct a new probability distribution – the risk-neutral probabilities – under which all assets may be valued at their expected payoff discounted at the risk-free rate. More formally

it would be said that we undertake a *transformation of measure* by which all assets are then expected to earn the risk-free rate. The key to our ability to find such a measure is that the financial markets exhibit no arbitrage opportunities.

Our setting was the standard Arrow-Debreu two-period equilibrium and we observed the intimate relationship between the risk-neutral probabilities and the relative prices of state claims. Here the practical applicability of the idea is limited. Applying these ideas to the real world would, after all, require a denumeration of all future states of nature and the contingent payoffs to all securities in order to compute the relevant risk-neutral probabilities, something for which there would be no general agreement.

Even so, this particular way of approaching the optimal portfolio problem was shown to be a source of useful insights. In more restrictive settings, it is also practically powerful and, as noted in Chapter 10, lies behind all modern derivatives pricing.

References

- Harrison, M., Kreps, D. (1979), “Martingales and Multi-Period Securities Market,” *Journal of Economic Theory* 20, 381–408.
- Pliska, S. R. (1997), *Introduction to Mathematical Finance: Discrete Time Models*, Basil Blackwell, Malden, Mass.

Appendix 11.1: Finding the Stock and Bond Economy That Is Directly Analogous to the Arrow-Debreu Economy in Which Only State Claims Are Traded

The Arrow-Debreu economy is summarized in Table 11.6. We wish to price the stock and bond with the payoff structures in Table A 11.1.

Table A11.1 Payoff Structure

$t = 0$	$t = 1$	
	θ_1	θ_2
$-q^e(0)$	1	3
$-q^b(0)$	1	1

In order for the economy in which the stock and bond are traded to be equivalent to the Arrow-Debreu economy where state claims are traded, we need the former to imply the same effective endowment structure. This is accomplished as follows:

Agent 1: Let his endowments of the stock and bond be denoted by \hat{z}_1^e and \hat{z}_1^b , then,

$$\text{In state } \theta_1: \hat{z}_1^b + \hat{z}_1^e = 1$$

$$\text{In state } \theta_2: \hat{z}_1^b + 3\hat{z}_1^e = 2$$

$$\text{Solution: } \hat{z}_1^e = \hat{z}_1^b = 1/2 \text{ (half a share and half a bond).}$$

Agent 2: Let his endowments of the stock and bond be denoted by \hat{z}_2^e and \hat{z}_2^b , then,

$$\text{In state } \theta_1: \hat{z}_2^b + \hat{z}_2^e = 4$$

$$\text{In state } \theta_2: \hat{z}_2^b + 3\hat{z}_2^e = 6$$

$$\text{Solution: } \hat{z}_2^e = 1, \hat{z}_2^b = 3$$

With these endowments the decision problems of the agent become:

Agent 1:

$$\max_{z_1^e, z_1^b} \frac{1}{2} \left(10 + \frac{1}{2}q^e + \frac{1}{2}q^b - z_1^e q^e - z_1^b q^b \right) + .9 \left(\frac{1}{3} \ln(z_1^e + z_1^b) + \frac{2}{3} \ln(3z_1^e + z_1^b) \right)$$

Agent 2:

$$\max_{z_2^e, z_2^b} \frac{1}{2} (5 + q^e + 3q^b - z_2^e q^e - z_2^b q^b) + .9 \left(\frac{1}{3} \ln(z_2^e + z_2^b) + \frac{2}{3} \ln(3z_2^e + z_2^b) \right)$$

The FOCs are

$$\begin{aligned} z_1^e : \frac{1}{2}q^e &= .9 \left(\frac{1}{3} \left(\frac{1}{z_1^e + z_1^b} \right) + \frac{2}{3} \left(\frac{1}{3z_1^e + z_1^b} \right) \right) (3) \\ z_1^b : \frac{1}{2}q^b &= .9 \left(\frac{1}{3} \left(\frac{1}{z_1^e + z_1^b} \right) + \frac{2}{3} \left(\frac{1}{3z_1^e + z_1^b} \right) \right) \\ z_2^e : \frac{1}{2}q^e &= .9 \left(\frac{1}{3} \left(\frac{1}{z_2^e + z_2^b} \right) + \frac{2}{3} \left(\frac{1}{3z_2^e + z_2^b} \right) \right) (3) \\ z_2^b : \frac{1}{2}q^b &= .9 \left(\frac{1}{3} \left(\frac{1}{z_2^e + z_2^b} \right) + \frac{2}{3} \left(\frac{1}{3z_2^e + z_2^b} \right) \right) (1) \end{aligned}$$

Since these securities span the space and since the period 1 and period 2 endowments are the same, the real consumption allocations must be the same as in the Arrow-Debreu economy:

$$c_1^1 = c_1^2 = 2.5$$

$$c_2^1 = c_2^2 = 4$$

Thus,

$$q^e = 2(.9) \left\{ \frac{1}{3} \left(\frac{1}{2.5} \right) + \frac{2}{3} \left(\frac{1}{4} \right) 3 \right\} = 1.14$$

$$q^b = 2(.9) \left\{ \frac{1}{3} \left(\frac{1}{2.5} \right) + \frac{2}{3} \left(\frac{1}{4} \right) \right\} = .54,$$

as computed previously.

To compute the corresponding security holding, observe that:

Agent 1:

$$\begin{aligned} z_1^e + z_1^b &= 2.5 \\ 3z_1^e + z_1^b &= 4 \end{aligned} \quad \Rightarrow \quad \begin{aligned} z_1^e &= .75 \\ z_1^b &= 1.75 \end{aligned}$$

Agent 2: (same holdings)

$$\begin{aligned} z_2^e &= .75 \\ z_2^b &= 1.75 \end{aligned}$$

Supply must equal demand in equilibrium:

$$\begin{aligned} \hat{z}_1^e + \hat{z}_2^e &= \frac{1}{2} + 1 = 1.5 = z_1^e + z_2^e \\ \hat{z}_1^b + \hat{z}_2^b &= \frac{1}{2} + 3 = 3.5 = z_1^b + z_2^b \end{aligned}$$

The period zero consumptions are identical to the earlier calculation as well.

Appendix 11.2: Proof of the Second Part of Proposition 11.6

Define $\hat{U}(x, \theta_j) = x \left\{ \frac{\pi_j^{RN}}{\pi_j(1+r_f)} \right\}$, where $\{\pi_j : j = 1, 2, \dots, J\}$ are the true objective state probabilities. This is a state-dependent utility function that is linear in wealth. We will show that for this function, Equation (11.13), indeed, has a solution. Consider an arbitrary allocation of wealth to the various fundamental assets $\{n_P^i : i = 1, 2, \dots, N\}$ and let P denote that portfolio. Fix the wealth at any level V_0 , arbitrary. We next compute the expected utility associated with this portfolio, taking advantage of representation (11.14):

$$\begin{aligned} E \hat{U}(\tilde{V}_P(\theta, 1)) &= E \hat{U} \left\{ (1+r_f) \left[V_0 + \sum_{i=1}^N n_P^i \left(\frac{\tilde{q}_i^e(\theta, 1)}{(1+r_f)} - q_i^e(0) \right) \right] \right\} \\ &= \sum_{j=1}^J \pi_j(1+r_f) \left\{ V_0 + \sum_{i=1}^N n_P^i \left(\frac{q_i^e(\theta_j, 1)}{(1+r_f)} - q_i^e(0) \right) \right\} \frac{\pi_j^{RN}}{\pi_j(1+r_f)} \\ &= \sum_{j=1}^J \pi_j^{RN} \left\{ V_0 + \sum_{i=1}^N n_P^i \left(\frac{q_i^e(\theta_j, 1)}{(1+r_f)} - q_i^e(0) \right) \right\} \\ &= \sum_{j=1}^J \pi_j^{RN} V_0 + \sum_{j=1}^J \pi_j^{RN} \sum_{i=1}^N n_P^i \left(\frac{q_i^e(\theta_j, 1)}{(1+r_f)} - q_i^e(0) \right) \\ &= V_0 + \sum_{i=1}^N n_P^i \left(\sum_{j=1}^J \pi_j^{RN} \left(\frac{q_i^e(\theta_j, 1)}{(1+r_f)} - q_i^e(0) \right) \right) \\ &= V_0 \end{aligned}$$

in other words, with this utility function, every trading strategy has the same value. Thus problem (11.13) has, trivially, a solution.

Chapter 12: The Martingale Measure in Discrete Time: Part II

12.1 Introduction

We return to the notion of risk-neutral valuation, which we now extend to settings with many time periods. This will be accomplished in two very different ways. First, we extend the concept to the CCAPM setting. Recall that this is a discrete time, general equilibrium framework: preferences and endowment processes must be specified and no-trade prices computed. We will demonstrate that, here as well, assets may be priced equal to the present value, discounted at the risk-free rate of interest, of their expected payoffs when expectations are computed using the set of risk-neutral probabilities. We would expect this to be possible. The CCAPM is an equilibrium model (hence there are no arbitrage opportunities and a set of risk-neutral probabilities must exist) with complete markets (hence this set is unique).

Second, we extend the idea to the partial equilibrium setting of equity derivatives (e.g., equity options) valuations. The key to derivatives pricing is to have an accurate model of the underlying price process. We hypothesize such a process (it is not derived from underlying fundamentals - preferences, endowments etc.; rather, it is a pure statistical model), and demonstrate that, in the presence of *local* market completeness and *local* no arbitrage situations, there exists a transformation of measure by which all derivatives written on that asset may be priced equal to the present value, discounted at the risk-free rate, of their expected payoffs computed using this transformed measure.¹ The Black-Scholes formula, for example, may be derived in this way.

12.2 Discrete Time Infinite Horizon Economies: A CCAPM Setting

As in the previous chapter, time evolves according to $t = 0, 1, \dots, T, T + 1, \dots$. We retain the context of a single good endowment economy and presume the existence of a complete markets Arrow-Debreu financial structure. In period t , any one of N_t possible states, indexed by θ_t , may be realized.

We will assume that a period t event is characterized by two quantities:

- (i) the actually occurring period t event as characterized by θ_t ,
- (ii) the unique history of events $(\theta_1, \theta_2, \dots, \theta_{t-1})$ that precedes it.

Requirement (ii), in particular, suggests an evolution of uncertainty similar to that of a tree structure in which the branches never join (two events always have distinct prior histories). While this is a stronger assumption than what underlies the CCAPM, it will allow us to avoid certain notational ambiguities; subsequently, assumption (ii) will be dropped. We are interested more in the

¹By *local* we mean that valuation is considered only in the context of the derivative, the underlying asset (a stock), and a risk-free bond.

idea than in any broad application, so generality is not an important consideration.

Let $\pi(\theta_t, \theta_{t+1})$ represent the probability of state θ_{t+1} being realized in period $t + 1$, given that θ_t is realized in period t . The financial market is assumed to be complete in the following sense: At every date t , and for every state θ_t , there exists a short-term contingent claim that pays one unit of consumption if state θ_{t+1} is realized in period $t + 1$ (and nothing otherwise). We denote the period t , state θ_t price of such a claim by $q(\theta_t, \theta_{t+1})$.

Arrow-Debreu long-term claims (relative to $t = 0$) are not formally traded in this economy. Nevertheless, they can be synthetically created by dynamically trading short-term claims. (In general, more trading can substitute for fewer claims). To illustrate, let $q(\theta_0, \theta_{t+1})$ represent the period $t = 0$ price of a claim to one unit of the numeraire, if and only if event θ_{t+1} is realized in period $t + 1$. It must be the case that

$$q(\theta_0, \theta_{t+1}) = \prod_{s=0}^t q(\theta_s, \theta_{s+1}), \quad (12.1)$$

where $(\theta_0, \dots, \theta_t)$ is the unique prior history of θ_{t+1} . By the uniqueness of the path to θ_{t+1} , $q(\theta_0, \theta_{t+1})$ is well defined. By no-arbitrage arguments, if the long-term Arrow-Debreu security were also traded, its price would conform to Equation (12.1). Arrow-Debreu securities can thus be effectively created via dynamic (recursive) trading, and the resulting financial market structure is said to be dynamically complete.² By analogy, the price in period t , state θ_t , of a security that pays one unit of consumption if state θ_{t+J} is observed in period $t + J$, $q(\theta_t, \theta_{t+J})$, is given by

$$q(\theta_t, \theta_{t+J}) = \prod_{s=t}^{t+J-1} q(\theta_s, \theta_{s+1}).$$

It is understood that θ_{t+J} is feasible from θ_t , that is, given that we are in state t in period t , there is some positive probability for the economy to find itself in state θ_{t+J} in period $t + J$; otherwise the claims price must be zero.

Since our current objective is to develop risk-neutral pricing representations, a natural next step is to define risk-free bond prices and associated risk-free rates. Given the current date-state is (θ_t, t) , the price, $q^b(\theta_t, t)$, of a risk-free one-period (short-term) bond is given by (no arbitrage)

$$q^b(\theta_t, t) = \sum_{\theta_{t+1}=1}^{N_{t+1}} q(\theta_t, \theta_{t+1}); \quad (12.2)$$

Note here that the summation sign applies across all N_{t+1} future states of nature. The corresponding risk-free rate must satisfy

$$(1 + r_f(\theta_t)) = \{q^b(\theta_t, t)\}^{-1}$$

²This fact suggests that financial markets may need to be “very incomplete” if incompleteness per se is to have a substantial effect on equilibrium asset prices and, for example, have a chance to resolve some of the puzzles uncovered in Chapter 9. See Telmer (1993).

Pricing a k -period risk-free bond is similar:

$$q^b(\theta_t, t+k) = \sum_{\theta_{t+k}=1}^{N_{t+k}} q(\theta_t, \theta_{t+k}). \quad (12.3)$$

The final notion is that of an *accumulation factor*, denoted by $g(\theta_t, \theta_{t+k})$, and defined for a specific path $(\theta_t, \theta_{t+1}, \dots, \theta_{t+k})$ as follows:

$$g(\theta_t, \theta_{t+k}) = \prod_{s=t}^{t+k-1} q^b(\theta_s, s+1). \quad (12.4)$$

The idea being captured by the accumulation factor is this: An investor who invests one unit of consumption in short-term risk-free bonds from date t to $t+k$, continually rolling over his investment, will accumulate $[g(\theta_t, \theta_{t+k})]^{-1}$ units of consumption by date $t+k$, if events $\theta_{t+1}, \dots, \theta_{t+k}$ are realized. Alternatively,

$$[g(\theta_t, \theta_{t+k})]^{-1} = \prod_{s=0}^{k-1} (1 + r_f(\theta_{t+s})). \quad (12.5)$$

Note that from the perspective of date t , state θ_t , the factor $[g(\theta_t, \theta_{t+k})]^{-1}$ is an uncertain quantity as the actual state realizations in the succeeding time periods are not known at period t . From the $t=0$ perspective, $[g(\theta_t, \theta_{t+k})]^{-1}$ is in the spirit of a (conditional) forward rate.

Let us illustrate with the two-date forward accumulation factor. We take the perspective of the investor investing one unit of the numeraire in a short-term risk-free bond from date t to $t+2$. His first investment is certain since the current state θ_t is known and it returns $(1 + r_f(\theta_t))$. At date $t+1$, this sum will be invested again in a one-period risk-free bond with return $(1 + r_f(\theta_{t+1}))$ contracted at $t+1$ and received at $t+2$. From the perspective of date t , this is indeed an uncertain quantity. The compounded return on the investment is: $(1 + r_f(\theta_t))(1 + r_f(\theta_{t+1}))$. This is the inverse of the accumulation factor $g(\theta_t, \theta_{t+1})$ as spelled out in Equation (12.5).

Let us next translate these ideas directly into the CCAPM settings.

12.3 Risk-Neutral Pricing in the CCAPM

We make two additional assumptions in order to restrict our current setting to the context of the CCAPM.

A12.1: There is one agent in the economy with time-separable VNM preferences represented by

$$\mathbb{U}(\tilde{c}) = E_0 \left(\sum_{t=0}^{\infty} U(\tilde{c}_t, t) \right),$$

where $U(\tilde{c}_t, t)$ is a family of strictly increasing, concave, differentiable period utility functions, with $U_1(c_t, t) > 0$ for all t , $\tilde{c}_t = c(\theta_t)$ is the uncertain period t consumption, and E_0 the expectations operator conditional on date $t = 0$ information.

This treatment of the agent's preferences is quite general. For example, $U(c_t, t)$ could be of the form $\delta^t U(c_t)$ as in earlier chapters. Alternatively, the period utility function could itself be changing through time in deterministic fashion, or some type of habit formation could be postulated. In all cases, it is understood that the set of feasible consumption sequences will be such that the sum exists (is finite).

A12.2: Output in this economy, $\tilde{Y}_t = Y_t(\theta_t)$ is exogenously given, and, by construction, represents the consumer's income. In equilibrium it represents his consumption as well.

Recall that equilibrium-contingent claims prices in the CCAPM economy are no-trade prices, supporting the consumption sequences $\{\tilde{c}_t\}$ in the sense that at these prices, the representative agent does not want to purchase any claims; that is, at the prevailing contingent-claims prices his existing consumption sequence is optimal. The loss in period t utility experienced by purchasing a contingent claim $q(\theta_t, \theta_{t+1})$ is exactly equal to the resultant increase in expected utility in period $t + 1$. There is no benefit to further trade. More formally,

$$\begin{aligned} U_1(c(\theta_t), t)q(\theta_t, \theta_{t+1}) &= \pi(\theta_t, \theta_{t+1})U_1(c(\theta_{t+1}), t+1), \text{ or} \\ q(\theta_t, \theta_{t+1}) &= \pi(\theta_t, \theta_{t+1}) \left\{ \frac{U_1(c(\theta_{t+1}), t+1)}{U_1(c(\theta_t), t)} \right\}. \end{aligned} \quad (12.6)$$

Equation (12.7) corresponds to Equation (8.1) of Chapter 8. State probabilities and inter-temporal rates of substitution appear once again as the determinants of equilibrium Arrow-Debreu prices. Note that the more general utility specification adopted in this chapter does not permit bringing out explicitly the element of time discounting embedded in the inter-temporal marginal rates of substitution. A short-term risk-free bond is thus priced according to

$$q^b(\theta_t, t+1) = \sum_{\theta_{t+1}=1}^{N_t} q(\theta_t, \theta_{t+1}) = \frac{1}{U_1(c(\theta_t), t)} E_t \{U_1(c(\theta_{t+1}), t+1)\}. \quad (12.7)$$

Risk-neutral valuation is in the spirit of discounting at the risk-free rate. Accordingly, we may ask: At what probabilities must we compute the expected payoff to a security in order to obtain its price by discounting that payoff at the risk-free rate? But which risk-free rates are we speaking about? In a multiperiod context, there are two possibilities and the alternative we choose will govern the precise form of the probabilities themselves.

The spirit of the dilemma is portrayed in Figure 12.1, which illustrates the case of a $t = 3$ period cash flow.

Insert Figure 12.1 about here

Under the first alternative, the cash flow is discounted at a series of consecutive short (one-period) rates, while in the second we discount back at the term structure of multiperiod discount bonds. These methods provide the same price, although the form of the risk-neutral probabilities will differ substantially. Here we offer a discussion of alternative 1; alternative 2 is considered in Appendix 12.1.

Since the one-period state claims are the simplest securities, we will first ask what the risk-neutral probabilities must be in order that they be priced equal to the present value of their expected payoff, discounted at the risk-free rate.³ As before, let these numbers be denoted by $\pi^{RN}(\theta_t, \theta_{t+1})$. They are defined by:

$$q(\theta_t, \theta_{t+1}) = \pi(\theta_t, \theta_{t+1}) \left\{ \frac{U_1(c(\theta_{t+1}), t+1)}{U_1(c(\theta_t), t)} \right\} = q^b(\theta_t, t+1) [\pi^{RN}(\theta_t, \theta_{t+1})].$$

The second equality reiterates the tight relationship found in Chapter 11 between Arrow-Debreu prices and risk-neutral probabilities. Substituting Equation (12.7) for $q^b(\theta_t, t+1)$ and rearranging terms, one obtains:

$$\begin{aligned} \pi^{RN}(\theta_t, \theta_{t+1}) &= \pi(\theta_t, \theta_{t+1}) \left\{ \frac{U_1(c(\theta_{t+1}), t+1)}{U_1(c(\theta_t), t)} \right\} \frac{U_1(c(\theta_t), t)}{E_t\{U_1(c(\theta_{t+1}), t+1)\}} \\ &= \pi(\theta_t, \theta_{t+1}) \left\{ \frac{U_1(c(\theta_{t+1}), t+1)}{E_t U_1(c(\theta_{t+1}), t+1)} \right\}. \end{aligned} \quad (12.8)$$

Since $U(c(\theta), t)$ is assumed to be strictly increasing, $U_1 > 0$ and $\pi^{RN}(\theta_t, \theta_{t+1}) > 0$ (without loss of generality we may assume $\pi(\theta_t, \theta_{t+1}) > 0$). Furthermore, by construction, $\sum_{\theta_{t+1}=1}^{N_{t+1}} \pi^{RN}(\theta_t, \theta_{t+1}) = 1$. The set $\{\pi^{RN}(\theta_t, \theta_{t+1})\}$ thus defines a set of conditional (on θ_t) risk-neutral transition probabilities. As in our earlier more general setting, if the representative agent is risk neutral, $U_1(c(\theta_t), t) \equiv$ constant for all t , and $\pi^{RN}(\theta_t, \theta_{t+1})$ coincides with $\pi(\theta_t, \theta_{t+1})$, the true probability. Using these transition probabilities, expected future consumption flows may be discounted at the intervening risk-free rates. Notice how the risk-neutral probabilities are related to the true probabilities: They represent the true probabilities scaled up or down by the relative consumption scarcities in the different states. For example, if, for some state θ_{t+1} , the representative agent's consumption is usually low, his marginal utility of consumption in that state will be much higher than average marginal utility and thus

$$\pi^{RN}(\theta_t, \theta_{t+1}) = \pi(\theta_t, \theta_{t+1}) \left\{ \frac{U_1(c(\theta_{t+1}), t+1)}{E_t U_1(c(\theta_{t+1}), t+1)} \right\} > \pi(\theta_t, \theta_{t+1}).$$

³Recall that since all securities can be expressed as portfolios of state claims, we can use the state claims alone to construct the risk-neutral probabilities.

The opposite will be true if a state has a relative abundance of consumption. When we compute expected payoffs to assets using risk-neutral probabilities we are thus implicitly taking into account both the (no-trade) relative equilibrium scarcities (prices) of their payoffs and their objective relative scarcities. This allows discounting at the risk-free rate: No further risk adjustment need be made to the discount rate as all such adjustments have been implicitly undertaken in the expected payoff calculation.

To gain a better understanding of this notion let us go through a few examples.

Example 12.1

Denote a stock's associated dividend stream by $\{d(\theta_t)\}$. Under the basic state-claim valuation perspective (Chapter 10, Section 10.2), its ex-dividend price at date t , given that θ_t has been realized, is:

$$q^e(\theta_t, t) = \sum_{s=t+1}^{\infty} \sum_{j=1}^{N_s} q(\theta_t, \theta_s(j)) d(\theta_s(j)), \quad (12.9)$$

or, with a recursive representation,

$$q^e(\theta_t, t) = \sum_{\theta_{t+1}} q(\theta_t, \theta_{t+1}) \{q^e(\theta_{t+1}, t+1) + d(\theta_{t+1})\} \quad (12.10)$$

Equation (12.10) may also be expressed as

$$q^e(\theta_t, t) = q^b(\theta_t, t+1) E_t^{RN} \{q^e(\tilde{\theta}_{t+1}, t+1) + d(\tilde{\theta}_{t+1})\}, \quad (12.11)$$

where E_t^{RN} denotes the expectation taken with respect to the relevant risk-neutral transition probabilities; equivalently,

$$q^e(\theta_t, t) = \frac{1}{1 + r_f(\theta_t)} E_t^{RN} \{q^e(\tilde{\theta}_{t+1}, t+1) + d(\tilde{\theta}_{t+1})\}.$$

Returning again to the present value expression, Equation (12.9), we have

$$\begin{aligned} q^e(\theta_t, t) &= \sum_{s=t+1}^{\infty} E_t^{RN} \{g(\theta_t, \tilde{\theta}_s) d(\tilde{\theta}_s)\} \\ &= \sum_{s=t+1}^{\infty} E_t^{RN} \frac{d(\tilde{\theta}_s)}{\prod_{j=0}^{s-1} (1 + r_f(\tilde{\theta}_{t+j}))}. \end{aligned} \quad (12.12)$$

What does Equation (12.12) mean? Any state $\hat{\theta}_s$ in period $s \geq t+1$ has a unique sequence of states preceding it. The product of the risk-neutral transition probabilities associated with the states along the path defines the (conditional) risk-neutral probability of $\hat{\theta}_s$ itself. The product of this probability and the

payment as $d(\hat{\theta}_s)$ is then discounted at the associated accumulation factor – the present value factor corresponding to the risk-free rates identified with the succession of states preceding $\hat{\theta}_s$. For each $s \geq t + 1$, the expectation represents the sum of all these terms, one for each θ_s feasible from θ_t .

Since the notational intensity tends to obscure what is basically a very straightforward idea, let us turn to a small numerical example.

Example 12.2

Let us value a two-period equity security, where $U(c_t, t) \equiv U(c_t) = \ln c_t$ for the representative agent (no discounting). The evolution of uncertainty is given by Figure 12.2 where

$$\begin{aligned}\pi(\theta_0, \theta_{1,1}) &= .6 & \pi(\theta_{1,1}, \theta_{2,1}) &= .3 \\ \pi(\theta_0, \theta_{1,2}) &= .4 & \pi(\theta_{1,1}, \theta_{2,2}) &= .7 \\ & & \pi(\theta_{1,2}, \theta_{2,3}) &= .6 \\ & & \pi(\theta_{1,2}, \theta_{2,4}) &= .4\end{aligned}$$

Insert Figure 12.2 about here

The consumption at each node, which equals the dividend, is represented as the quantity in parentheses. To value this asset risk neutrally, we consider three stages.

1. Compute the (conditional) risk-neutral probabilities at each node.

$$\begin{aligned}\pi^{RN}(\theta_0, \theta_{1,1}) &= \pi(\theta_0, \theta_{1,1}) \left\{ \frac{U_1(c(\theta_{1,1}))}{E_0\{U_1(c_1(\tilde{\theta}_1))\}} \right\} = \frac{(\frac{1}{5})}{(.6(\frac{1}{5}) + .4(\frac{1}{3}))} = .4737 \\ \pi^{RN}(\theta_0, \theta_{1,2}) &= 1 - \pi^{RN}(\theta_0, \theta_{1,1}) = .5263 \\ \pi^{RN}(\theta_{1,1}, \theta_{2,1}) &= \pi(\theta_{1,1}, \theta_{2,1}) \left\{ \frac{\frac{1}{4}}{(.3(\frac{1}{4}) + .7(\frac{1}{2}))} \right\} = .1765 \\ \pi^{RN}(\theta_{1,1}, \theta_{2,2}) &= 1 - \pi^{RN}(\theta_{1,1}, \theta_{2,1}) = .8235 \\ \pi^{RN}(\theta_{1,2}, \theta_{2,4}) &= \pi(\theta_{1,2}, \theta_{2,4}) \left\{ \frac{1}{(.6(\frac{1}{8}) + .4(1))} \right\} = .8421 \\ \pi^{RN}(\theta_{1,2}, \theta_{2,3}) &= .1579\end{aligned}$$

2. Compute the conditional bond prices.

$$\begin{aligned}q^b(\theta_0, 1) &= \frac{1}{U_1(c_0)} E_0\{U_1(c_1(\tilde{\theta}))\} = \frac{1}{(\frac{1}{2})} \left\{ .6(\frac{1}{5}) + .4(\frac{1}{3}) \right\} = .5066 \\ q^b(\theta_{1,1}, 2) &= \frac{1}{(\frac{1}{5})} \left\{ .3(\frac{1}{4}) + .7(\frac{1}{2}) \right\} = 2.125 \\ q^b(\theta_{1,2}, 2) &= \frac{1}{(\frac{1}{3})} \left\{ .6(\frac{1}{8}) + .4(\frac{1}{1}) \right\} = 1.425\end{aligned}$$

3. Value the asset.

$$\begin{aligned}
q^e(\theta_0, 0) &= \sum_{s=1}^2 E_0^{RN} \{g(\theta_0, \tilde{\theta}_s) d_s(\tilde{\theta}_s)\} \\
&= q^b(\theta_0, 1) \{ \pi^{RN}(\theta_0, \theta_{1,1})(5) + \pi^{RN}(\theta_0, \theta_{1,2})(3) \} \\
&+ q^b(\theta_0, 1) q^b(\theta_{1,1}, 2) \{ \pi^{RN}(\theta_0, \theta_{1,1}) \pi^{RN}(\theta_{1,1}, \theta_{2,1})(4) \\
&+ \pi^{RN}(\theta_0, \theta_{1,1}) \pi^{RN}(\theta_{1,1}, \theta_{2,2})(2) \} \\
&+ q^b(\theta_0, 1) q^b(\theta_{1,2}, 2) \{ \pi^{RN}(\theta_0, \theta_{1,2}) \pi^{RN}(\theta_{1,2}, \theta_{2,3})(8) \\
&+ \pi^{RN}(\theta_0, \theta_{1,2}) \pi^{RN}(\theta_{1,2}, \theta_{2,4})(1) \} \\
&= 4.00
\end{aligned}$$

At a practical level this appears to be a messy calculation at best, but it is not obvious how we might compute the no-trade equilibrium asset prices more easily. The Lucas tree methodologies, for example, do not apply here as the setting is not infinitely recursive. This leaves us to solve for the equilibrium prices by working back through the tree and solving for the no-trade prices at each node. It is not clear that this will be any less involved.

Sometimes, however, the risk-neutral valuation procedure does allow for a very succinct, convenient representation of specific asset prices or price interrelationship. A case in point is that of a long-term discount bond.

Example 12.3

To price at time t , state θ_t , a long-term discount bond maturing in date $t+k$, observe that the corresponding dividend $d_{t+k}(\theta_{t+k}) \equiv 1$ for every θ_{t+k} feasible from state θ_t . Applying Equation (12.12) yields

$$q^b(\theta_t, t+k) = E_t^{RN} g(\theta_t, \tilde{\theta}_{t+k}), \text{ or}$$

$$\frac{1}{(1+r(\theta_t, t+k))^k} = E_t^{RN} \left\{ \frac{1}{\prod_{s=t}^{t+k-1} (1+r(\theta_s, s+1))} \right\} \quad (12.13)$$

Equation (12.13), in either of its forms, informs us that the long term rate is the expectation of the short rates taken with respect to the risk-neutral transition probabilities. This is generally not true if the expectation is taken with the ordinary or true probabilities.

At this point we draw this formal discussion to a close. We now have an idea what risk-neutral valuation might mean in a CCAPM context. Appendix 12.1 briefly discusses the second valuation procedure and illustrates it with the pricing of call and put options.

We thus see that the notion of risk-neutral valuation carries over easily to a CCAPM context. This is not surprising: The key to the existence of a set of risk-neutral probabilities is the presence of a complete set of securities markets,

which is the case with the CCAPM. In fact, the somewhat weaker notion of dynamic completeness was sufficient.

We next turn our attention to equity derivatives pricing. The setting is much more specialized and not one of general equilibrium (though not inconsistent with it). One instance of this specialization is that the underlying stock's price is presumed to follow a specialized stochastic process. The term structure is also presumed to be flat. These assumptions, taken together, are sufficient to generate the existence of a unique risk-neutral probability measure, which can be used to value any derivative security written on the stock. That these probabilities are uniquely identified with the specific underlying stock has led us to dub them *local*.

12.4 The Binomial Model of Derivatives Valuation

Under the binomial abstraction we imagine a many-period world in which, at every date-state node only a stock and a bond are traded. With only two securities to trade, dynamic completeness requires that at each node there be only two possible succeeding states. For simplicity, we will also assume that the stock pays no dividend, in other words, that $d(\theta_t) \equiv 0$ for all $t \leq T$. Lastly, in order to avoid any ambiguity in the risk-free discount factors, it is customary to require that the risk-free rate be constant across all dates and states. We formalize these assumptions as follows:

A12.3: The risk-free rate is constant;

$$q^b(\theta_t, t+1) = \frac{1}{1+r_f} \text{ for all } t \leq T.$$

A12.4: The stock pays no dividends: $d(\theta_t) \equiv 0$ for all $t \leq T$.

A12.5: The rate of return to stock ownership follows an i.i.d. process of the form:

$$q^e(\theta_{t+1}, t+1) = \begin{cases} u q^e(\theta_t, t), & \text{with probability } \pi \\ d q^e(\theta_t, t), & \text{with probability } 1 - \pi, \end{cases}$$

where u (up) and d (down) represent gross rates of return. In order to preclude the existence of an arbitrage opportunity it must be the case that

$$u > R_f > d,$$

where, in this context, $R_f = 1 + r_f$.

There are effectively only two possible future states in this model ($\theta_t \in \{\theta_1, \theta_2\}$ where θ_1 is identified with u and θ_2 identified with d) and thus the evolution of the stock's price can be represented by a simple tree structure as seen in Figure 12.3.

Insert Figure 12.3 about here

Why such a simple setting should be of use is not presently clear but it will become so shortly.

In this context, the risk-neutral probabilities can be easily computed from Equation (12.11) specialized to accommodate $d(\theta_t) \equiv 0$:

$$\begin{aligned} q^e(\theta_t, t) &= q^b(\theta_t, t+1) E_t^{RN} \{q^e(\tilde{\theta}_{t+1}, t+1)\} \\ &= q^b(\theta_t, t+1) \{\pi^{RN} u q^e(\theta_t, t) + (1 - \pi^{RN}) d q^e(\theta_t, t)\} \end{aligned} \quad (12.14)$$

This implies

$$\begin{aligned} R_f &= \pi^{RN} u + (1 - \pi^{RN}) d, \text{ or} \\ \pi^{RN} &= \frac{R_f - d}{u - d}. \end{aligned} \quad (12.15)$$

The power of this simple context is made clear when comparing Equation (12.15) with Equation (12.8). Here risk-neutral probabilities can be expressed without reference to marginal rates of substitution, that is, to agents' preferences.⁴ This provides an immense simplification, which all derivative pricing will exploit in one way or another. Of course the same is true for one-period Arrow-Debreu securities since they are priced equal to the present value of their respective risk-neutral probabilities:

$$\begin{aligned} q(\theta_t, \theta_{t+1} = u) &= \left(\frac{1}{R_f} \right) \left(\frac{R_f - d}{u - d} \right), \text{ and} \\ q(\theta_t, \theta_{t+1} = d) &= \left(\frac{1}{R_f} \right) \left(1 - \frac{R_f - d}{u - d} \right) = \left(\frac{1}{R_f} \right) \left(\frac{u - R_f}{u - d} \right). \end{aligned}$$

Furthermore, since the risk-free rate is assumed constant in every period, the price of a claim to one unit of the numeraire to be received $T - t > 1$ periods from now if state θ_T is realized is given by

$$q(\theta_t, \theta_T) = \frac{1}{(1 + r_f(\theta_t, T))^{T-t}} \sum_{\{\theta_t, \dots, \theta_{T-1}\} \in \Omega} \prod_{s=t}^{T-1} \pi^{RN}(\theta_s, \theta_{s+1}),$$

where Ω represents the set of all time paths $\{\theta_t, \theta_{t+1}, \dots, \theta_{T-1}\}$ leading to θ_T . In the binomial setting this becomes

$$q(\theta_t, \theta_T) = \frac{1}{(R_f)^{T-t}} \binom{T-t}{s} (\pi^{RN})^s (1 - \pi^{RN})^{T-t-s}, \quad (12.16)$$

where s is the number of intervening periods in which the u state is observed on any path from θ_t to θ_T . The expression $\binom{T-t}{s}$ represents the number of ways s successes (u moves) can occur in $T-t$ trials. A standard result states

$$\binom{T-t}{s} = \frac{(T-t)!}{s!(T-t-s)!}.$$

⁴Notice that the risk-neutral probability distribution is i.i.d. as well.

The explanation is as follows. Any possible period T price of the underlying stock will be identified with a unique number of u and d realizations. Suppose, for example, that $s_1 u$ realizations are required. There are then $\binom{T-t}{s_1}$ possible paths, each of which has exactly $s_1 u$ and $T-t-s_1 d$ states, leading to the pre-specified period T price. Each path has the common risk-neutral probability $(\pi^{RN})^{s_1} (1-\pi^{RN})^{T-t-s_1}$. As an example, suppose $T-t=3$, and the particular final price is the result of 2 up-moves and 1 down-move. Then, there are $3 = \frac{3!}{2!1!} = \frac{3 \cdot 2 \cdot 1}{(2 \cdot 1)(1)}$ possible paths leading to that final state: uud, udu , and duu .

To illustrate the simplicity of this setting we again consider several examples.

Example 12.4

A European call option revisited: Let the option expire at $T > t$; the price of a European call with exercise price K , given the current date-state (θ_t, t) , is

$$\begin{aligned} C_{\mathcal{E}}(\theta_t, t) &= \left(\frac{1}{R_f}\right)^{T-t} E_t^{RN}(\max\{q^e(\theta_T, T) - K, 0\}) \\ &= \left(\frac{1}{R_f}\right)^{T-t} \sum_{s=0}^{T-t} \binom{T-t}{s} (\pi^{RN})^s (1-\pi^{RN})^{T-t-s} (\max\{q^e(\theta_t, t) u^s d^{T-t-s} - K, 0\}) \end{aligned}$$

When taking the expectation we sum over all possible values of $s \leq T-t$, thus weighting each possible option payoff by the risk-neutral probability of attaining it.

Define the quantity \hat{s} as the minimum number of intervening up states necessary for the underlying asset, the stock, to achieve a price in excess of K . The prior expression can then be simplified to:

$$C_{\mathcal{E}}(\theta_t, t) = \frac{1}{(R_f)^{T-t}} \sum_{s=\hat{s}}^{T-t} \binom{T-t}{s} (\pi^{RN})^s (1-\pi^{RN})^{T-t-s} [q^e(\theta_t, t) u^s d^{T-t-s} - K], \quad (12.17)$$

or

$$\begin{aligned} C_{\mathcal{E}}(\theta_t, t) &= \frac{1}{(R_f)^{T-t}} \sum_{s=\hat{s}}^{T-t} \binom{T-t}{s} (\pi^{RN})^s (1-\pi^{RN})^{T-t-s} q^e(\theta_t, t) u^s d^{T-t-s} \\ &\quad - \sum_{s=\hat{s}}^{T-t} \binom{T-t}{s} (\pi^{RN})^s (1-\pi^{RN})^{T-t-s} K \quad (12.18) \end{aligned}$$

The first term within the braces of Equation (12.18) is the risk-neutral expected value at expiration of the acquired asset if the option is exercised, while

the second term is the risk-neutral expected cost of acquiring it. The difference is the risk-neutral expected value of the call's payoff (value) at expiration.⁵ To value the call today, this quantity is then put on a present value basis by discounting at the risk-free rate R_f .

This same valuation can also be obtained by working backward, recursively, through the tree. Since markets are complete, in the absence of arbitrage opportunities any asset – the call included – is priced equal to its expected value in the succeeding time period discounted at R_f . This implies

$$C_{\mathcal{E}}(\theta_t, t) = q^b(\theta_t, t) E^{RN} C_{\mathcal{E}}(\tilde{\theta}_{t+1}, t+1). \quad (12.19)$$

Let us next illustrate how this fact may be used to compute the call's value in a simple three-period example.

Example 12.5

Let $u = 1.1$, $d = \frac{1}{u} = .91$, $q^e(\theta_t, t) = \$50$, $K = \$53$, $R_f = 1.05$, $T - t = 3$.

$$\pi^{RN} = \frac{R_f - d}{u - d} = \frac{1.05 - .91}{1.1 - .91} = .70$$

Insert Figure 12.4 about here

The numbers in parentheses in Figure 12.4 are the recursive values of the call, working backward in the manner of Equation (12.19). These are obtained as follows:

$$\begin{aligned} C_{\mathcal{E}}(u^2, t+2) &= \frac{1}{1.05} \{ .70(13.55) + .30(2) \} = 9.60 \\ C_{\mathcal{E}}(ud, t+2) &= \frac{1}{1.05} \{ .70(2) + .30(0) \} = 1.33 \\ C_{\mathcal{E}}(u, t+1) &= \frac{1}{1.05} \{ .70(9.60) + .30(1.33) \} = 6.78 \\ C_{\mathcal{E}}(d, t+1) &= \frac{1}{1.05} \{ .70(1.33) + .30(0) \} = .89 \\ C_{\mathcal{E}}(\theta_t, t) &= \frac{1}{1.05} \{ .70(6.78) + .30(.89) \} = 4.77 \end{aligned}$$

For a simple call, its payoff at expiration is dependent only upon the value of the underlying asset (relative to K) at that time, irrespective of its price history. For example, the value of the call when $q^e(\theta_T, T) = 55$ is the same if the price history is $(50, 55, 50, 55)$ or $(50, 45.5, 50, 55)$.

For other derivatives, however, this is not the case; they are *path dependent*. An Asian (path-dependent) option is a case in point. Nevertheless, the same

⁵Recall that there is no actual transfer of the security. Rather, this difference $q^e(\theta_T, T) - K$ represents the amount of money the writer (seller) of the call must transfer to the buyer at the expiration date if the option is exercised.

valuation methods apply: Its expected payoff is computed using the risk-neutral probabilities, and then discounted at the risk-free rate.

Example 12.6

A path dependent option: We consider an Asian option for which the payoff pattern assumes the form outlined in Table 12.1.

Table 12.1 Payoff Pattern – Asian Option

t	$t + 1$	$t + 2$	\dots	$T - 1$	T
0	0	0		0	$\max \{q_e^{AVG}(\theta_T, T) - K, 0\}$,

where $q_e^{AVG}(\theta_T, T)$ is the average price of the stock along the path from $q^e(\theta_t, t)$ to, and including, $q^e(\theta_T, T)$. We may express the period t value of such an option as

$$C_A(\theta_t, t) = \frac{1}{(R_f)^{T-t}} E_t^{RN} \max\{q_e^{AVG}(\theta_T, T) - K, 0\}$$

A simple numerical example with $T - t = 2$ follows. Let $q^e(\theta_t, t) = 100$, $K = 100$, $u = 1.05$, $d = \frac{1}{u} = .95$, and $R_f = 1.005$. The corresponding risk-neutral probabilities are

$$\pi^{RN} = \frac{R_f - d}{u - d} = \frac{1.005 - .95}{1.05 - .95} = .55; 1 - \pi^{RN} = .45$$

With two periods remaining, the possible evolutions of the stock's price and corresponding option payoffs are those found in Figure 12.5.

Insert Figure 12.5 about here

Thus,

$$\begin{aligned} C_A(\theta_t, t) &= \frac{1}{(1.005)^2} \{(.55)^2(5.083) + (.55)(.45)(1.67)\} \\ &= \$1.932 \end{aligned}$$

Note that we may as well work backward, recursively, in the price/payoff tree as shown in Figure 12.6.

Insert Figure 12.6 about here

$$\begin{aligned} \text{where } C_A(\theta_{t+1} = u, t + 1) = 3.53 &= \frac{1}{(1.005)} \{55(5.083) + .45(1.67)\}, \text{ and} \\ C_A(\theta_t, t) &= \frac{1}{(1.005)} \{.55(3.53) + .45(0)\} = \$1.932. \end{aligned}$$

A number of fairly detailed comments are presently in order. Note that with a path-dependent option it is not possible to apply, naively, a variation on Equation (12.18). Unlike with straightforward calls, the value of this type of option is not the same for all paths leading to the same final-period asset price.

Who might be interested in purchasing such an option? For one thing, they have payoff patterns similar in spirit to an ordinary call, but are generally less expensive (there is less upward potential in the average than in the price itself). This feature has contributed to the usefulness of path-dependent options in foreign exchange trading. Consider a firm that needs to provide a stream of payments (say, perhaps, for factory construction) in a foreign currency. It would want protection against a rise in the value of the foreign currency relative to its own, because such a rise would increase the cost of the payment stream in terms of the firm's own currency. Since many payments are to be made, what is of concern is the average price of the foreign currency rather than its price at any specific date. By purchasing the correct number of Asian calls on the foreign currency, the firm can create a payment for itself if, on average, the foreign currency's value exceeds the strike price – the level above which the firm would like to be insured. By analogous reasoning, if the firm wished to protect the average value of a stream of payments it was receiving in a foreign currency, the purchase of Asian puts would be one alternative.

We do not want to lose sight of the fact that risk-neutral valuation is a direct consequence of the dynamic completeness (at each node there are two possible future states and two securities available for trade) and the no-arbitrage assumption, a connection that is especially apparent in the binomial setting. Consider a call option with expiration one period from the present. Over this period the stock's price behavior and the corresponding payoffs to the call option are as found in Figure 12.7.

Insert Figure 12.7 about here

By the assumed dynamic completeness we know that the payoff to the option can be replicated on a state-by-state basis by a position in the stock and the bond. Let this position be characterized by a portfolio of Δ shares and a bond investment of value B (for simplicity of notation we suppress the dependence of these latter quantities on the current state and date). Replication requires

$$\begin{aligned} uq^e(\theta_t, t)\Delta + R_f B &= C_{\mathcal{E}}(u, t+1), \text{ and} \\ dq^e(\theta_t, t)\Delta + R_f B &= C_{\mathcal{E}}(d, t+1), \end{aligned}$$

from which follows

$$\begin{aligned} \Delta &= \frac{C_{\mathcal{E}}(u, t+1) - C_{\mathcal{E}}(d, t+1)}{(u-d)q^e(\theta_t, t)}, \text{ and} \\ B &= \frac{uC_{\mathcal{E}}(d, t+1) - dC_{\mathcal{E}}(u, t+1)}{(u-d)R_f}. \end{aligned}$$

By the no arbitrage assumption:

$$\begin{aligned}
C_{\mathcal{E}}(\theta_t, t) &= \Delta q^e(\theta_t, t) + B \\
&= \left\{ \frac{C_{\mathcal{E}}(u, t+1) - C_{\mathcal{E}}(d, t+1)}{(u-d)q^e(\theta_t, t)} q^e(\theta_t, t) + \frac{uC_{\mathcal{E}}(d, t+1) - dC_{\mathcal{E}}(u, t+1)}{(u-d)R_f} \right\} \\
&= \left(\frac{1}{R_f} \right) \left\{ \left(\frac{R_f - d}{u-d} \right) C_{\mathcal{E}}(u, t+1) + \left(\frac{u - R_f}{u-d} \right) C_{\mathcal{E}}(d, t+1) \right\} \\
&= \left(\frac{1}{R_f} \right) \{ \pi^{RN} C_{\mathcal{E}}(u, t+1) + (1 - \pi^{RN}) C_{\mathcal{E}}(d, t+1) \},
\end{aligned}$$

which is just a specialized case of Equation (12.18).

Valuing an option (or other derivative) using risk-neutral valuation is thus equivalent to pricing its replicating portfolio of stock and debt. Working backward in the tree corresponds to recomputing the portfolio of stock and debt that replicates the derivative's payoffs at each of the succeeding nodes. In the earlier example of the Asian option, the value 3.53 at the intermediate u node represents the value of the portfolio of stocks and bonds necessary to replicate the option's values in the second-period nodes leading from it (5.083 in the u state, 1.67 in the d state).

Let us see how the replicated portfolio evolves in the case of the Asian option written on a stock.

$$\begin{aligned}
\Delta_u &= \frac{C_{\mathcal{A}}(u^2, t+2) - C_{\mathcal{A}}(ud, t+2)}{(u-d)q^e(\theta_t, t)} = \frac{5.083 - 1.67}{(1.05 - .95)(105)} = .325 \\
B_u &= \frac{uC_{\mathcal{A}}(ud, t+2) - dC_{\mathcal{A}}(u^2, t+2)}{(u-d)R_f} = \frac{(1.05)(1.67) - (.95)(5.083)}{(1.05 - .95)(1.005)} = -30.60 \\
\Delta_d &= 0 \\
B_d &= 0 \quad (\text{all branches leading from the "d" node result in zero option value}) \\
\Delta &= \frac{C_{\mathcal{A}}(u, t+1) - C_{\mathcal{A}}(d, t+1)}{(u-d)q^e(\theta_t, t)} = \frac{3.53 - 0}{(1.05 - .95)(100)} = .353 \\
B &= \frac{uC_{\mathcal{A}}(d, t+1) - dC_{\mathcal{A}}(u, t+1)}{(u-d)R_f} = \frac{(1.05)(0) - (.95)(3.53)}{(1.05 - .95)(1.005)} = -33.33
\end{aligned}$$

We interpret these numbers as follows. In order to replicate the value of the Asian option, irrespective of whether the underlying stock's price rises to \$105 or falls to \$95.20, it is necessary to construct a portfolio composed of a loan of \$33.33 at R_f in conjunction with a long position of .353 share. The net cost is

$$.353(100) - 33.33 = \$1.97,$$

the cost of the call, except for rounding errors. To express this idea slightly differently, if you want to replicate, at each node, the value of the Asian option, borrow \$33.33 (at R_f) and, together with your own capital contribution of \$1.97, take this money and purchase .353 share of the underlying stock.

As the underlying stock's value evolves through time, this portfolio's value will evolve so that at any node it matches exactly the call's value. At the first u node, for example, the portfolio will be worth \$3.53. Together with a loan of \$30.60, this latter sum will allow the purchase of .325 share, with no additional capital contribution required. Once assembled, the portfolio is entirely self-financing, no additional capital need be added and none may be withdrawn (until expiration).

This discussion suggests that Asian options represent a levered position in the underlying stock. To see this, note that at the initial node the replicating portfolio consists of a \$1.97 equity contribution by the purchaser in conjunction with a loan of \$30.60. This implies a debt/equity ratio of $\frac{\$30.60}{1.97} \cong 15.5$! For the analogous straight call, with the same exercise price as the Asian and the same underlying price process, the analogous quantities are, respectively, \$3.07 and \$54.47, giving a debt/equity ratio of approximately 18. Call-related securities are thus attractive instruments for speculation! For a relatively small cash outlay, a stock's entire upward potential (within a limited span of time) can be purchased.

Under this pricing perspective there are no arbitrage opportunities within the universe of the underlying asset, the bond, or any derivative asset written on the underlying asset. We were reminded of this fact in the prior discussion! The price of the call at all times equals the value of the replicating portfolio. It does not, however, preclude the existence of such opportunities among different stocks or among derivatives written on different stocks.

These discussions make apparent the fact that binomial risk-neutral valuation views derivative securities, and call options in particular, as redundant assets, redundant in the sense that their payoffs can be replicated with a portfolio of preexisting securities. The presence or absence of these derivatives is deemed not to affect the price of the underlying asset (the stock) on which they are written. This is in direct contrast to our earlier motivation for the existence of options: their desirable property in assisting in the completion of the market. In principle, the introduction of an option has the potential of changing all asset values if it makes the market more complete.

This issue has been examined fairly extensively in the literature. From a theoretical perspective, Detemple and Selden (1991) construct a mean variance example where there is one risky asset, one risk-free asset, and an incomplete market. There the introduction of a call option is shown to increase the equilibrium price of the risky asset. In light of our earlier discussions, this is not entirely surprising: The introduction of the option enhances opportunities for risk sharing, thereby increasing demand and consequently the price of the risky asset. This result can be shown not to be fully applicable to all contexts, however. On the empirical side, Detemple and Jorion (1990) examine a large sample of options introductions over the period 1973 to 1986 and find that, on average, the underlying stock's price rises 3 percent as a result and its volatility diminishes.

12.5 Continuous Time: An Introduction to the Black-Scholes Formula

While the binomial model presents a transparent application of risk-neutral valuation, it is not clear that it represents the accurate description of the price evolution of any known security. We deal with this issue presently.

Fat tails aside, there is ample evidence to suggest that stock prices may be modeled as being lognormally distributed; more formally,

$$\ln q^e(\theta_T, T) \sim N(\ln q^e(\theta_t, t) + \mu(T - t), \sigma\sqrt{T - t}),$$

where μ and σ denote, respectively, the mean and standard deviation of the continuously compounded rate of return over the reference period, typically one year. Regarding t as the present time, this expression describes the distribution of stock prices at some time T in the future given the current price $q^e(\theta_t, t)$. The length of the time horizon $T - t$ is measured in years.

The key result is this: properly parameterized, the distribution of final prices generated by the binomial distribution can arbitrarily well approximate the prior lognormal distribution when the number of branches becomes very large. More precisely, we may imagine a binomial model in which we divide the period $T - t$ into n subintervals of equal length $\Delta t(n) = \frac{T-t}{n}$. If we adjust u , d , p (the true probability of a u price move) and R_f appropriately, then as $n \rightarrow \infty$, the distribution of period T prices generated by the binomial model will converge in probability to the hypothesized lognormal distribution. The adjustment requires that

$$\begin{aligned} u(n) &= e^{\sigma\sqrt{\Delta t(n)}}, d(n) = \frac{1}{u(n)}, p = \frac{e^{\mu\Delta t(n)} - d(n)}{u(n) - d(n)}, \text{ and} \\ R_f(n) &= (R_f)^{\frac{1}{n}} \end{aligned} \quad (12.20)$$

For this identification, the binomial valuation formula for a call option, Equation (12.18), converges to the Black-Scholes formula for a European call option written on a non-dividend paying stock:

$$C_{\mathcal{E}}(\theta_t, t) = q^e(\theta_t, T)N(d_1) - Ke^{-\hat{r}_f(T-t)}N(d_2) \quad (12.21)$$

where $N(\cdot)$ is the cumulative normal probability distribution function,

$$\begin{aligned} \hat{r}_f &= \ln(R_f) \\ d_1 &= \frac{\ln\left(\frac{q^e(\theta_t, t)}{K}\right) + (T - t)\left(\hat{r}_f + \frac{\sigma^2}{2}\right)}{\sigma\sqrt{T - t}} \\ d_2 &= d_1 - \sigma\sqrt{T - t} \end{aligned}$$

Cox and Rubinstein (1979) provide a detailed development and proof of this equivalence, but we can see the rudiments of its origin in Equation (12.18),

which we now present, modified to make apparent its dependence on the number of subintervals n :

$$C_{\mathcal{E}}(\theta_t, t; n) = \frac{1}{(R_f(n))^n} \left\{ \sum_{s=a(n)}^n \binom{n}{s} (\pi(n)^{RN})^s (1 - \pi(n)^{RN})^{n-s} q^e(\theta_t, t) - \sum_{s=a(n)}^n \binom{n}{s} (\pi(n)^{RN})^s (1 - \pi(n)^{RN})^{n-s} K \right\} \quad (12.22)$$

$$\text{where } \pi(n)^{RN} = \frac{R_f(n) - d(n)}{u(n) - d(n)}.$$

Rearranging terms yields

$$C_{\mathcal{E}}(\theta_t, t; n) = q^e(\theta_t, t) \sum_{s=a}^n \binom{n}{s} \left(\frac{\pi(n)^{RN}}{R_f(n)} \right)^s \left(\frac{1 - \pi(n)^{RN}}{R_f(n)} \right)^{n-s} - K \left(\frac{1}{R_f(n)} \right)^n \sum_{s=a}^n \binom{n}{s} (\pi(n)^{RN})^s (1 - \pi(n)^{RN})^{n-s}, \quad (12.23)$$

which is of the general form

$$C_{\mathcal{E}}(\theta_t, t; n) = q_e(\theta_t, t) \times \text{Probability} - (\text{present value factor}) \times K \times \text{Probability},$$

as per the Black-Scholes formula. Since, at each step of the limiting process (i.e., for each n , as $n \mapsto \infty$), the call valuation formula is fundamentally an expression of risk-neutral valuation, the same must be true of its limit. As such, the Black-Scholes formula represents the first hint at the translation of risk-neutral methods to the case of continuous time.

Let us conclude this section with a few more observations. The first concerns the relationship of the Black-Scholes formula to the replicating portfolio idea. Since at each step of the limiting process the call's value is identical to that of the replicating portfolio, this notion must carry over to the continuous time setting. This is indeed the case: in a context when investors may continuously and costlessly adjust the composition of the replicating portfolio, the initial position to assume (at time t) is one of $N(d_1)$ shares, financed in part by a risk-free loan of $Ke^{-R_f T} N(d_2)$. The net cost of assembling the portfolio is the Black-Scholes value of the call.

Notice also that neither the mean return on the underlying asset nor the true probabilities explicitly enter anywhere in the discussion.⁶ None of this is surprising. The short explanation is simply that risk-neutral valuation abandons the true probabilities in favor of the risk-neutral ones and, in doing so, all assets are determined to earn the risk-free rate. The underlying assets' mean

⁶They are implicitly present in the equilibrium price of the underlying asset.

return still matters, but it is now R_f . More intuitively, risk-neutral valuation is essentially no-arbitrage pricing. In a world with full information and without transaction costs, investors will eliminate all arbitrage opportunities irrespective of their objective likelihood or of the mean returns of the assets involved.

It is sometimes remarked that to purchase a call option is to buy volatility, and we need to understand what this expression is intended convey. Returning to the binomial approximation [in conjunction with Equation (12.18)], we observe first that a larger σ implies the possibility of a higher underlying asset price at expiration, with the attendant higher call payoff. More formally, σ is the only statistical characteristic of the underlying stock's price process to appear in the Black-Scholes formula. Given r_f , K , and $q^e(\theta_t, t)$, there is a unique identification between the call's value and σ . For this reason, estimates of an asset's volatility are frequently obtained from its corresponding call price by inverting the Black-Scholes formula. This is referred to as an *implied volatility estimate*.

The use of risk-neutral methods for the valuation of options is probably the area in which asset pricing theory has made the most progress. Indeed, Merton and Scholes were awarded the Nobel prize for their work (Fischer Black had died). So much progress has, in fact, been made that the finance profession has largely turned away from conceptual issues in derivatives valuation to focus on the development of fast computer valuation algorithms that mimic the risk-neutral methods. This, in turn, has allowed the use of derivatives, especially for hedging purposes, to increase so enormously over the past 20 years.

12.6 Dybvig's Evaluation of Dynamic Trading Strategies

Let us next turn to a final application of these methods: the evaluation of dynamic trading strategies. To do so, we retain the partial equilibrium setting of the binomial model, but invite agents to have preferences over the various outcomes. Note that under the pure pricing perspective of Section 12.4, preferences were irrelevant. All investors would agree on the prices of call and put options (and all other derivatives) regardless of their degrees of risk aversion, or their subjective beliefs as to the true probability of an up or down state. This is simply a reflection of the fact that any rational investor, whether highly risk averse or risk neutral, will seek to profit by an arbitrage opportunity, whatever the likelihood, and that in equilibrium, assets should thus be priced so that such opportunities are absent. In this section our goal is different, and preferences will have a role to play. We return to assumption A12.1.

Consider the optimal consumption problem of an agent who takes security prices as given and who seeks to maximize the present value of time-separable

utility (A12.1). His optimal consumption plan solves

$$\begin{aligned} \max E_0 \left(\sum_{t=0}^{\infty} U(\tilde{c}_t, t) \right) \\ \text{s.t.} \quad \sum_{t=0}^{\infty} \sum_{s \in N_t} q(\theta_0, \theta_t(s)) c(\theta_t(s)) \leq Y_0, \end{aligned} \quad (12.24)$$

where Y_0 is his initial period 0 wealth and $q(\theta_0, \theta_t(s))$ is the period $t = 0$ price of an Arrow-Debreu security paying one unit of the numeraire if state s is observed at time $t > 0$. Assuming a finite number of states and expanding the expectations operator to make explicit the state probabilities, the Lagrangian for this problem is

$$\begin{aligned} L(\cdot) = & \sum_{t=0}^{\infty} \sum_{s=1}^{N_t} \pi(\theta_0, \theta_t(s)) U(c(\theta_t(s), t)) \\ & + \lambda \left(Y_0 - \sum_{t=0}^{\infty} \sum_{s=1}^{N_t} q(\theta_0, \theta_t(s)) c(\theta_t(s)) \right), \end{aligned}$$

where $\pi(\theta_0, \theta_t(s))$ is the conditional probability of state s occurring, at time t and λ the Lagrange multiplier.

The first order condition is

$$U_1(c(\theta_t(s)), t) \pi(\theta_0, \theta_t(s)) = \lambda q(\theta_0, \theta_t(s)).$$

By the concavity of $U(\cdot)$, if $\theta_t(1)$ and $\theta_t(2)$ are two states, then

$$\frac{q(\theta_0, \theta_t(1))}{\pi(\theta_0, \theta_t(1))} > \frac{q(\theta_0, \theta_t(2))}{\pi(\theta_0, \theta_t(2))}, \text{ if and only if } c(\theta_t(1), t) < c(\theta_t(2), t). \quad (12.25)$$

It follows that if

$$\frac{q(\theta_0, \theta_t(1))}{\pi(\theta_0, \theta_t(1))} = \frac{q(\theta_0, \theta_t(2))}{\pi(\theta_0, \theta_t(2))}, \text{ then } c(\theta_t(1)) = c(\theta_t(2)).$$

The $q(\theta_0, \theta_t(s))/\pi(\theta_0, \theta_t(s))$ ratio measures the relative scarcity of consumption in state $\theta_t(s)$: A high ratio in some state suggests that the price of consumption is very high relative to the likelihood of that state being observed. This suggests that consumption is scarce in the high $q(\theta_0, \theta_t(s))/\pi(\theta_0, \theta_t(s))$ states. A rational agent will consume less in these states and more in the relatively cheaper ones, as Equation (12.25) suggests.

This observation is, in fact, quite general as Proposition 12.1 demonstrates.

Proposition 12.1 [Dybvig (1988)]

Consider the consumption allocation problem described by Equation (12.24). For any rational investor for which $U_{11}(c_t, t) < 0$, his optimal consumption plan is a decreasing function of $q(\theta_0, \theta_t(s))/\pi(\theta_0, \theta_t(s))$. Furthermore, for any consumption plan with this monotonicity property, there exists a rational investor with concave period utility function $U(c_t, t)$ for which the consumption plan is optimal in the sense of solving Equation (12.24).

Dybvig (1988) illustrates the power of this result most effectively in the binomial context where the price-to-probability ratio assumes an especially simple form. Recall that in the binomial model the state at time t is completely characterized by the number of up states, u , preceding it. Consider a state $\theta_t(s)$ where s denotes the number of preceding up states. The true conditional probability of $\theta_t(s)$ is

$$\pi(\theta_0, \theta_t(s)) = \pi^s (1 - \pi)^{t-s},$$

while the corresponding state claim has price

$$q(\theta_0, \theta_t(s)) = (R_f)^{-t} (\pi^{RN})^s (1 - \pi^{RN})^{t-s}.$$

The price/probability ratio thus assumes the form

$$\frac{q(\theta_0, \theta_t(s))}{\pi(\theta_0, \theta_t(s))} = (R_f)^{-t} \left(\frac{\pi^{RN}}{\pi} \right)^s \left(\frac{1 - \pi^{RN}}{1 - \pi} \right)^{t-s} = (R_f)^{-t} \left(\frac{\pi^{RN}(1 - \pi)}{(1 - \pi^{RN})\pi} \right)^s \left(\frac{1 - \pi^{RN}}{1 - \pi} \right)^t.$$

We now specialize the binomial process by further requiring the condition in assumption A12.6.

A12.6: $\pi u + (1 - \pi)d > R_f$, in other words, the expected return on the stock exceeds the risk-free rate.

Assumption A12.6 implies that

$$\pi > \frac{R_f - d}{u - d} = \pi^{RN},$$

so that

$$\frac{\pi^{RN}(1 - \pi)}{(1 - \pi^{RN})\pi} < 1,$$

for any time t , and the price probability ratio $q(\theta_0, \theta_t(s))/\pi(\theta_0, \theta_t(s))$ is a decreasing function of the number of preceding up moves, s . By Proposition 12.1 the period t level of optimal, planned consumption across states $\theta_t(s)$ is thus an *increasing function of the number of up moves, s , preceding it*.

Let us now specialize our agent's preferences to assume that he is only concerned with his consumption at some terminal date T , at which time he consumes his wealth. Equation (12.24) easily specializes to this case:

$$\begin{aligned} \max \quad & \sum_{s \in N_T} \pi(\theta_0, \theta_T(s)) U(c(\theta_T(s))) \\ \text{s.t.} \quad & \sum_{s \in N_T} q(\theta_0, \theta_T(s)) c(\theta_T(s)) \leq Y_0 \end{aligned} \tag{12.26}$$

In effect, we set $U(c_t, t) \equiv 0$ for $t \leq T$.

Remember also that a stock, from the perspective of an agent who is concerned only with terminal wealth, can be viewed as a portfolio of period t state claims. The results of Proposition 12.1 thus apply to this security as well.

Dybvig (1988) shows how these latter observations can be used to assess the optimality of many commonly used trading strategies. The context of his discussion is illustrated with the example in Figure 12.8 where the investor is presumed to consume his wealth at the end of the trading period.

Insert Figure 12.8 about here

For this particular setup, $\pi^{RN} = 1/3$. He considers the following frequently cited equity trading strategies:

1. Technical analysis: buy the stock and sell it after an up move; buy it back after a down move; invest at R_f (zero in this example) when out of the market. But under this strategy

$$c_4(\theta_t(s) | uuuu) = \$32, \text{ yet}$$

$$c_4(\theta_t(s) | udud) = \$48; \text{ in other words,}$$

the investor consumes more in the state with the fewer preceding up moves, which violates the optimality condition. This cannot be an optimal strategy.

2. Stop-loss strategy: buy and hold the stock, sell only if the price drops to \$8, and stay out of the market. Consider, again, two possible evolutions of the stock's price:

$$c_4(\theta_t(s) | duuu) = \$8$$

$$c_4(\theta_t(s) | udud) = \$16.$$

Once again, consumption is not a function of the number of up states under this trading strategy, which must, therefore, be suboptimal.

12.7 Conclusions

We have extended the notion of risk-neutral valuation to two important contexts: the dynamic setting of the general equilibrium consumption CAPM and the partial equilibrium binomial model. The return on our investment is particularly apparent in the latter framework. The reasons are clear: in the binomial context, which provides the conceptual foundations for an important part of continuous time finance, the risk-neutral probabilities can be identified independently from agents' preferences. Knowledge of the relevant inter-temporal marginal rates of substitution, in particular, is superfluous. This is the huge dividend of the twin modeling choices of binomial framework and arbitrage pricing. It has paved the way for routine pricing of complex derivative-based financial products and for their attendant use in a wide range of modern financial contracts.

References

Cox, J., Rubinstein, M. (1985), *Option Markets*, Prentice Hall, Upper Saddle River, N.J.

- Detemple, J., Jorion, P. (1990), “Option Listing and Stock Returns: An Empirical Analysis,” *Journal of Banking and Finance*, 14, 781–801.
- Detemple, J., Selden, L. (1991), “A General Equilibrium Analysis of Option and Stock Market Interactions,” *International Economic Review*, 32, 279–303.
- Dybvig, P. H. (1998), “Inefficient Dynamic Portfolio Strategies or How to Throw Away a Million Dollars in the Stock Market,” *The Review of Financial Studies*, 1, 67–88.
- Telmer, C. (1993), “Asset Pricing Puzzles and Incomplete Markets,” *Journal of Finance*, 48, 1803–1832.
- For an excellent text that deals with continuous time from an applications perspective, see Luenberger, D. (1998), *Investments*, Oxford University Press, New York.
- For an excellent text with a more detailed description of continuous time processes, see Dumas, B., Allaz, B. (1996), *Financial Securities*, Chapman and Hall, London.

Appendix 12.1: Risk-Neutral Valuation When Discounting at the Term Structure of Multiperiod Discount Bond

Here we seek a valuation formula where we discount not at the succession of one-period rates, but at the term structure. This necessitates a different set of risk-neutral probabilities with respect to which the expectation is taken.

Define the k -period, time adjusted risk-neutral transition probabilities as:

$$\hat{\pi}^{RN}(\theta_t, \theta_{t+k}) = \left\{ \frac{\pi^{RN}(\theta_t, \theta_{t+k})g(\theta_t, \theta_{t+k})}{q^b(\theta_t, \theta_{t+k})} \right\},$$

where $\pi^{RN}(\theta_t, \theta_{t+k}) = \prod_{s=t}^{t+k-1} \pi^{RN}(\theta_s, \theta_{s+1})$, and $\{\theta_t, \dots, \theta_{t+k-1}\}$ is the path of states preceding θ_{t+k} . Clearly, the $\hat{\pi}^{RN}(\)$ are positive since $\pi^{RN}(\) \geq 0$, $g(\theta_t, \theta_{t+k}) > 0$ and $q^b(\theta_t, \theta_{t+k}) > 0$. Furthermore, by Equation (12.13),

$$\begin{aligned} \sum_{\theta_{t+k}} \hat{\pi}^{RN}(\theta_t, \theta_{t+k}) &= \left(\frac{1}{q^b(\theta_t, \theta_{t+k})} \right) \sum_{\theta_{t+k}} \pi^{RN}(\theta_t, \theta_{t+k})g(\theta_t, \theta_{t+k}) \\ &= \frac{q^b(\theta_t, \theta_{t+k})}{q^b(\theta_t, \theta_{t+k})} = 1. \end{aligned}$$

Let us now use this approach to price European call and put options. A European call option contract represents the right (but not the obligation) to buy some underlying asset at some prespecified price (referred to as the exercise

Table A12.1 Payoff Pattern- European Call Option

t	$t+1$	$t+2$	\dots	$T-1$	T
0	0	0		0	$\max \{q^e(\theta_T, T) - K, 0\},$

or strike price) at some prespecified future date (date of contract expiration). Since such a contract represents a right, its payoff is as shown in Table A12.1.

where T represents the time of expiration and K the exercise price.

Let $C_{\mathcal{E}}(\theta_t, t)$ denote the period t , state θ_t price of the call option. Clearly,

$$\begin{aligned} C_{\mathcal{E}}(\theta_t, t) &= E_t^{RN} \{g(\theta_t, \tilde{\theta}_T)(\max \{q^e(\tilde{\theta}_T, T) - K, 0\})\} \\ &= q^b(\theta_t, T) \hat{E}_t^{RN} \{\max \{q^e(\tilde{\theta}_T, T) - K, 0\}\}, \end{aligned}$$

where \hat{E}_t^{RN} denotes the expectations operator corresponding to the $\{\hat{\pi}^{RN}\}$. A European put option is similarly priced according to

$$\begin{aligned} P_{\mathcal{E}}(\theta_t, t) &= E_t^{RN} \{g(\theta_t, \tilde{\theta}_T)(\max \{K - q^e(\tilde{\theta}_T, T), 0\})\} \\ &= q^b(\theta_t, T) \hat{K}_t^{RN} \{\max \{K - q^e(\tilde{\theta}_T, T), 0\}\} \end{aligned}$$

Chapter 13 : The Arbitrage Pricing Theory

13.1 Introduction

We have made two first attempts (Chapters 10 to 12) at asset pricing from an arbitrage perspective, that is, without specifying a complete equilibrium structure. Here we try again from a different, more empirically based angle. Let us first collect a few thoughts as to the differences between an arbitrage approach and equilibrium modeling.

In the context of general equilibrium theory, we make hypotheses about agents – consumers, producers, investors; in particular, we start with some form of rationality hypothesis leading to the specification of maximization problems under constraints. We also make hypotheses about markets: Typically we assume that supply equals demand in all markets under consideration.

We have repeatedly used the fact that at general equilibrium with fully informed optimizing agents, there can be no arbitrage opportunities, in other words, no possibilities to make money risklessly at zero cost. An arbitrage opportunity indeed implies that at least one agent can reach a higher level of utility without violating his/her budget constraint (since there is no extra cost).

In particular, our assertion that one can price any asset (income stream) from the knowledge of Arrow-Debreu prices relied implicitly on a no-arbitrage hypothesis: with a complete set of Arrow-Debreu securities, it is possible to replicate any given income stream and hence the value of a given income stream, the price paid on the market for the corresponding asset, cannot be different from the value of the replicating portfolio of Arrow-Debreu securities. Otherwise an arbitrageur could make arbitrarily large profits by short selling large quantities of the more expensive of the two and buying the cheaper in equivalent amount. Such an arbitrage would have zero cost and be riskless.

While general equilibrium implies the no-arbitrage condition, it is more restrictive in the sense of imposing a heavier structure on modeling. And the reverse implication is *not* true: No arbitrage opportunities¹ – the fact that all arbitrage opportunities have been exploited – does not imply that a general equilibrium in all markets has been obtained. Nevertheless, or precisely for that reason, it is interesting to see how far one can go in exploiting the less restrictive hypothesis that no arbitrage opportunities are left unexploited.

The underlying logic of the APT to be reviewed in this chapter is, in a sense, very similar to the fundamental logic of the Arrow-Debreu model and it is very much in the spirit of a complete market structure. It distinguishes itself in two major ways: First it replaces the underlying structure based on fundamental securities paying exclusively in a given state of nature with other fundamental securities exclusively remunerating some form of risk taking. More precisely, the APT abandons the analytically powerful, but empirically cumbersome, concept of states of nature as the basis for the definition of its primitive securities. It

¹An arbitrage portfolio is a self-financing (zero net-investment) portfolio. An arbitrage opportunity exists if an arbitrage portfolio exists that yields non-negative cash flows in all states of nature and positive cash flows in some states (Chapter 11).

replaces it with the hypothesis that there exists a (stable) set of factors that are essential and exhaustive determinants of all asset returns. The primitive security will then be defined as a security whose risk is exclusively determined by its association with one specific risk factor and totally immune from association with any other risk factor. The other difference with the Arrow-Debreu pricing of Chapter 8 is that the prices of the fundamental securities are not derived from primitives – supply and demand, themselves resulting from agents’ endowments and preferences – but will be deduced empirically from observed asset returns without attempting to explain them. Once the price of each fundamental security has been inferred from observed return distributions, the usual arbitrage argument applied to complex securities will be made (in the spirit of Chapter 10).²

13.2 Factor Models

The main building block of the APT is a factor model, also known as a return-generating process. As discussed previously, this is the structure that is to replace the concept of states of nature. The motivation has been evoked before: States of nature are analytically convincing and powerful objects. In practice, however, they are difficult to work with and, moreover, often not verifiable, implying that contracts cannot necessarily be written contingent on a specific state of nature. We discussed these shortcomings of the Arrow-Debreu pricing theory in Chapter 8. The temptation is thus irresistible to attack the asset pricing problem from the opposite angle and build the concept of primitive securities on an empirically more operational notion, abstracting from its potential theoretical credentials. This structure is what factor models are for.

The simplest conceivable factor model is a one-factor market model, usually labeled the *Market Model*, which asserts that ex-post returns on individual assets can be entirely ascribed either to their own specific stochastic components or to their common association in a single factor, which in the CAPM world would naturally be selected as the return on the market portfolio. This simple factor model can thus be summarized by following the equation (or process):³

$$\tilde{r}_j = \alpha_j + \beta_j \tilde{r}_M + \tilde{\varepsilon}_j, \quad (13.1)$$

with $E\tilde{\varepsilon}_j = 0$, $\text{cov}(\tilde{r}_M, \tilde{\varepsilon}_j) = 0, \forall j$, and $\text{cov}(\tilde{\varepsilon}_j, \tilde{\varepsilon}_k) = 0, \forall j \neq k$.

This model states that there are three components in individual returns: (1) an asset-specific constant α_j ; (2) a common influence, in this case the unique factor, the return on the market, which affects all assets in varying degrees, with β_j measuring the sensitivity of asset j ’s return to fluctuations in the market return; and (3) an asset-specific stochastic term $\tilde{\varepsilon}_j$ summarizing all other stochastic components of \tilde{r}_j that are unique to asset j .

²The arbitrage pricing theory was first developed by Ross (1976), and substantially interpreted by Huberman (1982) and Conner (1984) among others. For a presentation emphasizing practical applications, see Burmeister et al. (1994).

³Factors are frequently measured as deviations from their mean. When this is the case, α_j becomes an estimate of the mean return on asset j .

Equation (13.1) has no bite (such an equation can always be written) until one adds the hypothesis $\text{cov}(\tilde{\varepsilon}_j, \tilde{\varepsilon}_k) = 0, j \neq k$, which signifies that *all* return characteristics common to different assets are subsumed in their link with the market return. If this were empirically verified, the CAPM would be the undisputed end point of asset pricing. At an empirical level, one may say that it is quite unlikely that a single factor model will suffice.⁴ But the strength of the APT is that it is agnostic as to the number of underlying factors (and their identity). As we increase the number of factors, hoping that this will not require a number too large to be operational, a generalization of Equation (13.1) becomes more and more plausible. But let us for the moment maintain the hypothesis of one common factor for pedagogical purposes.⁵

13.2.1 About the Market Model

Besides serving as a potential basis for the APT, the Market Model, despite all its weaknesses, is also of interest on two grounds. First it produces estimates for the β 's that play a central role in the CAPM. Note, however, that estimating β 's from past data alone is useful only to the extent that some degree of stationarity in the relationship between asset returns and the return on the market is present. Empirical observations suggest a fair amount of stationarity is plausible at the level of portfolios, but not of individual assets. On the other hand, estimating the β 's does not require all the assumptions of the Market Model; in particular, a violation of the $\text{cov}(\tilde{\varepsilon}_i, \tilde{\varepsilon}_k) = 0, i \neq k$ hypothesis is not damaging.

The second source of interest in the Market Model, crucially dependent on the latter hypothesis being approximately valid, is that it permits economizing on the computation of the matrix of variances and covariances of asset returns at the heart of the MPT. Indeed, under the Market Model hypothesis, one can write (you are invited to prove these statements):

$$\begin{aligned}\sigma_j^2 &= \beta_j^2 \sigma_M^2 + \sigma_{\varepsilon j}^2, \forall j \\ \sigma_{ij} &= \beta_i \beta_j \sigma_M^2\end{aligned}$$

This effectively means that the information requirements for the implementation of MPT can be substantially weakened. Suppose there are N risky assets under consideration. In that case the computation of the efficient frontier requires knowledge of N expected returns, N variances, and $\frac{N^2 - N}{2}$ covariance terms (N^2 is the total number of entries in the matrix of variances and covariances, take away the N variance/diagonal terms and divide by 2 since $\sigma_{ij} = \sigma_{ji}, \forall i, j$).

⁴Recall the difficulty in constructing the empirical counterpart of M .

⁵Fama (1973), however, demonstrates that in its form (13.1) the Market Model is inconsistent in the following sense: the fact that the market is, by definition, the collection of all individual assets implies an exact linear relationship between the disturbances ε_j ; in other words, when the single factor is interpreted to be the market the hypothesis $\text{cov}(\tilde{\varepsilon}_j, \tilde{\varepsilon}_k) = 0, \forall j \neq k$ cannot be strictly valid. While we ignore this criticism in view of our purely pedagogical objective, it is a fact that if a single factor model had a chance to be empirically verified (in the sense of all the assumptions in (13.1) being confirmed) the unique factor could not be the market.

Working via the Market Model, on the other hand, requires estimating Equation (13.1) for the N risky returns producing estimations for the N β_j 's and the N $\sigma_{\varepsilon j}^2$ and estimating the variance of the market return, that is, $2N + 1$ information items.

13.3 The APT: Statement and Proof

13.3.1 A Quasi-Complete Market Hypothesis

To a return-generating process such as the Market Model, the APT superposes a second major hypothesis that is akin to assuming that the markets are “quasi-complete”. What is needed is the existence of a *rich* market structure with a large number of assets with different characteristics and a minimum number of trading restrictions. This market structure, in particular, makes it possible to form a portfolio P with the following three properties:

Property 1: P has zero cost; in other words, it requires no investment. This is the first requirement of an arbitrage portfolio.

Let us denote x_i as the *value* of the position in the i th asset in portfolio P . Portfolio P is then fully described by the vector $x^T = (x_1, x_2, \dots, x_N)$ and the zero cost condition becomes

$$\sum_{i=1}^N x_i = 0 = x^T \cdot \mathbf{1},$$

with $\mathbf{1}$ the (column) vector of 1's. (Positive positions in some assets must be financed by short sales of others.)

Property 2: P has zero sensitivity (zero beta) to the common factor:⁶

$$\sum_i x_i \beta_i = 0 = x^T \cdot \beta.$$

Property 3: P is a well-diversified portfolio. The specific risk of P is (almost)

totally eliminated:

$$\sum_i x_i^2 \sigma_{\varepsilon i}^2 \cong 0.$$

The APT builds on the assumed existence of such a portfolio, which requires a rich market structure.

⁶Remember that the beta of a portfolio is the weighted sum of the betas of the assets in the portfolio.

13.3.2 Statement and Proof of the APT

The APT relationship is the direct consequence of the factor structure hypothesis, the existence of a portfolio P satisfying these conditions, and the no-arbitrage assumption. Given that returns have the structure of Equation (13.1), Properties 2 and 3 imply that P is riskless. The fact that P has zero cost (Property 1) then entails that an arbitrage opportunity will exist unless:

$$\bar{r}_P = 0 = x^T \cdot \bar{r} \quad (13.2)$$

The APT theorem states, as a consequence of this succession of statements, that there must exist scalars λ_0, λ_1 , such that:

$$\begin{aligned} \bar{r} &= \lambda_0 \cdot \mathbf{1} + \lambda_1 \beta, \text{ or} \\ \bar{r}_i &= \lambda_0 + \lambda_1 \beta_i \text{ for all assets } i \end{aligned} \quad (13.3)$$

This is the main equation of the APT.

Equation (13.3) and Properties 1 and 2 are statements about 4 vectors: x , β , $\mathbf{1}$, and \bar{r} .

Property 1 states that x is orthogonal to $\mathbf{1}$. Property 2 asserts that x is orthogonal to β . Together these statements imply a geometric configuration that we can easily visualize if we fix the number of risky assets at $N = 2$, which implies that all vectors have dimension 2. This is illustrated in Figure 13.1.

Insert Figure 13.1

Equation (13.3) – no arbitrage – implies that x and \bar{r} are orthogonal. But this means that the vector \bar{r} must lie in the plane formed by $\mathbf{1}$ and β , or, that \bar{r} can be written as a linear combination of $\mathbf{1}$ and β , as Equation (13.3) asserts.

More generally, one can deduce from the triplet

$$\sum_i^N x_i = \sum_i^N x_i \beta_i = \sum_i^N x_i \bar{r}_i = 0$$

that there exist scalars λ_0, λ_1 , such that:

$$\bar{r}_i = \lambda_0 + \lambda_1 \beta_i \text{ for all } i.$$

This is a consequence of the orthonormal projection of the vector \bar{r}_i into the subspace spanned by the other two.

13.3.3 Meaning of λ_0 and λ_1

Suppose that there exists a risk-free asset or, alternatively, that the sufficiently rich market structure hypothesis permits constructing a fully diversified portfolio with zero-sensitivity to the common factor (but positive investment). Then

$$\bar{r}_f = r_f = \lambda_0.$$

That is, λ_0 is the return on the risk-free asset or the risk-free portfolio.

Now let us compose a portfolio Q with unitary sensitivity to the common factor $\beta = 1$. Then applying the APT relation, one gets:

$$\bar{r}_Q = r_f + \lambda_1 \cdot 1$$

Thus, $\lambda_1 = \bar{r}_Q - r_f$, the excess-return on the pure-factor portfolio Q . It is now possible to rewrite equation (13.3) as:

$$\bar{r}_i = r_f + \beta_i (\bar{r}_Q - r_f). \quad (13.4)$$

If, as we have assumed, the unique common factor is the return on the market portfolio, in which case $Q = M$ and $\bar{r}_Q \equiv \bar{r}_M$, then Equation (13.4) is simply the CAPM equation:

$$\bar{r}_i = r_f + \beta_i (\bar{r}_M - r_f).$$

13.4 Multifactor Models and the APT

The APT approach is generalizable to any number of factors. It does not, however, provide any clue as to what these factors should be, or any particular indication as to how they should be selected. This is both its strength and its weakness. Suppose we can agree on a two-factor model:

$$\tilde{r}_j = a_j + b_{j1}\tilde{F}_1 + b_{j2}\tilde{F}_2 + \tilde{e}_j \quad (13.5)$$

with $E\tilde{e}_j = 0$, $\text{cov}(\tilde{F}_1, \tilde{e}_j) = \text{cov}(\tilde{F}_2, \tilde{e}_j) = 0, \forall j$, and $\text{cov}(\tilde{e}_j, \tilde{e}_k) = 0, \forall j \neq k$.

As was the case for Equation (13.1), Equation (13.5) implies that one cannot reject, empirically, the hypothesis that the ex-post return on an asset j has two stochastic components: one specific, (\tilde{e}_j) , and one systematic, $(b_{j1}\tilde{F}_1 + b_{j2}\tilde{F}_2)$. What is new is that the systematic component is not viewed as the result of a single common factor influencing all assets. Common or systematic issues may now be traced to two fundamental factors affecting, in varying degrees, the returns on individual assets (and thus on portfolios as well). Without loss of generality we may assume that these factors are uncorrelated.

As before, an expression such as Equation (13.5) is useful only to the extent that it describes a relationship that is relatively stable over time. The two factors F_1 and F_2 must really summarize *all* that is common in individual asset returns.

What could these fundamental factors be? In an important article, Chen, Roll, and Ross (1986) propose that the systematic forces influencing returns must be those affecting discount factors and expected cash flows. They then isolate a set of candidates such as industrial production, expected and unexpected inflation, measures of the risk premium and the term structure, and even oil prices. At the end, they conclude that the most significant determinants of asset returns are industrial production (affecting cash flow expectations), changes in the risk premium measured as the spread between the yields on low- and high-risk corporate bonds (witnessing changes in the market risk appetite), and

twists in the yield curve, as measured by the spread between short- and long-term interest rates (representing movements in the market rate of impatience). Measures of unanticipated inflation and changes in expected inflation also play a (less important) role.

Let us follow, in a simplified way, Chen, Roll, and Ross's lead and decide that our two factors are industrial production (F_1) and changes in the risk premium (F_2). How would we go about implementing the APT? First we have to measure our two factors. Let $IP(t)$ denote the rate of industrial production in month t ; then $MP(t) = \log IP(t) - \log IP(t-1)$ is the monthly growth rate of IP . This is our first explanatory variable.

To measure changes in the risk premium, let us define

$$UPR(t) = \text{"Baa and under" bond portfolio return}(t) - LGB(t)$$

where $LGB(t)$ is the return on a portfolio of long-term government bonds. With these definitions we can rewrite Equation (13.5) as

$$\tilde{r}_{jt} = a_j + b_{j1}MP(t) + b_{j2}UPR(t) + \tilde{e}_j$$

The $b_{jk}, k = 1, 2$, are often called *factor loadings*. They can be estimated directly by multivariate regression. Alternatively, one could construct *pure factor portfolios* – well-diversified portfolios mimicking the underlying factors – and compute their correlation with asset j . The pure factor portfolio P_1 would be a portfolio with $b_{P_1} = 1$ and $b_{P_2} = \sigma_{e_{P_1}} = 0$; portfolio P_2 would be defined similarly to track the stochastic behavior of $UPR(t)$. Let us go on hypothesizing (wrongly according to Chen, Roll, and Ross) that this two-factor model satisfies the necessary assumptions ($\text{cov}(\tilde{e}_i, \tilde{e}_j) = 0, \forall i \neq j$) and further assume the existence of a risk-free portfolio P_f with zero sensitivity to either of our two factors and zero specific risk. Then the APT states that there exist scalars $\lambda_0, \lambda_1, \lambda_2$ such that:

$$\bar{r}_j = \lambda_0 + \lambda_1 b_{j1} + \lambda_2 b_{j2}.$$

That is, the expected return on an arbitrary asset j is perfectly and completely described by a linear function of asset j 's factor loadings b_{j1}, b_{j2} . This can appropriately be viewed as a (two-factor) generalization of the SML.

Furthermore the coefficients of the linear function are:

$$\begin{aligned} \lambda_0 &= r_f \\ \lambda_1 &= \bar{r}_{P_1} - r_f \\ \lambda_2 &= \bar{r}_{P_2} - r_f \end{aligned}$$

where P_1 and P_2 are our pure factor portfolios.

The APT agrees with the CAPM that the risk premium on an asset, $\bar{r}_j - \lambda_0$, is not a function of its specific or diversifiable risk. It potentially disagrees with the CAPM in the identification of the systematic risk. The APT decomposes the systematic risk into elements of risk associated with a particular asset's sensitivity to a few fundamental common factors.

Note the parallelism with the Arrow-Debreu pricing approach. In both contexts, every individual asset or portfolio can be viewed as a complex security, or a combination of primitive securities: Arrow-Debreu securities in one case, the pure factor portfolios in the other. Once the prices of the primitive securities are known, it is a simple step to compose replicating portfolios and, by a no-arbitrage argument, price complex securities and arbitrary cash flows. The difference, of course, resides in the identification of the primitive security. While the Arrow-Debreu approach sticks to the conceptually clear notion of states of nature, the APT takes the position that there exist a few common and stable sources of risk and that they can be empirically identified. Once the corresponding risk premia are identified, by observing the market-determined premia on the primitive securities (the portfolios with unit sensitivity to a particular factor and zero sensitivity to all others) the pricing machinery can be put to work.

Let us illustrate. In our two-factor examples, a security j with, say, $b_{j1} = 0.8$ and $b_{j2} = 0.4$ is like a portfolio with proportions of 0.8 of the pure portfolio P_1 , 0.4 of pure portfolio P_2 , and consequently proportion -0.2 in the riskless asset. By our usual (no-arbitrage) argument, the expected rate of return on that security must be:

$$\begin{aligned}\bar{r}_j &= -0.2r_f + 0.8\bar{r}_{P_1} + 0.4\bar{r}_{P_2} \\ &= -0.2r_f + 0.8r_f + 0.4r_f + 0.8(\bar{r}_{P_1} - r_f) + 0.4(\bar{r}_{P_2} - r_f) \\ &= r_f + 0.8(\bar{r}_{P_1} - r_f) + 0.4(\bar{r}_{P_2} - r_f) \\ &= \lambda_0 + b_{j1}\lambda_1 + b_{j2}\lambda_2\end{aligned}$$

The APT equation can thus be seen as the immediate consequence of the linkage between pure factor portfolios and complex securities in an arbitrage-free context. The reasoning is directly analogous to our derivation of the value additivity theorem in Chapter 10 and leads to a similar result: Diversifiable risk is not priced in a complete (or quasi-complete) market world.

While potentially more general, the APT does not necessarily contradict the CAPM. That is, it may simply provide another, more disaggregated, way of writing the expected return premium associated with systematic risk, and thus a decomposition of the latter in terms of its fundamental elements. Clearly the two theories have the same implications if (keeping with our two-factor model, the generalization is trivial):

$$\beta_j (\bar{r}_M - r_f) = b_{j1} (\bar{r}_{P_1} - r_f) + b_{j2} (\bar{r}_{P_2} - r_f) \quad (13.6)$$

Let β_{P_1} be the (market) beta of the pure portfolio P_1 and similarly for β_{P_2} . Then if the CAPM is valid, not only is the LHS of Equation (13.6) the expected risk premium on asset j , but we also have:

$$\begin{aligned}\bar{r}_{P_1} - r_f &= \beta_{P_1} (\bar{r}_M - r_f) \\ \bar{r}_{P_2} - r_f &= \beta_{P_2} (\bar{r}_M - r_f)\end{aligned}$$

Thus the APT expected risk premium may be written as:

$$b_{j1} [\beta_{P_1} (\bar{r}_M - r_f)] + b_{j2} [\beta_{P_2} (\bar{r}_M - r_f)] = (b_{j1}\beta_{P_1} + b_{j2}\beta_{P_2}) (\bar{r}_M - r_f)$$

which is the CAPM equation provided:

$$\beta_j = b_{j1}\beta_{P_1} + b_{j2}\beta_{P_2}$$

In other words, CAPM and APT have identical implications if the sensitivity of an arbitrary asset j with the market portfolio fully summarizes its relationship with the two underlying common factors. In that case, the CAPM would be another, more synthetic, way of writing the APT.⁷

In reality, of course, there are reasons to think that the APT with an arbitrary number of factors will always do at least as well in identifying the sources of systematic risk as the CAPM. And indeed Chen, Roll, and Ross observe that their five factors *cover* the market return in the sense that adding the return on the market to their preselected five factors does not help in explaining expected returns on individual assets.

13.5 Advantage of the APT for Stock or Portfolio Selection

The APT helps to identify the sources of systematic risk, or to split systematic risk into its fundamental components. It can thus serve as a tool for helping the portfolio manager modulate his risk exposure. For example, studies show that, among U.S. stocks, the stocks of chemical companies are much more sensitive to short-term inflation risk than stocks of electrical companies. This would be compatible with both having the same exposure to variations in the market return (same beta). Such information can be useful in at least two ways. When managing the portfolio of an economic agent whose natural position is very sensitive to short-term inflation risk, chemical stocks may be a lot less attractive than electricals, all other things equal (even though they may both have the same market beta). Second, conditional expectations, or accurate predictions, on short-term inflation may be a lot easier to achieve than predictions of the market's return. Such a refining of the information requirements needed to take aggressive positions can, in that context, be of great use.

13.6 Conclusions

We have now completed our review of asset pricing theories. At this stage it may be useful to draw a final distinction between the equilibrium theories covered in Chapters 7, 8, and 9 and the theories based on arbitrage such as the Martingale pricing theory and the APT. Equilibrium theories aim at providing a complete theory of value on the basis of *primitives*: preferences, technology, and market structure. They are inevitably *heavier*, but their weight is proportional to their ambition. By contrast, arbitrage-based theories can only provide a relative theory of value. With what may be viewed as a minimum of assumptions, they

⁷The observation in footnote 5, however, suggests this could be true as an approximation only.

- offer bounds on option values as a function of the price of the underlying asset, the stochastic behavior of the latter being taken as given (and unexplained);
- permit estimating the value of arbitrary cash flows or securities using risk-neutral measures extracted from the market prices of a set of fundamental securities, or in the same vein, using Arrow-Debreu prices extracted from a complete set of complex securities prices;
- explain expected returns on any asset or cash flow stream once the price of risk associated with pure factor portfolios has been estimated from market data on the basis of a postulated return-generating process.

Arbitrage-based theories currently have the upper hand in practitioners' circles where their popularity far outstrips the degree of acceptance of equilibrium theories. This, possibly temporary, state of affairs may be interpreted as a measure of our ignorance and the resulting need to restrain our ambitions.

References

- Burmeister, E., Roll, R., Ross, S.A. (1994), "A Practitioner's Guide to Arbitrage Pricing Theory," in *A Practitioner's Guide to Factor Models*, Research Foundation of the Institute of Chartered Financial Analysts, Charlottesville, VA.
- Chen, N. F., Roll R., Ross, S.A. (1986), "Economic Forces and the Stock Market," *Journal of Business* 59(3), 383-404.
- Connor, G. (1984), "A Unified Beta Pricing Theory," *Journal of Economic Theory* 34(1).
- Fama, E.F. (1973), "A Note on the Market Model and the Two-Parameter Model," *Journal of Finance* 28 (5), 1181-1185
- Huberman, G. (1982), "A Simple Approach to Arbitrage Pricing," *Journal of Economic Theory*, 28 (1982): 183-191.
- Ross, S. A. (1976), "The Arbitrage Pricing Theory," *Journal of Economic Theory*, 1, 341-360.

Part V

Extensions

Chapter 14 : Portfolio Management in the Long Run

14.1 Introduction

The canonical portfolio problem (Section 5.1) and the MPT portfolio selection problem embedded in the CAPM are both one-period utility-of-terminal-wealth maximization problems.

As such the advice to investors implicit in these theories is astonishingly straightforward:

(i) Be well diversified. Conceptually this recommendation implies that the risky portion of an investor's portfolio should resemble (be perfectly positively correlated with) the true market portfolio M . In practice, it usually means holding the risky component of invested wealth as a set of stock index funds, each one representing the stock market of a particular major market capitalization country with the relative proportions dependent upon the relevant ex-ante variance-covariance matrix estimated from recent historical data.

(ii) Be on the capital market line. That is, the investor should allocate his wealth between risk free assets and the aforementioned major market portfolio in proportions that are consistent with his subjective risk tolerance. Implicit in this second recommendation is that the investor first estimates her coefficient of relative risk aversion as per Section 4.5, and then solves a joint savings-portfolio allocation problem of the firm illustrated in Section 5.6.3. The risk free rate used in these calculations is customarily a one year T-bill rate in the U.S. or its analogue elsewhere.

But what should the investor do next period after this period's risky portfolio return realization has been observed? Our one period theory has nothing to say on this score except to invite the investor to repeat the above two step process possibly using an updated variance-covariance matrix and an updated risk free rate. This is what is meant by the investor behaving myopically. Yet we are uneasy about leaving the discussion at this level. Indeed, a number of important considerations seem purposefully to be ignored by following such a set of recommendations.

1) Equity return distributions have historically evolved in a pattern that is partially predictable. Suppose, for example, that a high return realization in the current period is on average followed by a low return realization in the subsequent period. This variation in conditional returns might reasonably be expected to influence intertemporal portfolio composition.

2) While known ex-ante relative to the start of a period, the risk free rate also varies through time (for the period 1928-1985 the standard deviation of the U.S. T-bill rate is 5.67%). From the perspective of a long term U.S. investor, the one period T-bill rate no longer represents a truly risk free return. Can any asset be viewed as risk free from a multiperiod perspective?

3) Investors typically receive labor income, and this fact will likely affect both the quantity of investable savings and the risky-risk free portfolio compo-

sition decision. The latter possibility follows from the observation that labor income may be viewed as the “dividend” on an implicit non-tradeable human capital asset, whose value may be differentially correlated with risky assets in the investor’s financial wealth portfolio.¹ If labor income were risk free (tenured professors!) the presence of a high value risk free asset in the investor’s overall wealth portfolio will likely tilt his security holdings in favor of a greater proportion in risky assets than would otherwise be the case.

4) There are other life-cycle considerations: savings for the educational expenses of children, the gradual disappearance of the labor income asset as retirement approaches, etc. How do these obligations and events impact portfolio choice?

5) There is also the issue of real estate. Not only does real estate (we are thinking of owner-occupied housing for the moment) provide a risk free service flow, but it is also expensive for an investor to alter his stock of housing. How should real estate figure into an investor’s multiperiod investment plan?

6) Other considerations abound. There are substantial taxes and transactions costs associated with rebalancing a portfolio of securities. Taking these costs into account, how frequently should a long term investor optimally alter his portfolio’s composition?

In this chapter we propose to present some of the latest research regarding these issues. Our perspective is one in which investors live for many periods (in the case of private universities, foundations or insurance companies, it is reasonable to postulate an infinite lifetime). For the moment, we will set aside the issue of real estate and explicit transactions costs, and focus on the problem of a long-lived investor confronted with jointly deciding, on a period by period basis, not only how much he should save and consume out of current income, but also the mix of assets, risky and risk free, in which his wealth should be invested.

In its full generality, the problem confronting a multiperiod investor-saver with outside labor income is thus:

$$\begin{aligned} \max_{\{a_t, S_t\}} E \left(\sum_{t=0}^T \delta^t U(C_t) \right) & \quad (14.1) \\ \text{s.t. } C_T &= S_{T-1}a_{T-1}(1 + \tilde{r}_T) + S_{T-1}(1 - a_{T-1})(1 + r_{f,T}) + \tilde{L}_T, \quad t = T \\ C_t + S_t &\leq S_{t-1}a_{t-1}(1 + \tilde{r}_t) + S_{t-1}(1 - a_{t-1})(1 + r_{f,t}) + \tilde{L}_t, \quad 1 \leq t \leq T-1 \\ C_0 + S_0 &\leq Y_0 + L_0, \quad t = 0 \end{aligned}$$

where \tilde{L}_t denotes the investor’s (possibly uncertain) period t labor income, and \tilde{r}_t represents the period t return on the risky asset which we shall understand

¹In particular, the value of an investor’s labor income asset is likely to be highly correlated with the return on the stock of the firm with whom he is employed. Basic intuition would suggest that the stock of one’s employer should not be held in significant amounts from a wealth management perspective.

to mean a well diversified stock portfolio.²

Problem (14.1) departs from our earlier notation in a number of ways that will be convenient for developments later in this chapter; in particular, C_t and S_t denote, respectively, period t consumption and savings rather than their lower case analogues (as in Chapter 5).

The fact that the risk free rate is indexed by t admits the possibility that this quantity, although known at the start of a period, can vary from one period to the next. Lastly, a_t will denote the *proportion* of the investor's savings assigned to the risky asset (rather than the absolute amount as before).

All other notation is standard; problem (14.1) is simply the multiperiod version of the portfolio problem in Section 5.6.3 augmented by the introduction of labor income. In what follows we will also assume that all risky returns are lognormally distributed, and that the investor's $U(C_t)$ is of the power utility CRRA class. The latter is needed to make certain that risk aversion is independent of wealth. Although investors have become enormously wealthier over the past 200 years, risk free rates and the return premium on stocks have not changed markedly, facts otherwise inconsistent with risk aversion dependent on wealth.

In its full generality, problem (14.1) is both very difficult to solve and begrudging of intuition. We thus restrict its scope and explore a number of special cases. The natural place to begin is to explore the circumstances under which the myopic solution of Section 5.3 carries over to the dynamic context of problem (14.1).

14.2 The Myopic Solution

With power utility, an investor's optimal savings to wealth ratio will be constant so the key to a fully myopic decision rule will lie in the constancy of the a ratio. Intuitively, if the same portfolio decisions are to be made, a natural sufficient condition would be to guarantee that the investor is confronted by the same opportunities on a period by period basis. Accordingly, we assume the return environment is not changing through time; in other words that $r_{f,t} \equiv r_f$ is constant and $\{\tilde{r}_t\}$ is independently and identically distributed. These assumptions guarantee that future prospects look the same period after period. Further exploration mandates that $L_t \equiv 0$ (with constant r_f , the value of this asset will otherwise be monotonically declining which is an implicit change in future wealth). We summarize these considerations as:

Theorem 14.1 (Merton, 1971)

Consider the canonical multiperiod consumption-saving-portfolio allocation problem (14.1); suppose $U(\cdot)$ displays CRRA, r_f is constant and $\{\tilde{r}_t\}$ is i.i.d.

²This portfolio might be the market portfolio M but not necessarily. Consider the case in which the investor's labor income is paid by one of the firms in M . It is likely that this particular firm's shares would be underweighted (relative to M) in the investor's portfolio.

Then the ratio a_t is time invariant.³

This is an important result in the following sense. It delineates the conditions under which a pure static portfolio choice analysis is generalizable to a multiperiod context. The optimal portfolio choice – in the sense of the allocation decision between the risk free and the risky asset – defined in a static one period context will continue to characterize the optimal portfolio decision in the more natural multiperiod environment. The conditions which are imposed are easy to understand: If the returns on the risky asset were not independently distributed, today’s realization of the risky return would provide information about the future return distribution which would almost surely affect the allocation decision. Suppose, for example, that returns are positively correlated. Then a good realization today would suggest high returns are more likely again tomorrow. It would be natural to take this into account by, say, increasing the share of the risky asset in the portfolio (beware, however, that, as the first sections of Chapter 5 illustrate, without extra assumption on the shape of the utility function – beyond risk aversion – the more intuitive result may not generally obtain. We will be reminded of this in the sequel of this chapter where, in particular, the log utility agent will stand out as a reference). The same can be said if the risk free rate is changing through time. In a period of high risk free rates, the riskless asset would be more attractive, all other things equal.

The need for the other assumption – the CRRA utility specification – is a direct consequence of Theorem 5.5. With another utility form than CRRA, Theorem 5.5 tells us that the share of wealth invested in the risky asset varies with the “initial” wealth level, that is, the wealth level carried over from the last period. But in a multiperiod context, the investable wealth, that is, the savings level, is sure to be changing over time, increasing when realized returns are favorable and decreasing otherwise. With a non-CRRA utility function, optimal portfolio allocations would consistently be affected by these changes.

Now let us illustrate the power of these ideas to evaluate an important practical problem. Consider the problem of an individual investor saving for retirement: at each period he must decide what fraction of his already accumulated wealth should be invested in stocks (understood to mean a well diversified portfolio of risky assets) and risk free bonds for the next investment period. We will maintain the $L_t \equiv 0$ assumption. Popular wisdom in this area can be summarized in the following three assertions:

(1) Early in life the investor should invest nearly all of his wealth in stocks (stocks have historically outperformed risk free assets over long (20 year) periods), while gradually shifting almost entirely into risk free instruments as retirement approaches in order to avoid the possibility of a catastrophic loss.

(2) If an investor is saving for a target level of wealth (such as, in the U.S., college tuition payments for children), he should gradually reduce his holdings

³If the investor’s period utility is log it is possible to relax the independence assumption. This important observation, first made by Samuelson (1969), will be confirmed later on in this chapter.

in stocks as his wealth approaches the target level in order to minimize the risk of a shortfall due to an unexpected market downturn.

(3) Investors who are working and saving from their labor income should rely more heavily on stocks early in their working lives, not only because of the historically higher returns that stocks provide but also because bad stock market returns, early on, can be offset by increased saving out of labor income in later years.

Following Jagannathan and Kocherlakota (1996), we wish to subject these assertions to the discipline imposed by a rigorous modeling perspective. Let us maintain the assumptions of Theorem (14.1) and hypothesize that the risk free rate is constant, that stock returns $\{\tilde{r}_t\}$ are i.i.d., and that the investor's utility function assumes the standard CRRA form.

To evaluate assertion (1), let us further simplify Problem (14.1) by abstracting away from the consumption-savings problem. This amounts to assuming that the investor seeks to maximize the utility of his terminal wealth, Y_T in period T , the planned conclusion of his working life. As a result, $S_t = Y_t$ for every period $t < T$ (no intermediate consumption). Under CRRA we know that the investor would invest the same fraction of his wealth in risky assets every period (disproving the assertion), but it is worthwhile to see how this comes about in a simple multiperiod setting.

Let \tilde{r} denote the (invariant) risky return distribution; the investor solves:

$$\begin{aligned} \max_{\{a_t\}} E \left\{ \frac{(Y_T)^{1-\gamma}}{1-\gamma} \right\} \\ \text{s.t.} \quad & Y_T = a_{T-1}Y_{T-1}(1 + \tilde{r}_T) + (1 - a_{T-1})Y_{T-1}(1 + r_f), \quad t = T \\ & Y_t = a_{t-1}Y_{t-1}(1 + \tilde{r}_t) + (1 - a_{t-1})Y_{t-1}(1 + r_f), \quad 1 \leq t \leq T-1 \\ & Y_0 \text{ given.} \end{aligned}$$

Problems of this type are most appropriately solved by working backwards: first solving for the $T-1$ decision, then solving for the $T-2$ decision conditional on the $T-1$ decision and so on. In period $T-1$ the investor solves:

$$\max_{a_{T-1}} E \left\langle (1-\gamma)^{-1} \{ [a_{T-1} Y_{T-1}(1 + \tilde{r}) + (1 - a_{T-1})Y_{T-1}(1 + r_f)]^{(1-\gamma)} \} \right\rangle$$

The solution to this problem, $a_{T-1} \equiv \hat{a}$, satisfies the first order condition

$$E \{ [\hat{a}(1 + \tilde{r}) + (1 - \hat{a})(1 + r_f)]^{-\gamma} (\tilde{r} - r_f) \} = 0$$

As expected, because of the CRRA assumption the optimal fraction invested in stocks is independent of the period $T-1$ wealth level. Given this result, we

can work backwards. In period $T - 2$, the investor rebalances his portfolio, knowing that in $T - 1$ he will invest the fraction \hat{a} in stocks. As such, this problem becomes:

$$\max_{a_{T-2}} E \langle (1 - \gamma)^{-1} \{ [a_{T-2} Y_{T-2} (1 + \tilde{r}) + (1 - a_{T-2}) Y_{T-2} (1 + r_f)] [\hat{a}(1 + \tilde{r}) + (1 - \hat{a})(1 + r_f)] \}^{1-\gamma} \rangle \quad (14.2)$$

Because stock returns are i.i.d., this objective function may be written as the product of expectations as per

$$E[\hat{a}(1 + \tilde{r}) + (1 - \hat{a})(1 + r_f)]^{1-\gamma} \cdot \max_{\{a_{T-2}\}} E \{ (1 - \gamma)^{-1} [a_{T-2} Y_{T-2} (1 + \tilde{r}) + (1 - a_{T-2}) Y_{T-2} (1 + r_f)]^{1-\gamma} \} \quad (14.3)$$

Written in this way the structure of the problem is no different from the prior one, and the solution is again $a_{T-2} \equiv \hat{a}$. Repeating the same argument it must be the case that $a_t = \hat{a}$ in every period, a result that depends critically not only on the CRRA assumption (wealth factors out of the first order condition) but also on the independence. The risky return realized in any period does not alter our belief about the future return distributions. There is no meaningful difference between the long (many periods) run and the short run (one period): agents invest the same fraction in stocks irrespective of their portfolio's performance history. Assertion (1) is clearly not generally valid.

To evaluate our second assertion, and following again Jagannathan and Kocherlakota (1996), let us modify the agent's utility function to be of the form

$$U(Y_T) = \begin{cases} \frac{(Y_T - \bar{Y})^{1-\gamma}}{1-\gamma} & \text{if } Y_T \geq \bar{Y} \\ -\infty & \text{if } Y_T < \bar{Y} \end{cases}$$

where \bar{Y} is the target level of wealth. Under this formulation it is absolutely essential that the target be achieved: as long as there exists a positive probability of failing to achieve the target the investor's expected utility of terminal wealth is $-\infty$. Accordingly we must also require that

$$Y_0(1 + r_f)^T > \bar{Y};$$

in other words, that the target can be attained by investing everything in risk free assets. If such an inequality were not satisfied, then every strategy would yield an expected utility of $-\infty$, with the optimal strategy thus being indeterminate.

A straightforward analysis of this problem yields the following two step solution:

Step 1: always invest sufficient funds in risk free assets to achieve the target wealth level with certainty, and

Step 2: invest a constant share a^* of any additional wealth in stock, where a^* is time invariant.

By this solution, the investor invests less in stocks than he would in the absence of a target, but since he invests in both stocks and bonds, his wealth will accumulate, on average, more rapidly than it would if invested solely at the risk free rate, and the stock portion of his wealth will, on average, grow faster. As a result, the investor will typically use proportionally less of his resources to guarantee achievement of the target. And, over time, targeting will tend to increase the share of wealth in stocks, again contrary to popular wisdom!

In order to evaluate assertion (3), we must admit savings from labor income into the analysis. Let $\{L_t\}$ denote the stream of savings out of labor income. For simplicity, we assume that the stream of future labor income is fully known at date 0. The investor's problem is now:

$$\begin{aligned} \max_{\{a_t\}} E \left\{ \frac{(Y_T)^{1-\gamma}}{1-\gamma} \right\} \text{ s.t.} \\ Y_T = L_T + a_{T-1}Y_{T-1}(1 + \tilde{r}_T) + (1 - a_{T-1})Y_{T-1}(1 + r_f), \quad t = T \\ Y_t \leq L_t + a_{t-1}Y_{t-1}(1 + \tilde{r}_t) + (1 - a_{t-1})Y_{t-1}(1 + r_f), \quad 1 \leq t \leq T-1 \\ Y_0; \{L_t\}_{t=0}^T \text{ given.} \end{aligned}$$

We again abstract away from the consumption-savings problem and focus on maximizing the expected utility of terminal wealth.

In any period, the investor now has two sources of wealth, financial wealth, Y_t^F , where

$$Y_t^F = L_t + a_{t-1}Y_{t-1}(1 + r_t) + (1 - a_{t-1})Y_{t-1}(1 + r_f)$$

(r_t is the period t realized value of \tilde{r}), and “labor income wealth”, Y_t^L , is measured by the present value of the future stream of labor income. As mentioned, we assume this income stream is risk free with present value,

$$Y_t^L = \frac{L_{t+1}}{(1 + r_f)} + \dots + \frac{L_T}{(1 + r_f)^{T-1}}.$$

Since the investor continues to have CRRA preferences, he will, in every period, invest a constant fraction of his total wealth \hat{a} in stocks, where \hat{a} depends only upon his CRRA and the characteristics of the return distributions \tilde{r} and r_f ; i.e.,

$$A_t = \hat{a}(Y_t^F + Y_t^L),$$

where A_t denotes the *amount* invested in the risky financial asset.

As the investor approaches retirement, his Y_t^L declines. In order to maintain the same fraction of wealth invested in risk free assets, the fraction of financial wealth invested in stocks,

$$\frac{A_t}{Y_t^F} = \hat{a} \left(1 + \frac{Y_t^L}{Y_t^F} \right)$$

must decline on average. Here at least the assertion has theoretical support, but for a reason different from what is commonly asserted.

In what follows we will consider the impact on portfolio choice of a variety of changes to the myopic context just considered. In particular, we explore the consequences of relaxing the constancy of the risk free rate and return independence for the aforementioned recommendations. In most (but not all) of the discussion we'll assume an infinitely lived investor ($T = \infty$ in Problem (14.1)). Recall that this amounts to postulating that a finitely lived investor is concerned for the welfare of his descendants. In nearly every cases it enhances tractability. As a device for tying the discussion together, we will also explore how robust the three investor recommendations just considered are to a more general return environment.

Our first modification admits a variable risk free rate; the second generalizes the return generating process on the risky asset (no longer i.i.d. but 'mean reverting'). Our remarks are largely drawn from a prominent recent publication, Campbell and Viceira (2002).

Following the precedents established by these authors, it will prove convenient to log-linearize the investor's budget constraint and optimality conditions. Simple and intuitive expression for optimal portfolio proportions typically result. Some of the underlying deviations are provided in an Appendix available on this text's website; others are simply omitted when they are lengthy and complex and where an attractive intuitive interpretation is available.

In the next section the risk free rate is allowed to vary, although in a particularly structured way.

14.3 Variations in the Risk Free Rate

Following Campbell and Viceira (2002) we specialize Problem (14.1) to admit a variable risk free rate. Other assumptions are:

- (i) $L_t \equiv 0$ for all t ; there is no labor income so that all consumption comes from financial wealth alone;
- (ii) $T = \infty$, that is, we explore the infinite horizon version of Problem (14.1); this allows a simplified description of the optimality conditions on portfolio choice;
- (iii) All relevant return random variables are lognormal with constant variances and covariances. This is an admittedly strong assumption as it mandates that the return on the investor's portfolio has a constant variance, and that the constituent assets have constant variances and covariances with the portfolio itself. Thus, the composition of the risky part of the investor's portfolio must itself be invariant. But this will be optimal only if the expected excess returns above the risk free rate on these same constituent assets are also constant. Expected returns can vary over time but, in effect, they must move in tandem with

the risk free rate. This assumption is somewhat specialized but it does allow for unambiguous conclusions.

(iv) The investor's period utility function is of the Epstein-Zin variety (cf. Section 4.7). In this case the intertemporal optimality condition for Problem (14.1) when $T = \infty$ and there are multiple risky assets, can be expressed as

$$1 = E_t \left\{ \left[\delta \left(\frac{C_{t+1}}{C_t} \right)^{-\frac{1}{\rho}} \right]^\theta \left[\frac{1}{\tilde{R}_{P,t+1}} \right]^{1-\theta} \tilde{R}_{i,t+1} \right\} \quad (14.4)$$

where $\tilde{R}_{i,t}$ is the period t gross return on any available asset (risk free or otherwise, including the portfolio itself) and $\tilde{R}_{P,t}$ is the period t overall risky portfolio's gross return. Note that, consumption C_t , and the various returns $\tilde{R}_{P,t}$ and $\tilde{R}_{i,t}$ are capitalized; we will henceforth denote the logs of these quantities by their respective lower case counterparts.⁴ Equation (14.4) is simply a restatement of equation (9.28) where γ is the risk aversion parameter, ρ is the elasticity of intertemporal substitution and $\theta = \frac{(1-\gamma)}{(1-\frac{1}{\rho})}$.

Bearing in mind assumptions (i) - (iv), we now proceed, first to the investor's budget constraint, and then to his optimality condition. The plan is to loglinearize each in a (rather lengthy) development to our ultimate goal, equation (14.20).

14.3.1 The budget constraint

In a model with period consumption exclusively out of financial wealth, the intertemporal budget constraint is of the form

$$Y_{t+1} = (R_{P,t+1})(Y_t - C_t), \quad (14.5)$$

where the risky portfolio P potentially contains many risky assets; equivalently,

$$\frac{Y_{t+1}}{Y_t} = (R_{P,t+1})\left(1 - \frac{C_t}{Y_t}\right), \quad (14.6)$$

or, taking the log of both sides of the equation,

$$\Delta y_{t+1} = \log Y_{t+1} - \log Y_t = \log(R_{P,t+1}) + \log(1 - \exp(\log C_t - \log Y_t)).$$

Recalling our identification of a lower case variable with the log of that variable we have

$$\Delta y_{t+1} = r_{P,t+1} + \log(1 - \exp(c_t - y_t)).$$

⁴At least with respect to returns, this new identification is consistent with our earlier notation in the following sense : in this chapter we identify $r_t \equiv_{def} \log R_t$. In earlier chapters r_t denoted the net return by the identification $R_t \equiv 1 + r_t$. However, $r_t \equiv_{def} \log R_t = \log(1 + r_t) \approx r_t$, for net returns that are not large. Thus even in this chapter we may think of r_t as the net period t return.

Assuming that the $\log(\frac{C_t}{Y_t})$ is not too variable (essentially this places us in the $\rho = \gamma = 1$ – the log utility case), then the right most term can be approximated around its mean to yield (see Taylor, Campbell and Viceira (2001)):

$$\Delta y_{t+1} = k_1 + r_{P,t+1} + (1 - \frac{1}{k_2})(c_t - y_t), \quad (14.7)$$

where k_1 and $k_2 < 1$ are constants related to $\exp(E(c_t - y_t))$.

Speaking somewhat informally in a fashion that would identify the log of a variable with the variable itself, equation (14.7) simply states that wealth will be higher next period ($t + 1$) in a manner that depends on both the portfolio's rate of return ($r_{P,t+1}$) over the next period and on this period's consumption relative to wealth. If c_t greatly exceeds y_t , wealth next period cannot be higher!

We next employ an identity to allow us to rewrite (14.7) in a more useful way; it is

$$\Delta y_{t+1} = \Delta c_{t+1} + (c_t - y_t) - (c_{t+1} - y_{t+1}) \quad (14.8)$$

where $\Delta c_{t+1} = c_{t+1} - c_t$. Substituting the R.H.S. of equation (14.8) into (14.7) and rearranging terms yields

$$(c_t - y_t) = k_2 k_1 + k_2(r_{P,t+1} - \Delta c_{t+1}) + k_2(c_{t+1} - y_{t+1}). \quad (14.9)$$

Equation (14.9) provides the same information as equation (14.7) albeit expressed differently. It states that an investor could infer his (log) consumption-wealth ratio ($c_t - y_t$) in period t from a knowledge of its corresponding value in period $t + 1$, ($c_{t+1} - y_{t+1}$) and his portfolio's return (the growth rate of his wealth) relative to the growth rate of his consumption ($r_{P,t+1} - \Delta c_{t+1}$). (Note that our use of language again informally identifies a variable with its log.)

Equation (14.9) is a simple difference equation which can be solved forward to yield

$$c_t - y_t = \sum_{j=1}^{\infty} (k_2)^j (r_{P,t+j} - \Delta c_{t+j}) + \frac{k_2 k_1}{1 - k_2}. \quad (14.10)$$

Equation (14.10) also has an attractive intuitive interpretation; a high (above average) consumption-wealth ratio ($(c_t - y_t)$ large and positive); i.e., a burst of consumption, must be followed either by high returns on invested wealth or lowered future consumption growth. Otherwise the investor's intertemporal budget constraint cannot be satisfied. But (14.10) holds *ex ante* relative to time t as well as *ex post*, its current form. Equation (14.11) provides the *ex ante* version:

$$c_t - y_t = E_t \sum_{j=1}^{\infty} (k_2)^j (r_{P,t+j} - \Delta c_{t+j}) + \frac{k_2 k_1}{1 - k_2}. \quad (14.11)$$

Substituting this expression twice into the R.H.S. of (14.8), substituting the R.H.S. of (14.7) for the L.H.S. of (14.8), and collecting terms yields our final representation for the log-linearized budget constraint equation :

$$\begin{aligned}
c_{t+1} - E_t c_{t+1} &= (E_{t+1} - E_t) \sum_{j=1}^{\infty} (k_2)^j r_{P,t+1+j} \\
&\quad - (E_{t+1} - E_t) \sum_{j=1}^{\infty} (k_2)^j \Delta c_{t+1+j}. \tag{14.12}
\end{aligned}$$

This equation again has an intuitive interpretation: if consumption in period $t + 1$ exceeds its period t expectation ($c_{t+1} > E_t c_{t+1}$, a positive consumption “surprise”), then this consumption increment must be “financed” either by an upward revision in expected future portfolio returns (the first term on the L.H.S. of (14.12)) or a downward revision in future consumption growth (as captured by the second term on the L.H.S. of (14.12)). If it were otherwise, the investor would receive “something for nothing” – as though his budget constraint could be ignored.

Since our focus is on deriving portfolio proportions and returns, it will be useful to be able to eliminate future consumption growth (the Δc_{t+1+j} terms) from the above equation, and to replace it with an expression related only to returns. The natural place to look for such an equivalence is the investor’s optimality equation, (14.4), which directly relates the returns on his choice of optimal portfolio to his consumption experience, log-linearized so as to be in harmony with (14.12).

14.3.2 The Optimality Equation

The log-linearized version of (14.4) is:

$$E_t \Delta c_{t+1} = \rho \log \delta + \rho E_t r_{P,t+1} + \frac{\theta}{2\rho} \text{var}_t(\Delta c_{t+1} - \rho r_{P,t+1}), \tag{14.13}$$

where we have specialized equation (14.4) somewhat by choosing the i th asset to be the portfolio itself so that $R_{i,t+1} = R_{P,t+1}$. The web appendix provides a derivation of this expression, but it is more important to grasp what it is telling us about an Epstein-Zin investor’s optimal behavior: in our partial equilibrium setting where investors take return processes as given, equation (14.13) states that an investor’s optimal expected consumption growth ($E_t(\Delta c_{t+1})$) is linearly (by the log linear approximation) related to the time preference parameter δ (an investor with a bigger δ will save more and thus his expected consumption growth will be higher), the portfolio returns he expects to earn ($E_t r_{P,t+1}$), and the miscellaneous effects of uncertainty as captured by the final term $\frac{\theta}{2\rho} \text{var}_t(\Delta c_{t+1} - \rho r_{P,t+1})$. A high intertemporal elasticity of substitution ρ means that the investor is willing to experience a steeper consumption growth profile if there are incentives to do so and thus ρ premultiplies both $\log \delta$ and $E_t r_{P,t+1}$. Lastly, if $\theta > 0$, an increase in the variance of consumption growth relative to portfolio returns leads to a greater expected consumption growth

profile. Under this condition the variance increase elicits greater precautionary savings in period t and thus a greater expected consumption growth rate.

Under assumption (iii) of this section, however, the variance term in (14.13) is constant, which leads to a much-simplified representation

$$E_t \Delta c_{t+1} = k_3 + \rho E_t r_{P,t+1}, \quad (14.14)$$

where the constant k_3 incorporates both the constant variance and the time preference term $\rho \log \delta$. Substituting (14.14) into (14.11) in the most straightforward way and rearranging terms yields

$$c_t - y_t = (1 - \rho) E_t \sum_{j=1}^{\infty} (k_2)^j r_{P,t+j} + \frac{k_2(k_1 - k_3)}{1 - k_2} \quad (14.15)$$

Not surprisingly, equation (14.15) suggests that the investor's (log) consumption to wealth ratio (itself a measure of how willing he is to consume out of current wealth) depends linearly on future discounted portfolio returns, negatively if $\rho > 1$ and positively if $\rho < 1$ where ρ is his intertemporal elasticity of substitution. The value of ρ reflects the implied dominance of the substitution over the income effect. If $\rho < 1$, the income effect dominates: if portfolio returns increase, the investor can increase his consumption permanently without diminishing his wealth. If the substitution effect dominates ($\rho > 1$), however, the investor will reduce his current consumption in order to take advantage of the impending higher expected returns. Substituting (14.15) into (14.12) yields

$$c_{t+1} - E_t c_{t+1} = r_{P,t+1} - E_t r_{P,t+1} + (1 - \rho)(E_{t+1} - E_t) \sum_{j=1}^{\infty} (k_2)^j r_{P,t+1+j}, \quad (14.16)$$

an equation that attributes period $t + 1$'s consumption surprise to (1) the unexpected contemporaneous component to the overall portfolio's return $r_{P,t+1} - E_t r_{P,t+1}$, plus, (2) the revision in expectation of future portfolio returns, $(E_{t+1} - E_t) \sum_{j=1}^{\infty} (k_2)^j r_{P,t+1+j}$. This revision either encourages or reduces consumption depending on whether, once again, the income or substitution effect dominates. This concludes the background on which the investor's optimal portfolio characterization rests. Note that equation (14.16) defines a relationship by which consumption may be replaced – in some other expression of interest – by a set of terms involving portfolio returns alone.

14.3.3 Optimal Portfolio Allocations

So far, we have not employed the assumption that the expected returns on all assets move in tandem with the risk free rate, and, indeed the risk free rate is not explicit in any of expressions (14.2) – (14.14). We address these issues presently.

In an Epstein-Zin context, recall that the risk premium on any risky asset over the safe asset, $E_t r_{P,t+1} - r_{f,t+1}$, is given by equation (9.32) which is recopied below:

$$E_t r_{t+1} - r_{f,t+1} + \frac{\sigma_t^2}{2} = \frac{\theta \text{cov}_t(r_{t+1}, \Delta c_{t+1})}{\rho} + (1 - \theta) \text{cov}_t(r_{t+1}, r_{P,t+1}) \quad (14.17)$$

where r_{t+1} denotes the return on the stock portfolio, $r_{P,t+1}$ is the return on the portfolio of all the investor's assets, that is, including the "risk free" one. Note that implicit in assumption (iii) is the recognition that all variances and covariances are constant despite the time dependency in notation.

From expression (14.16) we see that the covariance of (log) consumption with any variable (and we have in mind its covariance with the risky return variable of (14.17)) may be replaced by the covariance of that variable with the portfolio's contemporaneous return plus $(1 - \rho)$ times the expectations revisions concerning future portfolio returns. Eliminating in this way consumption from (14.17) via a judicious insertion of (14.16) yields

$$\begin{aligned} E_t r_{t+1} - r_{f,t+1} + \frac{\sigma_t^2}{2} &= \gamma \text{cov}_t(r_{t+1}, r_{P,t+1}) + \\ &(\gamma - 1) \text{cov}_t \left(r_{t+1}, (E_{t+1} - E_t) \sum_{j=1}^{\infty} (k_2)^j r_{P,t+1+j} \right). \end{aligned} \quad (14.18)$$

As noted in Campbell and Viceira (2002), equations (14.16) and (14.18) delineate in an elegant way the consequences of the Epstein and Zin separation of time and risk preferences. In particular, in equation (14.16) it is only the time preference parameter ρ which relates current consumption to future returns (and thus income) – a time preference effect, while, in (14.18), it is only γ , the risk aversion coefficient, that appears to influence the risk premium on the risky asset.

If we further recall (assumption (iii)) that variation in portfolio expected returns must be exclusively attributable to variation in the risk-free rate, it follows logically that revisions in expectations of the former must uniquely follow from revisions of expectations of the latter:

$$(E_{t+1} - E_t) \sum_{j=1}^{\infty} (k_2)^j r_{P,t+1+j} = (E_{t+1} - E_t) \sum_{j=1}^{\infty} (k_2)^j r_{f,t+1+j}. \quad (14.19)$$

In a model with one risky asset (in effect the risky portfolio whose composition we are a-priori holding constant),

$$\text{cov}_t(r_{t+1}, r_{p,t+1}) = a_t \sigma_t^2 = a_t \sigma_{P,t}^2$$

where a_t is, as before, the risky asset proportion in the portfolio.

Substituting both this latter expression and identification (14.19) into equation (14.18) and solving for a_t gives the optimal, time invariant portfolio weight

on the risky asset.

$$a_t \equiv a = \frac{1}{\gamma} \left[\frac{E_t r_{t+1} - r_{f,t+1} + \frac{\sigma_t^2}{2}}{\sigma_t^2} \right] + \left(1 - \frac{1}{\gamma}\right) \frac{1}{\sigma_t^2} \text{cov}_t \left(r_{t+1}, -(E_{t+1} - E_t) \sum_{j=1}^{\infty} (k_2)^j r_{f,t+1+j} \right), \quad (14.20)$$

our first portfolio result. Below we offer a set of interpretative comments related to it.

(1) The first term in (14.20) represents the myopic portfolio demand for the risky asset, myopic in the sense that it describes the fraction of wealth invested in the risky portfolio when the investor ignores the possibility of future risk free rate changes. In particular, the risky portfolio proportion is inversely related to the investor's CRRA (γ), and positively related to the risk premium. Note, however, that these rate changes are the fundamental feature of this economy in the sense that variances are fixed and all expected risky returns move in tandem with the risk free rate.

(2) The second term in (14.20) captures the risky asset demand related to its usefulness for hedging intertemporal interest rate risk. The idea is as follows. We may view the risk free rate as the "base line" return on the investor's wealth with the risky asset providing a premium on some fraction thereof. If expected future risk free returns are revised downwards, $(-(E_{t+1} - E_t) \sum_{j=1}^{\infty} (k_2)^j r_{f,t+1+j})$ increases), the investor's future income (consumption) stream will be reduced unless the risky asset's return increases to compensate. This will be so on average if the covariance term in equation (14.20) is positive. It is in this sense that risky asset returns (r_{t+1}) can hedge risk free interest rate risk. If the covariance term is negative, however, risky asset returns only tend to magnify the consequences of a downward revision in expected future risk free rates. As such, a long term investor's holding of risky assets would be correspondingly reduced.

These remarks have their counterpart in asset price changes: if risk free rates rise (bond prices fall) the investor would wish for changes in the risky portion of his portfolio to compensate via increased valuations.

(3) As the investor becomes progressively more risk averse ($\gamma \mapsto \infty$), he will continue to hold stocks in his portfolio, but only because of their hedging qualities, and not because of any return premium they provide. An analogous myopic investor would hold no risky assets.

(4) Note also that the covariance term in (14.20) depends on changes in expectations concerning the entire course of future interest rates. It thus follows that the investor's portfolio allocations will be much more sensitive to persistent changes in the expected risk free rate than to transitory ones.

Considering all the complicated formulae that have been developed, the conclusions thus far are relatively modest. An infinitely lived investor principally

consumes out of his portfolio's income and he wishes to maintain a stable consumption series. To the extent that risky equity returns can offset (hedge) variations in the risk free rate, investors are provided justification for increasing the share of their wealth invested in the high return risky asset. This leads us to wonder if any asset can serve as a truly risk free one for the long term investor.

14.3.4 The Nature of the Risk Free Asset

Implicit in the above discussion is the question of what asset, if any, best serves as the risk free one. From the long term investor's point of view it clearly cannot be a short-term money market instrument (e.g., a T-bill), because its well-documented rate variation makes uncertain the future reinvestment rates that the investor will receive.

We are reminded at this juncture, however, that it is not the return risk, per se, but the derived consumption risk that is of concern to investors. Viewed from the consumption perspective, a natural candidate for the risk free asset is an indexed consol bond which pays (the real monetary equivalent of) one unit of consumption every period. Campbell, Lo, and MacKinlay (1997) show that the (log) return on such a consol is given by

$$r_{c,t+1} = r_{f,t+1} + k_4 - (E_{t+1} - E_t) \sum_{j=1}^{\infty} (k_5)^j r_{f,t+1+j} \quad (14.21)$$

where k_4 is a constant measuring the (constant) risk premium on the consol, and k_5 is another positive constant less than one. Suppose, as well that we have an infinitely risk averse investor ($\gamma = \infty$) so that (14.20) reduces to

$$a = \frac{1}{\sigma_t^2} \text{cov}_t \left(r_{t+1}, -(E_{t+1} - E_t) \sum_{j=1}^{\infty} (k_2)^j r_{f,t+1+j} \right), \quad (14.22)$$

and that the single risky asset is the consol bond ($r_{t+1} = r_{c,t+1}$). In this case (substituting (14.21) into (14.22) and observing that constants do not matter for the computing of covariances), $a \equiv 1$: the highly risk averse investor will eschew short term risk free assets and invest entirely in indexed bonds. This alone will provide him with a risk free consumption stream, although the value of the asset may change from period to period.

14.3.5 The Role of Bonds in Investor Portfolios

Now that we allow the risk free rate to vary, let us return to the three life cycle portfolio recommendations mentioned in the myopic choice section of this chapter. Of course, the model – with an infinitely-lived investor – is, by construction, not the appropriate one for the life cycle issues of recommendation 2 and, being without labor income, nothing can be said regarding recommendation 3 either. This leaves the first recommendation, which really concerns the portfolio of choice for long term investors. The single message of this chapter

subsection must be that conservative long term investors should invest the bulk of their wealth in long term index bonds. If such bonds are not available, then in an environment of low inflation risk, long term government securities are a reasonable, second best substitute. For persons entering retirement – and likely to be very concerned about significant consumption risk – long term real bonds should be the investment vehicle of choice.

This is actually a very different recommendation from the static one period portfolio analysis which would argue for a large fraction of a conservative investor's wealth being assigned to risk free assets (T-bills). Yet we know that short rate uncertainty, which the long term investor would experience every time she rolled over her short term instruments, makes such an investment strategy inadvisable for the long term.

14.4 The Long Run Behavior of Stock Returns

Should the proportion of an investor's wealth invested in stocks differ systematically for long term versus short term investors? In either case, most of the attractiveness of stocks (by stocks we will continue to mean a well diversified stock portfolio) to investors lies in their high excess returns (recall the equity premium puzzle of Chapter 9). But what about long versus short term equity risk; that is, how does the ex ante return variance of an equity portfolio held for many periods compare with its variance in the short run?

The ex post historical return experience of equities versus other investments turns out to be quite unexpected in this regard.

From Table 14.1 it is readily apparent that, historically, over more than 20 year time horizons, stocks have never yielded investors a negative real annualized return, while for all other investment types this has been the case for some sample period. Are stocks in fact less risky than bonds for an appropriately "long run"?

In this section, we propose to explore this issue via an analysis of the following questions:

(1) What are the intertemporal equity return patterns in order that the outcomes portrayed in Table 14.1 be pervasive and not just represent the realizations of extremely low-probability events;

(2) Given a resolution of (1), what are the implications for the portfolio composition of long term versus short term investors; and,

(3) How does a resolution of questions (1) and (2) modify the myopic response to the long run investment advice of Section 14.2?

It is again impossible to answer these questions in full generality. Following Campbell and Viceira (1999), we elect to examine investor portfolios composed of one risk free and one risky asset (a diversified portfolio). Otherwise, the context is as follows:

(i) the investor is infinitely-lived with Epstein-Zin preferences so that (14.4) remains as the investor's intertemporal optimality condition; furthermore, the investor has no labor income;

Table 14.1⁽ⁱ⁾: Minimum and Maximum Actual Annualized Real Holding Period Returns for the Period 1802-1997 U.S. Securities Markets Variety of Investment Options

Maximum Observed Return			Minimum Observed Return
One Year Holding Period			
Stocks	66.6%		−38.6%
Bonds	35.1%		−21.9%
T-Bills	23.7%		−15.6%
Two Year Holding Period			
Stocks	41.0%		−31.6%
Bonds	24.7%		−15.9%
T-Bills	21.6%		−15.1%
Five Year Holding Period			
Stocks	26.7%		−11.0%
Bonds	17.7%		−10.1%
T-Bills	14.9%		−8.2%
Ten Year Holding Period			
Stocks	16.9%		−4.1% ⁽ⁱⁱ⁾
Bonds	12.4%		−5.4%
T-Bills	11.6%		−5.1%
Twenty Year Holding Period			
Stocks	12.6%		1.0%
Bonds	8.8%		−3.1%
T-Bills	8.3%		−3.0%
Thirty Year Holding Period			
Stocks	10.6%		2.6%
Bonds	7.4%		−2.0%
T-Bills	7.6%		−1.8%

⁽ⁱ⁾ Source: Siegel (1998), Figure 2-1

⁽ⁱⁱ⁾ Notice that beginning with a ten year horizon, the minimum observed stock return exceeded the corresponding minimum bill and bond returns.

(ii) the log real risk free rate is constant from period to period. Under this assumption, all risk free assets – long or short term – pay the same annualized return. The issues of Section 14.3 thus cannot be addressed;

(iii) the equity return generating process builds on the following observations:

First, note that the cumulative log return over T periods under an i.i.d. assumption is given by

$$r_{t+1} + r_{t+2} + \dots + r_{t+T},$$

so that

$$\text{var}(r_{t+1} + r_{t+2} + \dots + r_{t+T}) = T\text{var}(r_{t+1}) > T\text{var}(r_{f,t+1}).$$

For U.S. data, $\text{var}(r_{t+1}) \approx (.167)^2$ (taking the risky asset as the S&P500 market index) and $\text{var}(r_{f,t+1}) = (.057)^2$ (measuring the risk free rate as the one-year T-bill return). With a $T = 20$ year time horizon, the observed range of annualized relative returns given by Table 14.1 is thus extremely unlikely to have arisen from an i.i.d. process. What could be going on?

For an i.i.d. process, the large relative twenty-year variance arises from the possibility of long sequences, respectively, of high and low returns. But if cumulative stock returns are to be less variable than bond returns at long horizons, some aspect of the return generating process must be discouraging these possibilities. That aspect is referred to as “mean reversion”: the tendency of high returns today to be followed by low returns tomorrow on an expected basis and vice versa. It is one aspect of the “predictability” of stock returns and is well documented beyond the evidence in Table 14.1.⁵

⁵It is well known that stock returns are predicted by a number of disparate variables. Perhaps the most frequently cited predictive variable is $\log(Dt/Pt) = d_t - p_t$, the log of the dividend/price ratio at long horizons. In particular, regressions of the form

$$r_{t,t+k} \equiv r_{t+1} + \dots + r_{t+k} = \beta_k(d_t - p_t) + \varepsilon_{t,t+k}$$

obtain a R^2 of an order of magnitude of .3. In the above expression r_{t+j} denotes the log return on the value weighted index portfolio comprising all NYSE, AMEX, and NASDAQ stocks in month $t + j$, d_t is the log of the sum of all dividends paid on the index over the entire year preceding period t , and P_t denotes the period t value of the index portfolio. See Campbell et. al. (1997) for a detailed discussion. More recently, Santos and Veronesi (2004) study regressions whereby long horizon excess returns (above the risk free rate) are predicted by lagged values of the (U.S. data) aggregate labor income/consumption ratio:

$$r_{t+k} = \alpha_1 + \beta_k s_t^w + \varepsilon_{t+k},$$

where $s_t^w = w_t/c_t$; w_t is measured as period t total compensation to employees and c_t denotes consumption of non-durables plus services (quarterly data). For the period 1948-2001, for example, they obtain an adjusted R^2 of .42 for $k = 16$ quarters. Returns are computed in a manner identical to Campbell et al. (1997) just mentioned. The basic logic is as follows: when the labor income/consumption ratio is high, investors are less exposed to stock market fluctuations (equity income represents a small fraction of total consumption) and hence demand a lower premium. Stock prices are thus high. Since the s_t^w ratio is stationary (and highly persistent in the data) it will eventually return to its mean value suggesting a lower future tolerance for risk, a higher risk premium, lower equity prices and low future returns. Their statistical analysis concludes that the labor income/consumption ratio does indeed move in a direction opposite to long horizon returns. Campbell and Cochrane (1999)

Campbell and Viceira (1999) statistically model the mean reversion in stock returns in a particular way that facilitates the solution to the associated portfolio allocation problem. In particular, they assume the time variation in log return on the risky asset is captured by:

$$r_{t+1} - E_t r_{t+1} = u_{t+1}, \quad u_{t+1} \sim N(0, \sigma_u^2), \quad (14.23)$$

where u_{t+1} captures the unexpected risky return component or “innovation.” In addition, the expected premium on this risky asset is modeled as evolving according to:

$$E_t r_{t+1} - r_f + \frac{\sigma_u^2}{2} = x_t, \quad (14.24)$$

where x_t itself is a random variable following an AR(1) process with mean \bar{x} , persistence parameter ϕ , and random innovation $\tilde{\eta}_{t+1} \sim N(0, \sigma_\eta^2)$:

$$x_{t+1} = \bar{x} + \phi(x_t - \bar{x}) + \tilde{\eta}_{t+1}. \quad (14.25)$$

The x_t random variable thus moves slowly (depending on ϕ) with a tendency to return to its mean value. Lastly, mean reversion is captured by assuming $\text{cov}(\eta_{t+1}, u_{t+1}) = \sigma_{\eta u} < 0$, which translates, as per below, into a statement about risky return autocorrelations:

$$\begin{aligned} 0 > \sigma_{\eta u} = \text{cov}(u_{t+1}, \eta_{t+1}) &= \text{cov}_t[(r_{t+1} - E_t r_{t+1}), (x_{t+1} - \bar{x} - \phi(x_t - \bar{x}))] \\ &= \text{cov}_t(r_{t+1}, x_{t+1}) \\ &= \text{cov}_t(r_{t+1}, E_t r_{t+2} - r_f + \frac{\sigma_u^2}{2}) \\ &= \text{cov}_t(r_{t+1}, r_{t+2} - u_{t+2} - r_f + \frac{\sigma_u^2}{2}) \\ &= \text{cov}_t(r_{t+1}, r_{t+2}), \end{aligned}$$

a high return today reduces expected returns next period. Thus,

$$\text{var}_t(r_{t+1} + r_{t+2}) = 2\text{var}_t(r_{t+1}) + 2\text{cov}_t(r_{t+1}, r_{t+2}) < 2\text{var}_t(r_{t+1}),$$

in contrast to the independence case. More generally, for all horizons k ,

$$\frac{\text{var}_t(r_{t+1} + r_{t+2} + \dots + r_{t+k})}{k\text{var}_t(r_{t+1})} < 1 :$$

stocks will appear to be relatively safer to long term investors. This concludes assumption (iii) and its interpretation. We next explore what this environment implies for optimal investment proportions.

are also able to replicate many of the predictability results of the empirical literature using a slow moving external habit formation model.

⁶We are reminded that the presence of the $\frac{\sigma_u^2}{2}$ term in (14.24) follows from the fact that for any lognormal random variable z , $\log E_t z_{t+1} = E_t \log z_{t+1} + \frac{1}{2} \text{var}_t \log z_{t+1}$

14.4.1 Solving for Optimal Portfolio Proportions in a Mean Reversion Environment

Campbell and Viceira (1999) manipulate the log linearized version of (the optimality condition for Epstein-Zin utility (14.4) in conjunction with (14.23) - (14.25) to obtain:

$$a_t = a_0 + a_1 x_t \quad (14.26)$$

$$c_t - y_t = b_0 + b_1 x_t + b_2 x_t^2, \quad (14.27)$$

where a_t is the (time varying) wealth proportion invested in the risky asset and (14.27) describes the behavior of the (log) consumption-wealth ratio. The terms a_0 , a_1 , b_0 , b_1 , b_2 , are constants; of special interest are those defining the time varying risky asset proportion:

$$a_0 = (1 - \frac{1}{\gamma}) \left[\left(\frac{b_1}{1 - \rho} \right) + 2\bar{x}(1 - \phi) \left(\frac{b_2}{1 - \rho} \right) \right] \left(\frac{-\sigma_{\eta u}}{\sigma_u^2} \right), \text{ and} \quad (14.28)$$

$$a_1 = \frac{1}{\gamma \sigma_u^2} + (1 - \frac{1}{\gamma}) \left[2\phi \frac{b_2}{(1 - \rho)} \right] \left(\frac{-\sigma_{\eta u}}{\sigma_u^2} \right). \quad (14.29)$$

Note that in the case of a myopic investor ($\gamma = 1$) who holds stocks only because of the (conditional) excess returns they offer, $a_0 = 0$ and $a_1 = \frac{1}{\sigma_u^2}$. This suggests that all the other terms present in the expressions defining a_0 and a_1 must be present to capture some aspect of a non-myopic investor's *intertemporal hedging demand*.

What is meant by this latter expression in this particular context? With $\bar{x} > 0$, investors are typically long in stocks (the risky asset) in order to capture the excess returns they provide on average. Suppose in some period $t + 1$, stock returns are high, meaning that stock prices rose a lot from t to $t + 1$ (u_{t+1} is large). To keep the discussion less hypothetical, let's identify this event with the big run-up in stock prices in the late 1990s. Under the $\sigma_{\eta u} < 0$ assumption, expected future returns are likely to decline, and perhaps even become negative (η_{t+1} is small, possibly negative so that x_{t+1} is small and thus, via (14.24), so is Er_{t+1}). Roughly speaking this means stock prices are likely to decline - as they did in the 2000-2004 period! In anticipation of future price declines, long term investors would rationally wish to assemble a short position in the risky portfolio, since this is the only way to enhance their wealth in the face of falling prices (r_f is constant by assumption). Most obviously, this is a short position in the risky portfolio itself, since negative returns must be associated with falling prices.

These thoughts are fully captured by (14.26) -(14.27). Campbell and Viceira (2002) argue that the empirically relevant case is the one for which $\bar{x} > 0$, $\frac{b_1}{1 - \rho} > 0$, $\frac{b_2}{1 - \rho} > 0$, and $\sigma_{\eta u} < 0$. Under these circumstances, $a_0 > 0$, and $a_1 > 0$, for a sufficiently risk averse investor ($\gamma > 1$). If u_{t+1} is large then η_{t+1} is likely to be small - let's assume negative - and 'large' in absolute value if $|\sigma_{\eta u}|$

is itself large. Via portfolio allocation equation (14.26), the optimal $a_t < 0$ - a short position in the risky asset.

This distinguishing feature of long term risk averse investors is made more striking if we observe that with $\sigma_{\eta u} < 0$, such an investor will maintain a position in the risky asset if average excess returns, $\bar{x} = 0$: Even in this case $a_0 > 0$ (provided $\gamma > 1$). Thus if $x_t = 0$ (no excess returns to the risky asset), the proportion of the investor's wealth in stocks is still positive. In a one period CAPM investment universe, a mean-variance myopic investor would invest nothing in stocks under these circumstances. Neither would the myopic expected utility maximizer of Theorem 4.1.

All this is to observe that a risk averse rational long term investor will use whatever means are open to him, including shorting stocks, when he (rationally) expects excess future stock returns to be sufficiently negative to warrant it. A major caveat to this line of reasoning, however, is that it cannot illustrate an equilibrium phenomenon: if all investors are rational and equally well informed about the process generating equity returns (14.23) – (14.25), then all will want simultaneously to go long or short. The latter, in particular, is not feasible from an equilibrium perspective.

14.4.2 Strategic Asset Allocation

The expression “strategic asset allocation” is suggestive not only of long-term investing (for which intertemporal hedging is a concern), but also of portfolio weights assigned to broad classes of assets (e.g., “stocks”, “long term bonds”) each well diversified from the perspective of its own kind. This is exactly the setting of this chapter.

Can the considerations of this section, in particular, be conveniently contrasted with those of the preceding chapters? This is captured in Figure (14.1) below under the maintained assumptions of this subsection (itself a replica of Figure 4.1 in Campbell and Viceira (2002)).

Insert Figure 14.1 about here

The myopic buy and hold strategy assumes a constant excess stock return equal to the true unconditional mean ($(E_t r_{t+1} - r_f + \frac{\sigma_u^2}{2} \equiv \bar{x})$) with the investor solving a portfolio allocation problem as per Theorem 4.1. The line marked “tactical asset allocation” describes the portfolio allocations for an investor who behaves as a one period investor, conditional on his observation of x_t . Such an investor, by definition, will not take account of long term hedging opportunities. Consistent with the CAPM recommendation, such an investor will elect $a_t \equiv 0$ when $x_t = 0$ (no premium on the risky asset) but if the future looks good – even for just one period – he will increase his wealth proportions in the risky assets. Again, by construction, if $x_t = \bar{x}$, such an investor will adopt portfolio proportions consistent with the perpetually myopic investor.

Long term “strategic” investors, with rational expectations vis-à-vis the return generating process (i.e., they know and fully take account of (14.23) –

(14.25)) will always elect to hold a greater proportion of their wealth in the risky portfolio than will be the case for the “tactical” asset allocator. In itself this is not entirely surprising for only he is able to illustrate the “hedging demand.” But this demand is present in a very strong way; in particular, even if excess returns are zero, the strategic investor holds a positive wealth fraction in risky assets ($a_0 > 0$). Note also that the slope of the strategic asset allocation line exceeds that of the tactical asset allocation line. In the context of (14.23-14.25)), this is a reflection of the fact that $\phi = 0$ for the tactical asset allocator.

14.4.3 The Role of Stocks in Investor Portfolios

Stocks are less risky in the long run because of the empirically verified mean reversion in stock returns. But does this necessarily imply a 100% stock allocation in perpetuity for long term investors? Under the assumptions of Campbell and Viceira (2002), this is clearly not the case: long term investors should be prepared to take advantage of mean reversion by *timing the market* in a manner illustrated in Figure 14.1 But this in turn presumes the ability of investors to short stocks when their realized returns have recently been very high. Especially for small investors, shorting securities may entail prohibitive transactions costs. Even more significantly, this cannot represent an equilibrium outcome for all investors.

14.5 Background Risk: The Implications of Labor Income for Portfolio Choice

Background risks refer to uncertainties in the components of an investor’s income not directly related to his tradeable financial wealth and, in particular, his stock - bond portfolio allocation. Labor income risk is a principal component of background risk; variations in proprietary income (income from privately owned businesses) and in the value of owner-occupied real estate are the others.

In this section we explore the significance of labor income risk for portfolio choice. It is a large topic and one that must be dealt with using models of varying complexity. The basic insight we seek to develop is as follows: an investor’s labor income stream constitutes an element of his wealth portfolio. The desirability of the risky asset in the investor’s portfolio will therefore depend not only upon its excess return (above the risk free rate) relative to its variance (risk), but also the extent to which it can be used to hedge variations in the investor’s labor income. Measuring how the proportion of an investor’s financial wealth invested in the risky asset depends on its hedging attributes in the above sense is the principal focus of this section. Fortunately it is possible to capture the basic insights in a very simple framework. As discussed in Campbell and Viceira (2002) that framework makes a number of assumptions:

(i) the investor has a one period horizon, investing his wealth to enhance his consumption tomorrow (as such the focus is on the portfolio allocation decision exclusively; there is no $t = 0$ simultaneous consumption-savings decision);

(ii) the investor receives labor income L_{t+1} , tomorrow, which for analytical simplicity is assumed to be lognormally distributed: $\log \tilde{L}_{t+1} \equiv \tilde{\ell}_{t+1} \sim N(l, \sigma_\ell^2)$;

(iii) there is one risk free and one risky asset (a presumed-to-be well diversified portfolio). Following our customary notation, $r_f = \log(R_f)$ and $\tilde{r}_{t+1} = \log(\tilde{R}_{t+1})$. Furthermore, $\tilde{r}_{t+1} - E_t \tilde{r}_{t+1} = \tilde{u}_{t+1}$ where $\tilde{u}_{t+1} \sim N(0, \sigma_u^2)$. The possibility is admitted that the risky asset return is correlated with labor income in the sense that $\text{cov}(\tilde{\ell}_{t+1}, \tilde{r}_{t+1}) = \sigma_{\ell u} \neq 0$.

(iv) the investor's period $t+1$ utility function is of the CRRA-power utility type, with coefficient of relative risk aversion γ .

Since there is no labor-leisure choice, this model is implicitly one of fixed labor supply in conjunction with a random wage.

Accordingly, the investor solves the following problem:

$$\max_{\alpha_t} E_t \left[\delta \frac{C_{t+1}^{1-\gamma}}{1-\gamma} \right] \quad (14.30)$$

$$\text{s.t. } C_{t+1} = Y_t R_{P,t+1} + L_{t+1},$$

where

$$R_{P,t+1} = a_t(R_{t+1} - R_f) + R_f, \quad (14.31)$$

a_t represents the fraction of the investor's wealth assigned to the risky portfolio, and P denotes his overall wealth portfolio. As in nearly all of our problems to date, insights can be neatly obtained only if approximations are employed which take advantage of the lognormal setup. In particular, we first need to modify the portfolio return expression (14.31)⁷. Since

$$\begin{aligned} R_{P,t+1} &= a_t R_{t+1} + (1 - a_t) R_f, \\ \frac{R_{P,t+1}}{R_f} &= 1 + a_t \left(\frac{R_{t+1}}{R_f} - 1 \right). \end{aligned}$$

Taking the log of both sides of this equation yields

$$r_{P,t+1} - r_f = \log[1 + a_t(\exp(r_{t+1} - r_f) - 1)]. \quad (14.32)$$

The right hand side of this equation can be approximated using a second order Taylor expansion around $r_{t+1} - r_f = 0$, where the function to be approximated is

$$g_t(r_{t+1} - r_f) = \log[1 + a_t(\exp(r_{t+1} - r_f) - 1)].$$

By Taylor's Theorem

$$g_t(r_{t+1} - r_f) \approx g_t(0) + g'_t(0)(r_{t+1} - r_f) + \frac{1}{2} g''_t(0)(r_{t+1} - r_f)^2.$$

⁷The derivation to follow is performed in greater detail in Campbell and Viceira (2001b)

Clearly $g_t(0) \equiv 0$; Straightforward calculations (simple calculus) yield $g'_t(0) = a_t$, and $g''_t(0) = a_t(1 - a_t)$. Substituting the Taylor expansion, with the indicated coefficient values, for the left hand side of (14.32) yields,

$$r_{P,t+1} - r_f = a_t(r_{t+1} - r_f) + \frac{1}{2}a_t(1 - a_t)\sigma_t^2$$

where $(r_{t+1} - r_f)^2$ is replaced by its conditional expectation. By the special form of the risky return generating process, $\sigma_t^2 = \sigma_u^2$, which yields

$$r_{P,t+1} = a_t(r_{t+1} - r_f) + r_f + \frac{1}{2}a_t(1 - a_t)\sigma_u^2. \quad (14.33)$$

We next modify the budget constraint to problem (14.30).

$$\frac{C_{t+1}}{L_{t+1}} = \frac{Y_t}{L_{t+1}}(R_{P,t+1}) + 1,$$

or taking the log of both sides of the equation,

$$\begin{aligned} c_{t+1} - \ell_{t+1} &= \log[\exp(y_t + r_{P,t+1} - \ell_{t+1}) + 1] \\ &\approx k + \xi(y_t + r_{P,t+1} - \ell_{t+1}), \end{aligned} \quad (14.34)$$

where k and ξ , $0 < \xi < 1$, are constants of approximation. Adding log labor income $-\ell_{t+1}$ to both sides of the equation yields

$$c_{t+1} = k + \xi(y_t + r_{P,t+1}) + (1 - \xi)\ell_{t+1}, \quad 1 > \xi > 0. \quad (14.35)$$

In other words, (log-) end of period consumption is a constant plus a weighted average of (log-) end-of-period financial wealth and (log-) labor income, with the weights $\xi, 1 - \xi$ serving to describe the respective elasticities of consumption with respect to these individual wealth components.

So far, nothing has been said regarding optimality. Problem (14.30) is a one period optimization problem. The first order necessary and sufficient condition for this problem with respect to a_t , the proportion of financial wealth invested in the portfolio, is given by:

$$E_t \left[\delta(\tilde{C}_{t+1})^{-\gamma}(\tilde{R}_{t+1}) \right] = E_t \left[\delta(\tilde{C}_{t+1})^{-\gamma}(R_f) \right]. \quad (14.36)$$

In loglinear form, equation (14.36) has the familiar form:

$$E_t(\tilde{r}_{t+1} - r_f) + 1/2\sigma_t^2 = \gamma \text{cov}_t(\tilde{r}_{t+1}, \tilde{c}_{t+1}).$$

Substituting the expression in (14.35) for c_{t+1} yields

$$E_t(\tilde{r}_{t+1} - r_f) + 1/2\sigma_t^2 = \gamma \text{cov}_t(\tilde{r}_{t+1}, k + \xi(y_t + \tilde{r}_{P,t+1}) + (1 - \xi)\tilde{\ell}_{t+1}).$$

After substituting (14.33) for $r_{P,t+1}$, we are left with

$$E_t(\tilde{r}_{t+1} - r_f) + 1/2\sigma_t^2 = \gamma \left[\xi a_t \sigma_t^2 + (1 - \xi) \text{cov}_t(\tilde{\ell}_{t+1}, \tilde{r}_{P,t+1}) \right].$$

from which we can solve directly for a_t .

Recall that our objective was to explore how the hedging (with respect to labor income) features of risky securities influence the proportion of financial wealth invested in the risky asset. Accordingly it is convenient first to simplify the expression via the following identifications: let

- (i) $\mu = E_t(\tilde{r}_{t+1} - r_f)$;
- (ii) $\sigma_t^2 = \sigma_u^2$, since $r_{t+1} - E_t r_{t+1} = u_{t+1}$;
- (iii) $\text{cov}(\tilde{\ell}_{t+1}, \tilde{r}_{t+1}) = \text{cov}(\tilde{\ell}_{t+1}, \tilde{r}_{t+1} - E_t \tilde{r}_{t+1}) = \text{cov}(\tilde{\ell}_{t+1}, \tilde{u}_{t+1}) = \sigma_{\ell u}$

With these substitutions the above expression reduces to:

$$\mu + \frac{1}{2}\sigma_u^2 = \gamma[\xi a_t \sigma_u^2 + (1 - \xi)\sigma_{\ell u}].$$

Straightforwardly solving for a_t yields

$$a_t = \frac{1}{\xi} \left(\frac{\mu + \frac{\sigma_u^2}{2}}{\gamma \sigma_u^2} \right) + \left(1 - \frac{1}{\xi} \right) \frac{\sigma_{\ell u}}{\sigma_u^2}, \quad (14.37)$$

an expression with an attractive interpretation. The first term on the right hand side of (14.37) represents the fraction in the risky asset if labor income is uncorrelated with the risky asset return ($\sigma_{\ell u} = 0$). It is positively related to the adjusted return premium ($\mu + \frac{\sigma_u^2}{2}$) and inversely related to the investor's risk aversion coefficient γ . The second term represents the hedging component: if $\sigma_{\ell u} < 0$, then since $\xi < 1$, demand for the risky asset is enhanced since it can be employed to diversify away some of the investor's labor income risk. Or, to express the same idea from a slightly different perspective, if the investor's labor income has a "suitable" statistical pattern vis-a-vis the stock market, he can reasonably take on greater financial risk.

It is perhaps even more striking to explore further the case where $\sigma_{\ell u} = 0$: since $\xi < 1$, even in this case the optimal fraction invested in the risky portfolio

$$a_t = \frac{1}{\xi} \left(\frac{\mu + \frac{\sigma_u^2}{2}}{\gamma \sigma_u^2} \right) > \left(\frac{\mu + \frac{\sigma_u^2}{2}}{\gamma \sigma_u^2} \right),$$

where the rightmost ratio represents the fraction the investor places in the risky portfolio were there to be no labor income at all. If $\sigma_{\ell u} = 0$, then at least one of the following is true: $\text{corr}(u, \ell) = 0$ or $\sigma_\ell = 0$, and each leads to a slightly different interpretation of the optimal a_t . First, if $\sigma_\ell > 0$ (there is variation in labor income), then the independence of labor and equity income allows for a good deal of overall risk reduction, thereby implying a higher optimal risky asset portfolio weight. If $\sigma_\ell = 0$ – labor income is constant – then human capital wealth is a non-tradeable risk free asset in the investor's overall wealth portfolio. Ceteris paribus, this also allows the investor to rebalance his portfolio in favor of a greater fraction held in risky assets. If, alternatively $\sigma_{\ell u} > 0$ – a situation in which the investor's income is closely tied to the behavior of the stock market

– then the investor should correspondingly reduce his position in risky equities. In fact, if the investor’s coefficient of relative risk aversion is sufficiently high and $\sigma_{\ell u}$ large and positive (say, if the investor’s portfolio contained a large position in his own firm’s stock) then $a_t < 0$; i.e., the investor should hold a short position in the overall equity market.

These remarks formalize, though in a very simple context, the idea that an investor’s wage income stream represents an asset and that its statistical covariance with the equity portion of his portfolio should matter for his overall asset allocation. To the extent that variations in stock returns are offset by variations in the investor’s wage income, stocks are effectively less risky (so also is wage income less risky) and he can comfortably hold more of them. The reader may be suspicious, however, of the one period setting. We remedy this next.

Viceira (2001) extends these observations to a multiperiod infinite horizon setting by adopting a number of special features. There is a representative investor-worker who saves for retirement and who must take account in his portfolio allocation decisions of the expected length of his retirement period. In any period there is a probability π^r that the investor will retire; his probability of remaining employed and continuing to receive labor income is $\pi^e = 1 - \pi^r$, with constant probability period by period. With this structure of uncertainty, the expected number of periods until an investor’s retirement period is $\frac{1}{\pi^r}$. Once retired (zero labor income) the period constant probability of death is π^d ; in a like manner the expected length of his retirement is $\frac{1}{\pi^d}$. Viceira (2001) also assumes that labor income is growing in the manner of

$$\Delta \ell_{t+1} = \log L_{t+1} - \log L_t = g + \tilde{\varepsilon}_{t+1}, \quad (14.38)$$

where $g > 0$ and $\tilde{\varepsilon}_{t+1} \sim N(0, \sigma_\varepsilon^2)$. In expression (14.38), g represents the mean growth in labor income (for the U.S. this figure is approximately 2%) while $\tilde{\varepsilon}_t$ denotes random variations about the mean. The return on the risky asset is assumed to follow the same hypothetical process as in the prior example. In this case,

$$\sigma_{u\ell} = \text{cov}_t(r_{t+1}, \Delta \ell_{t+1}) = \text{cov}_t(u_{t+1}, \varepsilon_{t+1}) = \sigma_{u\varepsilon}.$$

With an identical asset structure as in the previous model, the investor’s problem appears deceptively similar to (14.30):

$$\begin{aligned} \max E_t \left(\sum_{i=0}^{\infty} \delta^i \frac{C_{t+i}^{1-\gamma}}{1-\gamma} \right) \\ \text{s.t. } Y_{t+1} = (Y_t + L_t - C_t)R_{P,t+1}. \end{aligned} \quad (14.39)$$

The notation in problem (14.39) is identical to that of the previous model. Depending on whether an agent is employed or retired, however, the first order optimality condition will be different, reflecting the investor’s differing probability structure looking forward. If the investor is retired, for any asset i (the

portfolio P , or the risk free asset):

$$1 = E_t[(1 - \pi^d)\delta \left(\frac{C_{t+1}^r}{C_t^r}\right)^{-\gamma} R_{i,t+1}]. \quad (14.40)$$

The interpretation of equation (14.40) is more or less customary: the investor trades off the marginal utility lost in period t by investing one more consumption unit against the expected utility gain in period $t + 1$ for having done so. The expectation is adjusted by the probability $(1 - \pi^d)$ that the investor is, in fact, still living next period. Analytically, its influence on the optimality condition is the same as a reduction in his subjective discount factor δ . In equations (14.40) and (14.41) to follow, the (not log) consumption is superscripted by e or r , depending upon its enjoyment in the investor's period of employment or retirement, respectively. If the investor is employed, but with positive probability of retirement and subsequent death, then each asset i satisfies:

$$1 = E_t \left\{ \left[\pi^e \delta \left(\frac{C_{t+1}^e}{C_t^e}\right)^{-\gamma} + (1 - \pi^e)(1 - \pi^d) \delta \left(\frac{C_{t+1}^r}{C_t^r}\right)^{-\gamma} \right] (R_{i,t+1}) \right\} \quad (14.41)$$

Equation (14.41)'s interpretation is analogous to (14.40) except that the investor must consider the likelihood of his two possible states next period: either he is employed (probability π^e) or retired and still living (probability $(1 - \pi^e)(1 - \pi^d)$). Whether employed or retired, these equations implicitly characterize the investor's optimal risk free - risky portfolio proportions as those for which his expected utility gain to a marginal dollar invested in either one is the same.

Viceira (2001) log linearizes these equations and their associated budget constraints to obtain the following expressions for log consumption and the optimal risky portfolio weight in both retirement and employment; for a retired investor:

$$c_t^r = b_0^r + b_1^r y_t, \text{ and} \quad (14.42)$$

$$a^r = \frac{\mu + \frac{\sigma_u^2}{2}}{\gamma b_1^r \sigma_u^2}, \quad (14.43)$$

where $b_1^r = 1$ and b_0^r is a complicated (in terms of the model's parameters) constant of no immediate concern; for an employed investor, the corresponding expressions are:

$$c_t^e = b_0^e + b_1^e y_t + (1 - b_1^e) \ell_t, \quad (14.44)$$

$$a^e = \frac{\mu + \frac{\sigma_u^2}{2}}{\gamma \bar{b}_1 \sigma_u^2} - \left(\frac{\pi^e (1 - b_1^e)}{\bar{b}_1} \right) \frac{\sigma_{eu}}{\sigma_u^2} \quad (14.45)$$

with $0 < b_1^e < 1$, $\bar{b}_1 = \pi^e b_1^e + (1 - \pi^e) b_1^r$, and b_0^e , again, a complex constant whose precise form is not relevant for the discussion.

These formulae allow a number of observations:

(1) Since $b_1^r > b_1^e$, (log) consumption is more sensitive to (log) wealth changes for the retired (equation 14.42) as compared with the employed (equation 14.44). This is not surprising as the employed can hedge this risk via his labor income. The retired cannot.

(2) As in the prior model with labor income, there are two terms which together comprise the optimal risky asset proportions for the employed, a^e .

The first $\frac{\mu + \frac{\sigma_u^2}{2}}{\gamma b_1 \sigma_u^2}$ reflects the proportion when labor income is independent of risky returns ($\sigma_{\varepsilon u} = 0$). The second, $-(\frac{\pi^e(1-b_1^e)}{b_1}) \frac{\sigma_{\varepsilon u}}{\sigma_u^2}$, accounts for the hedging component. If $\sigma_{\varepsilon u} < 0$, then the hedge that labor income provides to the risky component of the investor's portfolio is more powerful: the optimal a^e is thus higher, while the opposite is true if $\sigma_{\varepsilon u} > 0$. With a longer expected working life (greater π^e) the optimal hedging component is also higher: the present value of the gains to diversification provided by labor income variation correspondingly increase. Note also that the hedging feature is very strong in the sense that even if the mean equity premium, $\mu = 0$, the investor will retain a positive allocation in risky assets purely for their diversification effect vis-a-vis labor income.

(3) Let us next separate the hedging effect by assuming $\sigma_\varepsilon = 0$ (and thus $\sigma_{\varepsilon u} = 0$). Since $b_1^e < b_1^r$, $\bar{b}_1 < b_1^r$,

$$a^e = \frac{\mu + \frac{\sigma_u^2}{2}}{\gamma \bar{b}_1 \sigma_u^2} > \frac{\mu + \frac{\sigma_u^2}{2}}{\gamma b_1^r \sigma_u^2} = a^r,$$

for any level of risk aversion γ : even if labor income provides no hedging services, the employed investor will hold a greater fraction of his wealth in the risky portfolio than will the retired investor. This is the labor income wealth effect. Ceteris paribus, a riskless labor income stream contributes a valuable riskless asset and its presence allows the investor to tilt the financial component of his wealth in favor of a greater proportion in stocks. It also enhances his average consumption suggesting less aversion to risk. If $\sigma_{\varepsilon u} = 0$ because $\rho_{\varepsilon u} = 0$, $a^e > a^r$ can be justified on the basis of diversification alone. This latter comment is strengthened (weakened) when greater (lesser) diversification is possible: $\sigma_{\varepsilon u} < 0$ ($\sigma_{\varepsilon u} > 0$).

Before summarizing these thoughts, we return to a consideration of the initial three life-cycle portfolio recommendations. Strictly speaking Problem (14.39) is not a life cycle model. Life cycle considerations can be dealt with to a good approximation, however, if we progressively resolve problem (14.39) for a variety of choices of π^e, π^d . If π^d is increased the expected length of the period of retirement falls. If π^r is increased (π^e decreased) it is as if the investor's expected time to retirement was declining as he "aged." Campbell and Viceira (2002) present the results of such an exercise which we report in Table 14.2 for a selection of realistic risk aversion parameters.

Here we find more general theoretical support for at least the 1st and 3rd of our original empirical observations. As investors approach retirement, the fraction of their wealth invested in the risky portfolio does decline strongly. Notice also that it is optimal for mildly risk averse young investors ($\gamma = 2$) to short

Table 14.2: Optimal Percentage Allocation to Stocks^{(i),(ii)}

	Employed				Retired
	Expected Time to Retirement (years)				
	35	25	10	5	
Panel A: $\text{corr}(r_{P,t+1}, \Delta l_{t+1}) = 0$					
$\gamma = 2$	184	156	114	97	80
$\gamma = 5$	62	55	42	37	32
Panel B: $\text{corr}(r_{P,t+1}, \Delta l_{t+1}) = .35$					
$\gamma = 2$	155	136	116	93	80
$\gamma = 5$	42	39	35	33	32

⁽ⁱ⁾ $r_f = .02$, $E r_{P, t+1} - r_{f, t+1} = \mu = .04$, $\sigma_u = .157$, $g = .03$, $\sigma_\varepsilon = .10$

⁽ⁱⁱ⁾ Table 14.2 is a subset of Table 6.1 in Campbell and Viceira (2002)

dramatically the risk free asset to buy more of the risky one in order to “capture” the return supplement inherent in the equity premium. In actual practice, however, such a leverage level is unlikely to be feasible for young investors without a high level of collateral assets. However, the “pull of the premium” is so strong that even retired persons with $\gamma = 5$ (the upper bound for which there is empirical support) will retain roughly one third of their wealth in the risky equity index. In this sense the latter aspect of the first empirical assertion is not borne out, at least for this basic preference specification.

We conclude this section by summarizing the basic points.

1. Riskless labor income creates a tilt in investor portfolios toward risky equities. This is not surprising as the labor income stream in this case contributes a risk free asset in the investor’s portfolio. There are two effects going on. One is a wealth effect: ceteris paribus an investor with a labor income stream is wealthier than an investor without one, and with CRRA utility some of that additional wealth will be assigned to equities. This is complemented by a pure portfolio effect: the risk free asset alters overall portfolio proportions in a way that is manifest as an increased share of financial wealth in risky assets.

2. These same effects are strengthened by the hedging effect if $\sigma_{\varepsilon u} \leq 0$ (effectively this means $\sigma_{r\ell} \leq 0$). Stocks and risky labor income co-vary in a way that each reduces the effective risk of the other. Only if $\sigma_{\varepsilon u}$ is large and positive will the presence of labor income risk reduce the fraction of financial wealth invested in risky assets.

3. Although not discussed explicitly, the ability of an investor to adjust his labor supply - and thus his labor income - only enhances these effects. In this case the investor can elect not only to save more but also to work more if he

experiences an unfavorably risky return realization. His ability to hedge averse risky return realizations is thus enriched, and stocks appear effectively less risky.

14.6 An Important Caveat

The accuracy and usefulness of the notions developed in the preceding sections, especially as regards applications of the formulae for practical portfolio allocations should not be overemphasized. Their usefulness depends in every case on the accuracy of the forecast means, variances, and covariances which represent the inputs to them: garbage in; garbage out still applies! Unfortunately these quantities – especially expected risky returns – have been notoriously difficult to forecast accurately, even one year in advance. Errors in these estimates can have substantial significance for risky portfolio proportions, as these are generally computed using a formula of the generic form

$$\mathbf{a}_t = \frac{1}{\gamma} \mathbf{\Sigma}^{-1} (E_t \mathbf{r}_{t+1} - \mathbf{r}_{f,t+1} \mathbf{1}),$$

where bold face letters represent vectors and $\mathbf{\Sigma}^{-1}$ is a matrix of ‘large’ numbers. Errors in $E_t \mathbf{r}_{t+1}$, the return vector forecasts are magnified accordingly in the portfolio proportion choice.

In a recent paper Uppal (2004) evaluates a number of complex portfolio strategies against a simple equal-portfolio-weights-buy-and-hold strategy. Using the same data set as Campbell and Viceira (2002) use, the equal weighting strategy tends to dominate all the others, simply because, under this strategy, the forecast return errors (which tend to be large) do not affect the portfolio’s makeup.

14.7 Another Background Risk: Real Estate

In this final section we explore the impact of real estate holdings on an investor’s optimal stock-bond allocations. As before, our analysis will be guided by two main principles: (1) all assets – including human capital wealth – should be explicitly considered as components of the investor’s overall wealth portfolio and (2) it is the correlation structure of cash flows from these various income sources that will be paramount for the stock-bond portfolio proportions. Residential real estate is important because it represents roughly half of the U.S. aggregate wealth, and it is not typically included in empirical stand-ins for the U.S. market portfolio M .

Residential real estate also has features that make it distinct from pure financial assets. In particular, it provides a stream of housing services which are inseparable from the house itself. Houses are indivisible assets: one may acquire a small house but not 1/2 of a house. Such indivisibilities effectively place minimum bounds on the amount of real estate that can be acquired. Furthermore, houses cannot be sold without paying a substantial transactions fee, variously estimated to be between 8 % and 10 % of the value of the unit being exchanged. As the purchase of a home is typically a leveraged transaction,

most lenders require minimum “down payments” or equity investments by the purchaser in the house. Lastly, investors may be forced to sell their houses for totally exogenous reasons, such as a job transfer to a new location.

Cocco (2004) studies the stock-bond portfolio allocation problem in the context of a model with the above features yet which is otherwise very similar to the ones considered thus far in this chapter. Recall that our perspective is one of partial equilibrium where, in this section, we seek to understand how the ownership of real estate influences an investor’s other asset holdings, given assumed return processes on the various assets. Below we highlight certain aspects of Cocco’s (2004) modeling of the investor’s problem.

The investor’s objective function, in particular, is

$$\max_{\{S_t, B_t, D_t, F, C_t\}} E \left\{ \sum_{t=0}^T \beta^t \frac{(C_t^\theta H_t^{1-\theta})^{1-\gamma}}{1-\gamma} + \beta^T \frac{(Y_{T+1})^{1-\gamma}}{1-\gamma} \right\}$$

where, as before, C_t is his period t (non-durable) consumption (not logged; in fact no variables will be logged in the problem description), H_t denotes period t housing services (presumed proportional to housing stock with a proportionality constant of one), and Y_t the investor’s period t wealth. Under this formulation non-durable consumption and housing services complement one another with the parameter θ describing the relative preference of one to the other⁸. Investor risk sensitivity to variations in the non durable consumption-housing joint consumption decision is captured by γ (the investor displays CRRA with respect to the composite consumption product). The right most term, $\frac{(Y_{T+1})^{1-\gamma}}{1-\gamma}$, is to be interpreted as a bequest function: the representative investor receives utility from non-consumed terminal wealth which is presumed to be bequeathed to the next generation with the same risk preference structure applying to this quantity as well. In order to capture the idea that houses are indivisible assets, Cocco (2004) imposes a minimum size constraint

$$H_t \geq H_{\min};$$

to capture the fact that there exist transactions costs to changing one’s stock of housing, the agent is assumed to receive only

$$(1 - \lambda)P_t H_{t-1}$$

if he sells his housing stock H_{t-1} , in period t for a price P_t . In his calibration, λ - the magnitude of the transaction cost - is fixed at .08, a level for which there is substantial empirical support in U.S. data. Note the apparent motivation for a bequest motive: given the minimum housing stock constraint, an investor in the final period of his life would otherwise own a significant level of housing stock for which the disposition at his death would be ambiguous⁹.

⁸The idea is simply that an investor will “enjoy his dinner more if he eats it in a warm and spacious house.”

⁹An alternative device for dealing with this modeling feature would be to allow explicitly for reverse mortgages.

Let \tilde{R}_t, R_f and R_D denote the gross random exogenous return on equity, (constant) risk free rate, and the (constant) mortgage interest rate. If the investor elects not to alter his stock of housing in period t relative to $t - 1$, his budget constraint for that period is:

$$S_t + B_t = \tilde{R}_t S_{t-1} + R_f B_t - R_D D_{t-1}^M + L_t - C_t - \chi_t^{FC} F - \Omega P_t H_{t-1} + D_t^M = Y_t \quad (14.46)$$

where the notation is suggestive: S_t and B_t denote his period t stock and bond holdings, D_t^M the level of period t mortgage debt, Ω is a parameter measuring the maintenance cost of home ownership and F is a fixed cost of participating in the financial markets. The indicator function χ_t^{FC} assumes values $\chi_t^{FC} = 1$, if the investor alters his stock or bond holdings relative to period $t - 1$ and 0 otherwise. This device is meant to capture the cost of participating in the securities markets. In the event the investor wishes to alter his stock of housing in period t , his budget constraint is modified in the to-be-expected way (most of it is unchanged except for the addition of the costs of trading houses):

$$S_t + B_t = Y_t + (1 - \lambda)P_t H_{t-1} - P_t H_t, \text{ and} \quad (14.47)$$

$$D_t \leq (1 - d)P_t H_t \quad (14.48)$$

The additional terms in equation (14.47) relative to (14.46) are simply the net proceeds from the sale of the ‘old’ house, $(1 - \lambda)P_t H_{t-1}$, less the costs of the ‘new’ one $P_t H_t$. Constraint (14.48) reflects the down payment equity requirement and the consequent limits to mortgage debt (in his simulation Cocco (2004) chooses $d = .15$).

Cocco (2004) numerically solves the above problem given various assumptions on the return and house price processes which are calibrated to historical data. In particular, he allows for house prices and aggregate labor income shocks to be perfectly positively correlated, and for labor income to have both random and determinate components.¹⁰

¹⁰In particular, Cocco (2004) assumes

$$\tilde{L}_t = \begin{cases} f(t) + \tilde{u}_t, & t \leq T \\ f(t) & t > T \end{cases},$$

where T is the retirement date and the deterministic component $f(t)$ is chosen to replicate the hump shape earnings pattern typically observed. The random component \tilde{u}_t has aggregate ($\tilde{\eta}_t$) and idiosyncratic components ($\tilde{\omega}_t$) where

$$\tilde{u}_t = \tilde{\eta}_t + \tilde{\omega}_t, \text{ and} \quad (14.49)$$

$$\tilde{\eta}_t = \kappa_\eta \tilde{P}_t \quad (14.50)$$

where P_t is the log of the average house price. In addition he assumes $R_f = 1.02$ is fixed for the period $[0, T]$ as is the mortgage rate $R_D = 1.04$. The return on equity follows

$$r_t = \log(\tilde{R}_t) = E \log \tilde{R} + \tilde{\iota}_t$$

with $\tilde{\iota}_t \sim N(0; \sigma_\iota^2)$, $\sigma_\iota = .1674$ and $E \log \tilde{R} = .10$.

Cocco (2004) uses his model to comment upon a number of outstanding financial puzzles of which we will review three: (1) considering the magnitude of the equity premium and the mean reversion in equity returns, why do all investors not hold at least some of their wealth as a well-diversified equity portfolio? Simulations of the model reveal that the minimum housing level H_{\min} (which is calculated at 20,000 US\$) in conjunction with the down payment requirement make it non-optimal for lower labor income investors to pay the fixed costs of entering the equity markets. This is particularly the case for younger investors who remain liquidity constrained. (2) While the material in Section 14.5 suggests that the investors' portfolio share invested in stocks should decrease in later life (as the value of labor income wealth declines), the empirical literature finds that for most investors the portfolio share invested in stocks is increasing over their life cycle. Cocco's (2004) model implies the share in equity investments increasing over the life cycle. As noted above, early in life, housing investments keep investors' liquid assets low and they choose not to participate in the markets. More surprisingly, he notes that the presence of housing can prevent a decline in the share invested in stocks as investors age: as housing wealth increases, investors are more willing to accept equity risk as that risk is not highly correlated with this component. Lastly (3), Cocco deals with the cross sectional observation that the extent of leveraged mortgage debt is highly positively correlated with equity asset holdings. His model is able to replicate this phenomenon as well because of the consumption dimension of housing: investors with more human capital acquire more expensive houses and thus borrow more. Simultaneously, the relatively less risky human capital component induces a further tilt towards stock in high labor income investor portfolios.

14.8 Conclusions

The analysis in this chapter has brought us conceptually to the state of the art in modern portfolio theory. It is distinguished by (1) the comprehensive array of asset classes that must be explicitly considered in order properly to understand an investor's *financial* asset allocations. Labor income (human capital wealth) and residential real estate are two principal cases in point. And to some extent these two asset classes provide conflicting influences on an investor's stock-bond allocations. On the one hand, as relatively riskless human capital diminishes as an investor ages, then *ceteris paribus*, his financial wealth allocation to stocks should fall. On the other hand, if his personal residence has dramatically increased in value over the investor's working years, this fact argues for increased equity holdings given the low correlation between equity and real estate returns. Which effectively dominates is unclear. (2) Second, long run portfolio analysis is distinguished by its consideration of security return paths beyond the standard one-period-ahead mean, variance and covariance characterization. Mean reversion in stock returns suggests intertemporal hedging opportunities as does the long run variation in the risk free rate.

References

- Campbell, J., Cochrane, J. (1999), "By Force of Habit: A Consumption Based Explanation of Aggregate Stock Market Behavior", *Journal of Political Economy*, 107, 205-251.
- Campbell J., Lo, A., MacKinlay, C. (1997), *The Econometrics of Financial Markets*, Princeton University Press.
- Campbell, J., Viceira, L. (1999), "Consumption and Portfolio Decisions when Expected Returns are Time Varying," *Quarterly Journal of Economics*, 114, 433-495.
- Campbell, J., Viceira, L. (2001), Appendix to *Strategic asset Allocation*, <http://kuznets.fas.harvard.edu/Campbell/papers.html>.
- Campbell, J., Viceira, L. (2002), *Strategic Asset Allocation*, Oxford University Press: New York.
- (2001), "Who Should Buy Long-Term Bonds?," *American Economic Review*, 91, 99-127.
- Cocco, D. (2005), "Portfolio Choice in the Presence of Housing," forthcoming, *Review of Financial Studies*.
- Jagannathan, R., Kocherlakota, N.R. (1996), "Why Should Older People Invest Less in Stocks than Younger People?" *Federal Reserve Bank of Minneapolis Quarterly Review*, Summer, 11-23
- Merton, R.C. (1971), "Optimum Consumption and Portfolio Rules in a Continuous Time Model," *Journal of Economic Theory*, 3, 373-413.
- Samuelson, P.A. (1969), "Lifetime Portfolio Selection by Dynamic Stochastic Programming," *Review of Economics and Statistics*, 51, 239-246.
- Santos, T., Veronesi, P. (2004), "Labor Income and Predictable Stock Returns", mimeo, Columbia University.
- Siegel, J. (1998), *Stocks for the Long Run*, McGraw Hill: New York.
- Uppal, R. (2004), "How Inefficient are Simple Asset Allocation Strategies," mimeo, London Business School.
- Viceira, L. (2001), "Optimal Portfolio Choice for Long-Term Investors with Nontradable Labor Income," *Journal of Finance*, 56, 433-470.

Chapter 15 : Financial Structure and Firm Valuation in Incomplete Markets

15.1 Introduction

We have so far motivated the creation of financial markets by the fundamental need of individuals to transfer income across states of nature and across time periods. In Chapter 8 (Section 8.5), we initiated a discussion of the possibility of market failure in financial innovation. There we raised the possibility that coordination problems in the sharing of the benefits and the costs of setting up a new market could result in the failure of a Pareto-improving market to materialize. In reality, however, the bulk of traded securities are issued by firms with the view of raising capital for investment purposes rather than by private individuals. It is thus legitimate to explore the incentives for security issuance taking the viewpoint of the corporate sector. This is what we do in this chapter. Doing so involves touching upon a set of fairly wide and not fully understood topics. One of them is the issue of *security design*. This term refers to the various forms financial contracts can take (and to their properties), in particular, in the context of managing the relationship between a firm and its managers on the one hand, and financiers and owners on the other. We will not touch on these incentive issues here but will first focus on the following two questions.

What Securities Should a Firm Issue If the Value of the Firm Is to be Maximized?

This question is, of course, central to standard financial theory and is usually resolved under the heading *Modigliani-Miller (MM) Theorem* (1958). The *MM* Theorem tells us that under a set of appropriate conditions, if markets are complete the financial decisions of the firm are irrelevant (recall our discussion in Chapter 2). Absent any tax considerations in particular, whether the firm is financed by debt or equity has no impact on its valuation. Here we go one step further and rephrase the question in a context where markets are incomplete and a firm's financing decision modifies the set of available securities. In such a world, the financing decisions of the firm are important for individuals as they may affect the possibilities offered to them for transferring income across states. In this context is it still the case that the firm's financing decisions are irrelevant for its valuation? If not, can we be sure that the interests of the firm's owners as regards the firm's financing decisions coincide with the interests of society at large?

In a second step, we cast the same *security design* issue in the context of inter-temporal investment that can be loosely connected with the finance and growth issues touched upon in Chapter 1. Specifically, we raise the following complementary question.

What Securities Should a Firm Issue If It Is to Grow As Rapidly

As Possible?

We first discuss the connection between the supply of savings and the financial market structure and then consider the problem of a firm wishing to raise capital from the market. The questions raised are important: Is the financial structure relevant for a firm's ability to obtain funds to finance its investments? If so, are the interests of the firm aligned with those of society?

15.2 Financial Structure and Firm Valuation

Our discussion will be phrased in the context of the following simple example. We assume the existence of a unique firm owned by an entrepreneur who wishes only to consume at date $t = 0$; for this entrepreneur, $U'(c_0) > 0$. The assumption of a single entrepreneur circumvents the problem of *shareholder unanimity*: If markets are incomplete, the firm's objective does not need to be the maximization of market value: shareholders cannot reallocate income across all dates and states as they may wish. By definition, there are missing markets. But then shareholders may well have differing preferred payment patterns by the firm – over time and across states – depending on the specificities of their own endowments. One shareholder, for example, may prefer investment project A because it implies the firm will flourish and pay high dividends in future circumstances where he himself would otherwise have a low income. Another shareholder would prefer the firm to undertake some other investment project or to pay higher current dividends because her personal circumstances are different. Furthermore, there may be no markets where the two shareholders could insure one another.

The firm's financial structure consists of a finite set of claims against the firm's period 1 output. These securities are assumed to exhaust the returns to the firm in each state of nature. Since the entrepreneur wishes to consume only in period 0, yet his firm creates consumption goods only in period 1, he will want to sell claims against period 1 output in exchange for consumption in period 0.

The other agents in our economy are agents 1 and 2 of the standard Arrow-Debreu setting of Chapter 8 and we retain the same general assumptions:

1. There are two dates: 0, 1.
2. At date 1, N possible states of nature, indexed $\theta = 1, 2, \dots, N$, with probabilities π_θ , may be realized. In fact, for nearly all that we wish to illustrate $N = 2$ is sufficient.
3. There is one consumption good.
4. Besides the entrepreneur, there are 2 consumers, indexed $k = 1, 2$, with preferences given by

$$U_0^k(c_0^k) + \delta^k \sum_{\theta=1}^N \pi_\theta U^k(c_\theta^k) = \alpha c_0^k + E \ln c_\theta^k$$

and endowments $e_0^k, (e_\theta^k)_{\theta=1,2,\dots,N}$. We interpret c_θ^k to be the consumption of

agent k if state θ should occur, and c_0^k his period zero consumption. Agents' period utility functions are all assumed to be concave, α is the constant date 0 marginal utility, which, for the moment, we will specify to be 0.1, and the discount factor is unity (there is no time discounting). The endowment matrix for the two agents is assumed to be as shown in Table 15.1.

Table 15.1: Endowment Matrix

	Date $t = 0$	Date $t = 1$	
		State $\theta = 1$	State $\theta = 2$
Agent $k = 1$	4	1	5
Agent $k = 2$	4	5	1

Each state has probability 1/2 (equally likely) and consumption in period 0 cannot be stored and carried over into period 1. Keeping matters as simple as possible, let us further assume the cash flows to the firm are the same in each state of nature, as seen in Table 15.2.

Table 15.2: Cash Flows at Date $t = 1$

	$\theta = 1$	$\theta = 2$
Firm	2	2

There are at least two different financial structures that could be written against this output vector:

$F_1 = \{(2, 2)\}$ – pure equity;¹

$F_2 = \{(2, 0), (0, 2)\}$ – Arrow-Debreu securities.²

From our discussion in Chapter 8, we expect financial structure F_2 to be more desirable to agents 1 and 2, as it better allows them to effect income (consumption) stabilization: F_2 amounts to a complete market structure with the two required Arrow-Debreu securities. Let us compute the value of the firm (what the claims to its output could be sold for) under both financial structures. Note that the existence of either set of securities affords an opportunity to shift consumption between periods. This situation is fundamentally different, in this way, from the pure reallocation examples in the pure exchange economies of Chapter 8.

15.2.1 Financial Structure F_1

Let p denote the price (in terms of date 0 consumption) of equity – security $\{(2, 2)\}$ – and let z_1, z_2 respectively, be the quantities demanded by agents 1

¹Equity is risk-free here. This is the somewhat unfortunate consequence of our symmetry assumption (same output in the two date $t = 1$ states). The reader may want to check that our message carries over with a state $\theta = 2$ output of 3.

²Of course, we could have assumed, equivalently, that the firm issues 2 units of the two conceivable *pure* Arrow-Debreu securities, $\{(1, 0), (0, 1)\}$.

and 2. In equilibrium, $z_1 + z_2 = 1$ since there is one unit of equity issued; holding z units of equity entitles the owner to a dividend of $2z$ both in state 1 and in state 2.

Agent 1 solves:

$$\max_{pz_1 \leq 4} (.1)(4 - pz_1) + 1/2[\ln(1 + 2z_1) + \ln(5 + 2z_1)].$$

Agent 2 solves:

$$\max_{pz_2 \leq 4} (.1)(4 - pz_2) + 1/2[\ln(5 + 2z_2) + \ln(1 + 2z_2)].$$

Assuming an interior solution, the FOCs for agents 1 and 2 are, respectively,

$$\begin{aligned} z_1 : \quad & \left(\frac{1}{10}\right)p = \frac{1}{2} \left[\frac{2}{1 + 2z_1} \right] + \frac{1}{2} \left[\frac{2}{5 + 2z_1} \right] \\ & \frac{p}{10} = \left[\frac{1}{1 + 2z_1} + \frac{1}{5 + 2z_1} \right] \\ z_2 : \quad & \frac{p}{10} = \left[\frac{1}{5 + 2z_2} + \frac{1}{1 + 2z_2} \right] \end{aligned}$$

Clearly $z_1 = z_2 = 1/2$, and $\frac{p}{10} = \left[\frac{1}{1+1} + \frac{1}{5+1} \right] = \left[\frac{1}{2} + \frac{1}{6} \right] = \frac{2}{3}$ or $p = 20/3$. Thus, $V_{F1} = p = 20/3 = 6\frac{2}{3}$, and the resulting equilibrium allocation is displayed in Table 15.3.

Table 15.3: Equilibrium Allocation

	$t = 0$	$t = 1$	
		θ_1	θ_2
Agent 1:	$4 - 3\frac{1}{3}$	$1+1$	$5+1$
Agent 2:	$4 - 3\frac{1}{3}$	$5+1$	$1+1$

Agents are thus willing to pay a large proportion of their period 1 consumption in order to increase period 2 consumption. On balance, agents (except the entrepreneur) wish to shift income from the present (where $MU = \alpha = 0.1$) to the future and now there is a device by which they may do so.

Since markets are incomplete in this example, the competitive equilibrium need not be Pareto optimal. That is the case here. There is no way to equate the ratios of the two agents' marginal utilities across the 2 states: In state 1, the MU ratio is $\frac{1/2}{1/6} = 3$ while it is $\frac{1/6}{1/2} = \frac{1}{3}$ in state 2. A transfer of one unit of consumption from agent 2 to agent 1 in state 1 in exchange for one unit of consumption in the other direction in state 2 would obviously be Pareto

improving. Such a transfer cannot, however, be effected with the limited set of financial instruments available. This is the reality of incomplete markets.

Note that our economy is one of three agents: agents 1 and 2, and the original firm owner. From another perspective, the equilibrium allocation under F_1 is not a Pareto optimum because a redistribution of wealth between agents 1 and 2 could be effected making them both better off in ex- ante expected utility terms while not reducing the utility of the firm owner (which is, presumably, directly proportional to the price he receives for the firm). In particular the allocation that dominates the one achieved under F_1 is shown in Table 15.4.

Table 15.4: A Pareto-superior Allocation

	$t = 0$	$t = 1$	
		θ_1	θ_2
Agent 1	2/3	4	4
Agent 2	2/3	4	4
Owner	$6\frac{2}{3}$	0	0

15.2.2 Financial Structure F_2

This is a complete Arrow-Debreu financial structure. It will be notationally clearer here if we deviate from our usual notation and denote the securities as $X = (2, 0)$, $W = (0, 2)$ with prices q_X , q_W respectively (q_X thus corresponds to the price of 2 units of the state-1 Arrow-Debreu security while q_W is the price of 2 units of the state-2 Arrow-Debreu security), and quantities $z_X^1, z_X^2, z_W^1, z_W^2$. The problems confronting the agents are as follows.

Agent 1 solves:

$$\begin{aligned} \max(1/10)(4 - q_X z_X^1 - q_W z_W^1) + [1/2 \ln(1 + 2z_X^1) + 1/2 \ln(5 + 2z_W^1)] \\ q_X z_X^1 + q_W z_W^1 \leq 4 \end{aligned}$$

Agent 2 solves:

$$\begin{aligned} \max(1/10)(4 - q_X z_X^2 - q_W z_W^2) + [1/2 \ln(5 + 2z_X^2) + 1/2 \ln(5 + 2z_W^2)] \\ q_X z_X^2 + q_W z_W^2 \leq 4 \end{aligned}$$

The FOCs are:

$$\begin{aligned} \text{Agent 1: } \begin{cases} (i) \frac{1}{10} q_X = \frac{1}{2} \left(\frac{1}{1+2z_X^1} \right) 2 \\ (ii) \frac{1}{10} q_W = \frac{1}{2} \left(\frac{1}{5+2z_W^1} \right) 2 \end{cases} \\ \text{Agent 2: } \begin{cases} (iii) \frac{1}{10} q_X = \frac{1}{2} \left(\frac{1}{1+2z_X^2} \right) 2 \\ (iv) \frac{1}{10} q_W = \frac{1}{2} \left(\frac{1}{5+2z_W^2} \right) 2. \end{cases} \end{aligned}$$

By equation (i): $\frac{1}{10}q_X = \frac{1}{1+2z_X^1} \Rightarrow 1 + 2z_X^1 = \frac{10}{q_X} \Rightarrow z_X^1 = \frac{5}{q_X} - \frac{1}{2}$.

By equation (iii): $\frac{1}{10}q_X = \frac{1}{5+2z_X^2} \Rightarrow 5 + 2z_X^2 = \frac{10}{q_X} \Rightarrow z_X^2 = \frac{5}{q_X} - \frac{5}{2}$.

With one security of each type issued:

$$\begin{aligned} z_X^1 + z_X^2 &= 1 \quad (z_X^1 \geq 0; z_X^2 \geq 0) \\ \frac{5}{q_X} - \frac{1}{2} + \frac{5}{q_X} - \frac{5}{2} &= 1 \Rightarrow \frac{10}{q_X} = 4 \Rightarrow q_X = \frac{10}{4}. \end{aligned}$$

Similarly, $q_W = \frac{10}{4}$ (by symmetry) and $V_F = q_X + q_W = \frac{10}{4} + \frac{10}{4} = \frac{20}{4} = 5$.

So we see that V_F has declined from $6\frac{2}{3}$ in the F_1 case to 5. Let us further examine this result. Consider the allocations implied by the complete financial structure:

$$\begin{aligned} z_X^1 &= \frac{5}{q_X} - \frac{1}{2} = \frac{5}{5/2} - \frac{1}{2} = 2 - \frac{1}{2} = 1\frac{1}{2} \\ z_X^2 &= \frac{5}{q_X} - \frac{5}{2} = \frac{5}{5/2} - \frac{5}{2} = 2 - \frac{5}{2} = -\frac{1}{2} \\ z_W^1 &= -\frac{1}{2}, \quad z_W^2 = 1\frac{1}{2} \text{ by symmetry} \end{aligned}$$

Thus, agent 1 wants to short sell security 2 while agent 2 wants to short sell security 1. Of course, in the case of financial structure $F_1(2, 2)$, there was no possibility of short selling since every agent, in equilibrium must have the same security holdings. The post-trade allocation is found in Table 15.5.

Table 15.5: Post-Trade Allocation

$t = 0$

Agent 1: $4 - (1\frac{1}{2})q_x + \frac{1}{2}q_w = 4 - \frac{3}{2}(\frac{10}{4}) + \frac{1}{2}(\frac{10}{4}) = 4 - \frac{10}{4} = 1\frac{1}{2}$

Agent 2: $4 + \frac{1}{2}q_x - \frac{3}{2}q_w = 4 + \frac{1}{2}(\frac{10}{4}) - \frac{3}{2}(\frac{10}{4}) = 4 - \frac{10}{4} = 1\frac{1}{2}$

$t = 1$

Agent 1: $(1, 5) + 1\frac{1}{2}(2, 0) - \frac{1}{2}(0, 2) = (4, 4)$

Agent 2: $(5, 1) + (-\frac{1}{2})(2, 0) + 1\frac{1}{2}(0, 2) = (4, 4)$,

This, unsurprisingly, constitutes a Pareto optimum.³

We have thus reached an important result that we summarize in Propositions 15.1 and 15.2.

Proposition 15.1:

When markets are incomplete, the Modigliani-Miller theorem fails to hold and the financial structure of the firm may affect its valuation by the market.

³Note that our example also illustrates the fact that the addition of new securities in a financial market does not necessarily improve the welfare of *all* participants. Indeed, the firm owner is made worse off by the transition from F_1 to F_2 .

Proposition 15.2:

When markets are incomplete, it may not be in the interest of a value-maximizing manager to issue the socially optimal set of securities.

In our example the issuing of the *right* set of securities by the firm leads to completing the market and making a Pareto-optimal allocation attainable. The impact of the financial decision of the firm on the set of markets available to individuals in the economy places us outside the realm of the *MM* theorem and, indeed, the value of the firm is not left unaffected by the choice of financing. Moreover, it appears that it is not, in this situation, in the private interest of the firm's owner to issue the socially optimal set of securities. Our example thus suggests that there is no reason to necessarily expect that value-maximizing firms will issue the set of securities society would find preferable.⁴

15.3 Arrow-Debreu and Modigliani-Miller

In order to understand why V_F declines when the firm issues the richer set of securities, it is useful to draw on our work on Arrow-Debreu pricing (Chapter 8). Think of the economy under financial structure F_2 . This is a complete Arrow-Debreu structure in which we can use the information on equilibrium endowments to recompute the pure Arrow-Debreu prices as per Equation (15.1),

$$q_\theta = -\frac{\delta \pi_\theta \frac{\partial U^k}{\partial c_\theta^k}}{\frac{\partial U_0^k}{\partial c_0^k}}, \theta = 1, 2, \quad (15.1)$$

which, in our example, given the equilibrium allocation (4 units of commodity in each state for both agent) reduces to

$$q_\theta = \frac{1(\frac{1}{2})(\frac{1}{4})}{.1} = \frac{5}{4}, \theta = 1, 2,$$

which corresponds, of course, to

$$q_X = q_W = \frac{10}{4},$$

and to $V_F = 5$.

This Arrow-Debreu complete markets equilibrium is unique: This is generically the case in an economy such as ours, implying there are no other allocations satisfying the required conditions and no other possible prices for the Arrow-Debreu securities. This implies the Modigliani-Miller proposition as the following reasoning illustrates. In our example, the firm is a mechanism to produce 2 units of output in date 1, both in state 1 and in state 2. Given that the

⁴The reader may object that our example is just that, an example. Because it helps us reach results of a negative nature, this example is, however, a fully general *counterexample*, ruling out the proposition that the *MM* theorem continues to hold and that firms' financial structure decisions will always converge with the social interest.

date 0 price of one unit of the good in state 1 at date 1 is $5/4$ and the price of one unit of the good in state 2 at date 1 is $5/4$ as well, it must of necessity be that the price (value) of the firm is 4 times $5/4$, that is, 5. In other words, absent any romantic love for this firm, no one will pay more than 5 units of the current consumption good (which is the numeraire) for the title of ownership to this production mechanism knowing that the same bundle of goods can be obtained for 5 units of the numeraire by purchasing 2 units of each Arrow-Debreu security. A converse reasoning guarantees that the firm will not sell for less either. The value of the firm is thus given by its fundamentals and is independent of the specific set of securities the entrepreneur choses to issue: This is the essence of the Modigliani-Miller theorem!

Now let us try to understand how this reasoning is affected when markets are incomplete and why, in particular, the value of the firm is higher in that context. The intuition is as follows. In the incomplete market environment of financial structure F_1 , security $\{(2, 2)\}$ is desirable for two reasons: to transfer income across time *and* to reduce date 1 consumption risk. In this terminology, the firm in the incomplete market environment is more than a mechanism to produce 2 units of output in either states of nature in date 1. The security issued by the entrepreneur is also the only available vehicle to reduce second period consumption risk. Individual consumers are willing to pay something, that is to sacrifice current consumption, to achieve such risk reduction. To see that trading of security $\{(2, 2)\}$ provides some risk reduction in the former environment, we need only compare the range of date 1 utilities across states after trade and before trade for agent 1 (agent 2 is symmetric). See Table 15.6.

Table 15.6: Agent 1 State Utilities Under F_1

	Before Trade	$\{(2, 2)\}; z^1 = 0.5$ (Equilibrium Allocation)
State 1	$U^1(c_1^1) = \ln 1 = 0$	$U^1(c_1^1) = \ln 2 = 0.693$
State 2	$U^1(c_2^1) = \ln 5 = 1.609$	$U^1(c_2^1) = \ln 6 = 1.792$
	Difference = 1.609	Difference = 1.099

The premium paid for the equity security, over and above the value of the firm in complete markets, thus originates in the dual role it plays as a mechanism for consumption risk smoothing and as a title to two units of output in each future state. A question remains: Given that the entrepreneur, by his activity and security issuance, plays this dual role, why can't he reap the corresponding rewards independently of the security structure he choses to issue? In other words, why is it that his incentives are distorted away from the socially optimal financial structure? To understand this, notice that if any amount of Arrow-Debreu-like securities, such as in $F_2 = \{(2, 0), (0, 2)\}$ is issued, no matter how small, the market for such securities has effectively been created. With no further trading restrictions, the agents can then supply additional amounts of these securities to one another. This has the effect of empowering them to trade, entirely independently of the magnitude of the firm's security issuance, to the

following endowment allocation (see Table 15.7).

Table 15.7: Allocation When the Two Agents Trade Arrow-Debreu Securities Among Themselves

	$t = 0$	$t = 1$	
		θ_1	θ_2
Agent 1:	4	3	3
Agent 2:	4	3	3

In effect, investors can eliminate all *second-period* endowment uncertainty *themselves*. Once this has been accomplished and markets are effectively completed (because there is no further demand for across-state income redistribution, it is irrelevant to the investor whether the firm issues $\{(2, 2)\}$ or $\{(2, 0), (0, 2)\}$, since either package is equally appropriate for transferring income *across time periods*. Were $\{(2, 0), (0, 2)\}$ to be the package of securities issued, the agents would each buy equal amounts of $(2, 0)$, and $(0, 2)$, effectively repackaging them as $(2, 2)$. To do otherwise would be to reintroduce date 1 endowment uncertainty. Thus the relative value of the firm under either financial structure, $\{(2, 2)\}$ or $\{(2, 0), (0, 2)\}$, is determined solely by whether the security $(2, 2)$ is worth more to the investors in the environment of period two endowment uncertainty or when all risk has been eliminated as in the environment noted previously.

Said otherwise, once the markets have been completed, the value of the firm is fixed at 5 as we have seen before, and there is nothing the entrepreneur can do to appropriate the extra insurance premium. If investors can eliminate all the risk themselves (via short selling) there is no premium to be paid to the firm, in terms of value enhancement, for doing so. This is confirmed if we examine the value of the firm when security $\{(2, 2)\}$ is issued *after* the agents have traded among themselves to equal second-period allocation $(3, 3)$. In this case $V_F = 5$ also.

There is another lesson to be gleaned from this example and that leads us back to the CAPM. One of the implications of the CAPM was that securities could not be priced in isolation: Their prices and rates of return depended on their interactions with other securities as measured by the covariance. This example follows in that tradition by confirming that the value of the securities issued by the firm is not independent of the other securities available on the market or which the investors can themselves create.

15.4 On the Role of Short Selling

From another perspective (as noted in Allen and Gale, 1994), short selling expands the supply of securities and provides additional opportunities for risk sharing, but in such a way that the benefits are not internalized by the innovating firm. When deciding what securities to issue, however, the firm only takes into account the impact of the security issuance on its own value; in other words, it

only considers those benefits it can internalize. Thus, in an incomplete market setting, the firm may not issue the socially optimal package of securities.

It is interesting to consider the consequence of forbidding or making it impossible for investors to increase the supply of securities $(2, 0)$ and $(0, 2)$ via short selling. Accordingly, let us impose a no-short-selling condition (by requiring that all holdings of all securities by all agents are positive). Agent 1 wants to short sell $(0, 2)$; agent 2 wants to short sell $(2, 0)$. So, we know that the constrained optimum will have (simply setting $z = 0$ wherever the unconstrained optimum had a negative z and anticipating the market clearing condition):

$$\begin{aligned} z_X^2 &= 0 & z_W^1 &= 0 \\ z_X^1 &= 1 & z_W^2 &= 1 \\ \frac{1}{10}q_X &= MU_1 = \frac{1}{2} \left(\frac{1}{1+2(1)} \right) 2 = \frac{1}{3} \\ \frac{1}{10}q_W &= MU_2 = \frac{1}{2} \left(\frac{1}{1+2(1)} \right) 2 = \frac{1}{3} \\ q_X &= \frac{10}{3}, \quad q_W = \frac{10}{3} \\ V_F &= \frac{20}{3} = 6\frac{2}{3}, \end{aligned}$$

which is as it was when the security $(2, 2)$ was issued.

The fact that V_F rises when short sales are prohibited is not surprising as it reduces the supply of securities $(2, 0)$ and $(0, 2)$. With demand unchanged, both q_X and q_W increase, and with it, V_F . In some sense, now the firm has a monopoly in the issuance of $(2, 0)$ and $(0, 2)$, and that monopoly position has value. All this is in keeping with the general reasoning developed previously. While it is, therefore, not surprising that the value of the firm has risen with the imposition of the short sales constraint, the fact that its value has returned precisely to what it was when it issued $\{(2, 2)\}$ is striking and possibly somewhat of a coincidence.

Is the ruling out of short selling realistic? In practice, short selling on the U.S. stock exchanges is costly, and only a very limited amount of it occurs. The reason for this is that the short seller must deposit as collateral with the lending institution, as much as 100 percent of the value of the securities he borrows to short sell. Under current practice in the United States, the interest on this deposit is less than the T-bill rate even for the largest participants, and for small investors it is near zero. There are other exchange-imposed restrictions on short selling. On the NYSE, for example, investors are forbidden to short sell on a down-tick in the stock's price.⁵

⁵Brokers must obtain permission from clients to borrow their shares and relend them to a short seller. In the early part of 2000, a number of high technology firms in the United States asked their shareholders to deny such permission as it was argued short sellers were depressing prices! Of course if a stock's price begins rising, short sellers may have to enter the market to buy shares to cover their short position. This boosts the share price even further.

15.5 Financing and Growth

Now we must consider our second set of issues, which we may somewhat more generally characterize as follows: How does the degree of completeness in the securities markets affect the level of capital accumulation? This is a large topic, touched upon in our introductory chapter, for which there is little existing theory. Once again we pursue our discussion in the context of examples.

Example 15.1

Our first example serves to illustrate the fact that while a more complete set of markets is unambiguously good for welfare, it is not necessarily so for growth. Consider the following setup. Agents own firms (have access to a productive technology) while also being able to trade state-contingent claims with one another (net supply is zero). We retain the two-agent, two-period setting. Agents have state-contingent consumption endowments in the second period. They also have access to a productive technology which, for every k units of period one consumption foregone, produces \sqrt{k} in date 1 in either state of nature⁶ (see Table 15.8).

Table 15.8: The Return From Investing k Units

$t = 1$	
θ_1	θ_2
\sqrt{k}	\sqrt{k}

The agent endowments are given in Table 15.9.

Table 15.9: Agent Endowments

	$t = 1$	$t = 2$	
		θ_1	θ_2
Agent 1:	3	5	1
Agent 2:	3	1	5

Prob(θ_1)=Prob(θ_2) = 1/2

and the agent preference orderings are now (identically) given by

$$EU(c_0, c_\theta) = \ln(c_0) + \frac{1}{2} \ln(c_1) + \frac{1}{2} \ln(c_2).$$

In this context, we compute the agents' optimal savings levels under two alternative financial structures. In one case, there is a complete set of contingent

⁶Such a technology may not look very interesting at first sight! But, at the margin, agents may be very grateful for the opportunity it provides to smooth consumption across time periods.

claims, in the other, the productive technology is the only possibility for redistributing purchasing power across states (as well as across time) among the two agents.

15.5.1 No Contingent Claims Markets

Each agent acts autonomously and solves:

$$\max_k \ln(3 - k) + \frac{1}{2} \ln(5 + \sqrt{k}) + \frac{1}{2} \ln(1 + \sqrt{k}).$$

Assuming an interior solution, the optimal level of savings k^* solves:

$$-\frac{1}{3 - k^*} + \left\{ \frac{1}{2} \left(\frac{1}{5 + \sqrt{k^*}} \right) \frac{1}{2} (k^*)^{-\frac{1}{2}} + \frac{1}{2} \left(\frac{1}{1 + \sqrt{k^*}} \right) \frac{1}{2} (k^*)^{-\frac{1}{2}} \right\} = 0$$

which, after several simplifications, yields

$$3(k^*)^{\frac{3}{2}} + 15k^* + 7\sqrt{k^*} - 9 = 0.$$

The solution to this equation is $k^* = 0.31$. With two agents in the economy, *economy-wide savings* are 0.62. Let us now compare this result with the case in which the agents also have access to contingent claims markets.

15.5.2 Contingent Claims Trading

Let q_1 be the price of a security that pays 1 unit of consumption if state 1 occurs, and let q_2 be the price of a security that pays 1 unit of consumption if state 2 occurs. Similarly, let $z_1^1, z_2^1, z_1^2, z_2^2$ denote, respectively, the quantities of these securities demanded by agents 1 and 2. These agents continue to have simultaneous access to the technology.

Agent 1 solves:

$$\max_{k_1, z_1^1, z_2^1} \ln(3 - k_1 - q_1 z_1^1 - q_2 z_2^1) + \frac{1}{2} \ln(5 + \sqrt{k_1 + z_1^1}) + \frac{1}{2} \ln(1 + \sqrt{k_1 + z_2^1}).$$

Agent 2's problem is essentially the same:

$$\max_{k_2, z_1^2, z_2^2} \ln(3 - k_2 - q_1 z_1^2 - q_2 z_2^2) + \frac{1}{2} \ln(1 + \sqrt{k_2 + z_1^2}) + \frac{1}{2} \ln(5 + \sqrt{k_2 + z_2^2}).$$

By symmetry, in equilibrium

$$\begin{aligned} k_1 &= k_2; \quad q_1 = q_2; \\ z_1^1 &= z_2^2 = -z_1^2, \quad z_2^1 = z_1^1 = -z_2^2. \end{aligned}$$

Using these facts and the FOCs (see the Appendix), it can be directly shown that

$$-2 = z_1^1$$

and, it then follows that $k_1 = 0.16$. Thus, $total\ savings = k_1 + k_2 = 2k_1 = 0.32$.

Savings have thus been substantially reduced. This result also generalizes to situations of more general preference orderings, and to the case where the uncertainty in the states is in the form of uncertainty in the production technology rather than in the investor endowments. The explanation for this phenomenon is relatively straightforward and it parallels the mechanism at work in the previous sections. With the opening of contingent claims markets, the agents can eliminate all second-period risk. In the absence of such markets, it is real investment that alone must provide for any risk reduction as well as for income transference across time periods – a dual role. In a situation of greater uncertainty, resulting from the absence of contingent claims markets, more is saved and the extra savings take, necessarily, the form of productive capital: There is a precautionary demand for capital. Jappelli and Pagano (1994) find traces of a similar behavior in Italy prior to recent measures of financial deregulation.

Example 15.2

This result also suggests that if firms want to raise capital in order to invest for date 1 output, it may not be value maximizing to issue a more complete set of securities, an intuition we confirm in our second example.

Consider a firm with access to a technology with the output pattern found in Table 15.10.

Table 15.10: The Firm's Technology

$t = 0$	$t = 1$	
	θ_1	θ_2
$-k$	\sqrt{k}	\sqrt{k}

Investor endowments are given in Table 15.11.

Table 15.11: Investor Endowments

	$t = 0$	$t = 1$	
		θ_1	θ_2
Agent 1:	12	1/2	10
Agent 2:	12	10	1/2

Their preference orderings are both of the form:

$$EU(c_1, c_2) = \frac{1}{12}c_0 + \frac{1}{2}\ln(c_1) + \frac{1}{2}\ln(c_2).$$

15.5.3 Incomplete Markets

Suppose a security of the form $(1, 1)$ is traded, at a price p ; agents 1 and 2 demand, respectively, z_1 and z_2 . The agent maximization problems that define their demand are as follows:

Agent 1:

$$\max_{pz_1 \leq 12} (1/12)(12 - pz_1) + \frac{1}{2} \ln(1/2 + z_1) + \frac{1}{2} \ln(10 + z_1) ;$$

Agent 2:

$$\max_{pz_2 \leq 12} (1/12)(12 - pz_2) + \frac{1}{2} \ln(10 + z_2) + \frac{1}{2} \ln(1/2 + z_2) .$$

It is obvious that $z_1 = z_2$ at equilibrium. The first order conditions are (again assuming an interior solution):

$$\begin{aligned} \text{Agent 1: } \frac{p}{12} &= \frac{1}{2} \frac{1}{(1/2 + z_1)} + \frac{1}{2} \frac{1}{(10 + z_1)} \\ \text{Agent 2: } \frac{p}{12} &= \frac{1}{2} \frac{1}{(10 + z_2)} + \frac{1}{2} \frac{1}{(1/2 + z_2)} \end{aligned}$$

In order for the technological constraint to be satisfied, it must also be that

$$\begin{aligned} [p(z_1 + z_2)]^{1/2} &= z_1 + z_2, \text{ or} \\ p &= z_1 + z_2 = 2z_1 \text{ as noted earlier.} \end{aligned}$$

Substituting for p in the first agent's FOC gives:

$$\begin{aligned} \frac{2z_1}{12} &= \frac{1}{2} \frac{1}{(1/2 + z_1)} + \frac{1}{2} \frac{1}{(10 + z_1)}, \text{ or} \\ 0 &= z_1^3 + 10.5z_1^2 - z_1 - 31.5. \end{aligned}$$

Trial and error gives $z_1 = 1.65$. Thus $p = 3.3$ and total investment is

$p = z_1 + z_2 = (3.3)(3.3) = 10.89 = V_F$; date 1 output in each state is thus $\sqrt{10.89} = 3.3$.

15.5.4 Complete Contingent Claims

Now suppose securities $R = (1, 0)$ and $S = (0, 1)$ are traded at prices q_R and q_S and denote quantities demanded respectively as $z_R^1, z_R^2, z_S^1, z_S^2$. The no short sales assumption is retained. With this assumption, agent 1 buys only R while agent 2 buys only security S . Each agent thus prepares himself for his worst possibility.

$$\begin{aligned} \text{Agent 1: } \max_{0 \leq q_R z_R^1} & (1/12)(12 - q_R z_R^1) + \frac{1}{2} \ln(1/2 + z_R^1) + \frac{1}{2} \ln(10) \end{aligned}$$

$$\begin{aligned} \text{Agent 2: } \max_{0 \leq q_S z_S^2} & (1/12)(12 - q_S z_S^2) + \frac{1}{2} \ln(10) + \frac{1}{2} \ln(1/2 + z_S^2) \end{aligned}$$

The FOCs are thus:

$$\begin{aligned}\text{Agent 1 : } \frac{q_R}{12} &= \frac{1}{2} \frac{1}{(1/2 + z_R^1)} \\ \text{Agent 2 : } \frac{q_S}{12} &= \frac{1}{2} \frac{1}{(1/2 + z_S^2)}.\end{aligned}$$

Clearly $q_R = q_S$ by symmetry, and $z_R^1 = z_S^2$; by the technological constraints:

$$\begin{aligned}(q_R z_R^1 + q_S z_S^2)^{1/2} &= \left(\frac{z_R^1 + z_S^2}{2} \right), \text{ or} \\ q_R &= \frac{z_R^1}{2}.\end{aligned}$$

Solving for z_R^1 :

$$\begin{aligned}\frac{q_R}{12} &= \frac{z_R^1}{24} = \frac{1}{2} \frac{1}{(1/2 + z_R^1)} = \frac{1}{1 + 2z_R^1} \\ z_R^1 (1 + 2z_R^1) &= 24 \\ z_R^1 &= \frac{-1 \pm \sqrt{1 - 4(2)(-24)}}{4} = \frac{-1 \pm \sqrt{1 + 192}}{4} = \frac{-1 \pm 13.892}{4}\end{aligned}$$

(taking positive root)

$$\begin{aligned}z_R^1 &= 3.223 \\ z_S^2 &= 3.223 \\ q_R &= 1.61, \text{ and} \\ q_R (z_R^1 + z_S^2) &= 1.61 (6.446) = 10.378 = V_F.\end{aligned}$$

As suspected, this is less than what the firm could raise issuing only (1,1).

Much in the spirit of our discussion of Section 15.2, this example illustrates the fact that, for a firm wishing to maximize the amount of capital levied from the market, it may not be a good strategy to propose contracts leading to a (more) complete set of markets. This is another example of the failure of the Modigliani-Miller theorem in a situation of incomplete markets and the reasoning is the same as before: In incomplete markets, the firm's value is not necessarily equal to the value, computed at Arrow-Debreu prices, of the portfolio of goods it delivers in future date-states. This is because the security it issues may, in addition, be valued by market participants for its unintended role as an insurance mechanism, a role that disappears if markets are complete. In the growth context of our last examples, this may mean that more savings will be forthcoming when markets are incomplete, a fact that may lead a firm wishing to raise capital from the markets to refrain from issuing the optimal set of securities.

15.6 Conclusions

We have reached a number of conclusions in this chapter.

1. In an incomplete market context, it may not be value maximizing for firms to offer the socially optimal (complete) set of securities. This follows from the fact that, in a production setting, securities can be used not only for risk reduction but also to transfer income across dates. The value of a security will depend upon its usefulness in accomplishing these alternative tasks.

2. The value of securities issued by the firm is not independent of the supply of similar securities issued by other market participants. To the extent that others can increase the supply of a security initially issued by the firm (via short selling), its value will be reduced.

3. Finally, welfare is, but growth may not be, promoted by the issuance of a more complete set of markets.⁷ As a result, it may not be in the best interest of a firm aiming at maximizing the amount of capital it wants to raise, to issue the most socially desirable set of securities.

All these results illustrate the fact that if markets are incomplete, the link between private interests and social optimality is considerably weakened. Here lies the intellectual foundation for financial market regulation and supervision.

References

- Allen, F., Gale, D. (1994), *Financial Innovation and Risk Sharing*, MIT Press, Cambridge, Mass.
- Hart, O. (1975), "On the Optimality of Equilibrium When Market Structure is Incomplete," *Journal of Economic Theory*, 11, 418–443.
- Jappelli, T., Pagano, M. (1994), "Savings, Growth and Liquidity Constraints," *Quarterly Journal of Economics*, 109, 83–109.
- Modigliani, F., Miller, M. (1958), "The Cost of Capital, Corporation Finance, and the Theory of Investment," *American Economic Review*, 48, 261–297.

Appendix: Details of the Solution of the Contingent Claims Trade Case of Section 15.5

Agent 1 solves:

$$\max_{k_1, z_1^1, z_2^1} \ln(3 - k_1 - q_1 z_1^1 - q_2 z_2^1) + \frac{1}{2} \ln(5 + \sqrt{k_1 + z_1^1}) + \frac{1}{2} \ln(1 + \sqrt{k_1 + z_2^1})$$

⁷The statement regarding welfare is strictly true only when financial innovation achieves full market completeness. Hart (1975) shows that it is possible that everyone is made worse off when the markets become more complete but not fully complete (say, going from 9 to 10 linearly independent securities when 15 would be needed to make the markets complete).

$$k_1 : \frac{-1}{3 - k_1 - q_1 z_1^1 - q_2 z_2^1} + \frac{1}{2} \left(\frac{1}{5 + \sqrt{k_1} + z_1^1} \right) \frac{1}{2} k_1^{-\frac{1}{2}} + \frac{1}{2} \left(\frac{1}{1 + \sqrt{k_1} + z_2^1} \right) \frac{1}{2} k_1^{-\frac{1}{2}} = 0 \quad (15.2)$$

$$z_1^1 : \frac{-q_1}{3 - k_1 - q_1 z_1^1 - q_2 z_2^1} + \frac{1}{2} \left(\frac{1}{5 + \sqrt{k_1} + z_1^1} \right) = 0 \quad (15.3)$$

$$z_2^1 : \frac{-q_2}{3 - k_1 - q_1 z_1^1 - q_2 z_2^1} + \frac{1}{2} \left(\frac{1}{1 + \sqrt{k_1} + z_2^1} \right) = 0 \quad (15.4)$$

Agent 2's problem and FOC are essentially the same:

$$\max_{k_2, z_1^2, z_2^2} \ln(3 - k_2 - q_1 z_1^2 - q_2 z_2^2) + \frac{1}{2} \ln(1 + \sqrt{k_2} + z_1^2) + \frac{1}{2} \ln(5 + \sqrt{k_2} + z_2^2)$$

$$k_2 : \frac{-1}{3 - k_2 - q_1 z_1^2 - q_2 z_2^2} + \frac{1}{2} \left(\frac{1}{1 + \sqrt{k_2} + z_1^2} \right) \frac{1}{2} k_2^{-\frac{1}{2}} + \frac{1}{2} \left(\frac{1}{5 + \sqrt{k_2} + z_2^2} \right) \frac{1}{2} k_2^{-\frac{1}{2}} = 0 \quad (15.5)$$

$$z_1^2 : \frac{-q_1}{3 - k_2 - q_1 z_1^2 - q_2 z_2^2} + \frac{1}{2} \left(\frac{1}{1 + \sqrt{k_2} + z_1^2} \right) = 0 \quad (15.6)$$

$$z_2^2 : \frac{-q_2}{3 - k_2 - q_1 z_1^2 - q_2 z_2^2} + \frac{1}{2} \left(\frac{1}{5 + \sqrt{k_2} + z_2^2} \right) = 0 \quad (15.7)$$

By symmetry, in equilibrium

$$k_1 = k_2; \quad q_1 = q_2; \\ z_1^1 = z_2^2 = -z_1^2, \quad z_2^1 = z_1^2 = -z_2^2$$

By Equations (15.3) and (15.6), using the fact that $z_1^1 + z_2^1 = z_2^2 + z_1^2$:

$$\frac{1}{5 + \sqrt{k_1} + z_1^1} = \frac{1}{1 + \sqrt{k_2} + z_1^2}$$

Equations (15.4) and (15.7):

$$\frac{1}{1 + \sqrt{k_1} + z_2^1} = \frac{1}{5 + \sqrt{k_2} + z_2^2}$$

The equations defining k_1 and z_1^1 are thus reduced to

$$k_1 : \frac{1}{3 - k_1 - q_1 z_1^1 - q_2 z_2^1} + \frac{1}{4} \frac{1}{\sqrt{k_1}} \left(\frac{1}{5 + \sqrt{k_1} + z_1^1} \right) + \frac{1}{4} \frac{1}{\sqrt{k_1}} \left(\frac{1}{1 + \sqrt{k_1} + z_1^1} \right) = 0 \quad (15.8)$$

$$z_1^1 : \frac{1}{5 + \sqrt{k_1} + z_1^1} = \frac{1}{1 + \sqrt{k_1} + z_1^1} \quad (15.9)$$

Solving for k_1, z_1^1 , yields from Equation (15.8)

$$\begin{aligned} 1 + \sqrt{k_1} - z_1^1 &= 5 + \sqrt{k_1} + z_1^1 \\ -4 &= 2z_1^1 \\ -2 &= z_1^1 \end{aligned}$$

Substituting this value into Equation (15.8) gives

$$\begin{aligned} \frac{1}{3 - k_1} &= +\frac{1}{4} \frac{1}{\sqrt{k_1}} \left\{ \frac{1}{5 + \sqrt{k_1} - 2} + \frac{1}{1 + \sqrt{k_1} + 2} \right\} \\ \frac{1}{3 - k_1} &= +\frac{1}{4} \frac{1}{\sqrt{k_1}} \left\{ \frac{1}{3 + \sqrt{k_1}} \right\} \\ 4\sqrt{k_1} \left\{ 3 + \sqrt{k_1} \right\} &= 2(3 - k_1), \text{ or, simplifying,} \\ -6 + 12\sqrt{k_1} + 6k_1 &= 0 \\ -1 + 2\sqrt{k_1} + k_1 &= 0 \end{aligned}$$

Let $X = \sqrt{k_1}$

$$\begin{aligned} X &= \frac{-2 \pm \sqrt{4 - 4(1)(-1)}}{2} = \frac{-2 \pm \sqrt{8}}{2} \\ X &= -1 + \sqrt{2} = -1 + 1.4 \\ X &= 0.4 \end{aligned}$$

$k_1 = 0.16$ and total savings $= k_1 + k_2 = 2k_1 = 0.32$

Chapter 16: Financial Equilibrium with Differential Information

16.1 Introduction

The fact that investors often disagree about expected future returns or the evaluation of the risks associated with specific investments is probably the foremost determinant of financial trading in the sense of explaining the larger fraction of trading volume. Yet we have said very little so far about the possibility of such disagreements and, more generally, of differences in investors' information. In fact, two of the equilibrium models we have reviewed have explicitly assumed investors have identical information sets. In the case of the CAPM, it is assumed that all investors' expectations are summarized by the same vector of expected returns and the same variance-covariance matrix. It is this assumption that gives relevance to the single efficient frontier. Similarly, the assumption of a single representative decision maker in the CCAPM is akin to assuming the existence of a large number of investors endowed with identical preferences and information sets.¹ The Rational Expectations hypothesis, which is part of the CCAPM, necessarily implies that, at equilibrium, all investors share the same objective views about future returns.

Both the APT and the Martingale pricing models are nonstructural models which, by construction, are agnostic about the background information (or preferences) of the investors. In a sense they thus go beyond the homogenous information assumption, but without being explicit as to the specific implications of such an extension. The Arrow-Debreu model is a structural model equipped to deal, at least implicitly, with heterogeneously informed agents. In particular, it can accommodate general utility representations defined on state-contingent commodities where, in effect, the assumed state probabilities are embedded in the specific form taken by the individual's utility function.² Thus, while agents must agree on the relevant states of the world, they could disagree on their probabilities. We did not exploit this degree of generality, however, and typically made our arguments on the basis of time-additive and state-additive utility functions with explicit, investor-homogenous, state probabilities.

In this chapter we relax the assumption that all agents in the economy have the same subjective probabilities about states of nature or the same expectations about returns, or that they know the objective probability distributions. In so doing we open a huge and fascinating, yet incomplete, chapter in financial economics, part of which was selectively reviewed in Chapter 2. We will again be very selective in the topics we choose to address under this heading and will concentrate on the issue of market equilibrium with differentially informed traders. This is in keeping with the spirit of this book and enables us to revisit the last important pillar of traditional financial theory left untouched thus far: the efficient market hypothesis.

¹Box 9.1 discussed the extent to which this interpretation can be relaxed as far as utility functions are concerned.

²Such preference structures are, strictly speaking, not expected utility.

The import of differential information for understanding financial markets, institutions, and contracts, however, goes much beyond market efficiency. Since Akerlof (1970), asymmetric information – a situation where agents are differentially informed with, moreover, one or a subgroup having *superior* information – is known potentially to lead to the failure of a market to exist. This *lemons* problem is a relevant one in financial markets: One may be suspicious of purchasing a stock from a better informed intermediary, or, a fortiori, from the primary issuer of a security who may be presumed to have the best information about the exact value of the underlying assets. One may suspect that the issuer would be unwilling to sell at a price lower than the fundamental value of the asset. What is called the *winner's curse* is applicable here: if the transaction is concluded, that is, if the better-informed owner has agreed to sell, is it not likely that the buyer will have paid too much for the asset? This reasoning might go some way toward explaining the fact that capital raised by firms in equity markets is such a small proportion of total firm financing [on this, see Greenwald and Stiglitz (1993)].

Asymmetric information may also explain the phenomenon of credit rationing. The idea here is that it may not be to the advantage of a lender, confronted with a demand for funds larger than he can accommodate, to increase the interest rate he charges as would be required to balance supply and demand: doing so the lender may alter the pool of applicants in an unfavorable way. Specifically, this possibility depends on the plausible hypothesis that the lender does not know the degree of riskiness of the projects for which borrowers need funds and that, in the context of a debt contract, a higher hurdle rate may eliminate the less profitable, but consequently, also the less risky, projects. It is easy to construct cases where the creditor is worse off lending his funds at a higher rate because at the high rate the pool of borrowers becomes riskier [Stiglitz and Weiss (1981)].

Asymmetric information has also been used to explain the prevalence of debt contracts relative to contingent claims. We have used the argument before (Chapter 8): States of nature are often costly to ascertain and verify for one of the parties in a contract. When two parties enter into a contract, it may be more efficient, as a result, to stipulate noncontingent payments most of the time, thus economizing on verification costs. Only states leading to bankruptcy or default are recognized as resulting in different rights and obligations for the parties involved [Townsend (1979)].

These are only a few of the important issues that can be addressed with the asymmetric information assumption. A full review would deserve a whole book in itself. One reason for the need to be selective is that there is a lack of a unifying framework in this literature. It has often proceeded with a set of specific examples rather than more encompassing models. We refer interested readers to Hirshleiffer and Riley (1992) for a broader review of this fascinating and important topic in financial economics.

16.2 On the Possibility of an Upward Sloping Demand Curve

There are plenty of reasons to believe that differences in information and beliefs constitute an important motivation for trading in financial markets. It is extremely difficult to rationalize observed trading volumes in a world of homogeneously informed agents. The main reason for having neglected what is without doubt an obvious fact is that our equilibrium concept, borrowed from traditional supply and demand analysis (the standard notion of Walrasian equilibrium), must be thoroughly updated once we allow for heterogeneous information.

The intuition is as follows: The Walrasian equilibrium price is necessarily some function of the orders placed by traders. Suppose traders are heterogeneously informed and that their private information set is a relevant determinant of their orders. The equilibrium price will, therefore, reflect and, in that sense, transmit at least a fraction of the privately held information. In this case, the equilibrium price is not only a signal of relative scarcity, as in a Walrasian world; it also reflects the agents' information. In this context, the price quoted for a commodity or a security may be high because the demand for it is objectively high and/or the supply is low. But it may also be high because a group of investors has private information suggestive that the commodity or security in question will be expensive tomorrow. Of course, this information about the future value of the item is of interest to all. Presumably, except for liquidity reasons, no one will want to sell something at a low price that will likely be of much higher value tomorrow. This means that when the price quoted on the market is high (in the fiction of standard microeconomics, when the *Walrasian auctioneer* announces a high price), a number of market participants will realize that they have sent in their orders on the basis of information that is probably not shared by the *rest of the market*. Depending on the confidence they place in their own information, they may then want to revise their orders, and to do so in a paradoxical way: Because the announced price is higher than they thought it would be, they may want to buy more! Fundamentally, this means that what was thought to be the equilibrium price is not, in fact, an equilibrium.

This is a new situation and it requires a departure from the Walrasian equilibrium concept. In this chapter we will develop these ideas with the help of an example. We first illustrate the notion of a Rational Expectations Equilibrium (REE), a concept we have used more informally in preceding chapters (e.g., Chapter 9), in a context where all participants share the same information. We then extend it to encompass situations where agents are heterogeneously informed. We provide an example of a fully revealing rational expectations equilibrium which may be deemed to be the formal representation of the notion of an *informationally efficient market*. We conclude by discussing some weaknesses of this equilibrium concept and possible extensions.

16.3 An Illustration of the Concept of REE: Homogeneous Information³

Let us consider the joint equilibrium of a spot market for a given commodity and its associated futures market. The context is the familiar now and then, two-date economy. The single commodity is traded at date 1. Viewed from date 0, the date at which producers must make their production decisions, the demand for this commodity, emanating from final users, is stochastic. It can be represented by a linear demand curve shocked by a random term as in

$$D(p, \tilde{\eta}) = a - cp + \tilde{\eta},$$

where $D(\cdot)$ represents the quantity demanded, p is the (spot) price for the commodity in question, a and c are positive constants, and $\tilde{\eta}$ is a stochastic demand-shifting element.⁴ This latter quantity is centered at (has mean value) zero, at which point the demand curve assumes its average position, and it is normally distributed with variance $\sigma_{\tilde{\eta}}^2$, in other words, $h(\tilde{\eta}) = N(0; \sigma_{\tilde{\eta}}^2)$ where $h(\cdot)$ is the probability density function on $\tilde{\eta}$. See Figure 16.1 for an illustration. At date 0, the N producers decide on their input level x – the input price is normalized at 1 – knowing that $g(x)$ units of output will then be available after a one-period production lag at date 1. The production process is thus non-stochastic and the only uncertainty originates from the demand side. Because of the latter feature, the future sale price \tilde{p} is unknown at the time of the input decision.

Insert Figure 16.1 about here

We shall assume the existence of a futures or forward market⁵ that our producers may use for hedging or speculative purposes. Specifically, let $f > 0$ (< 0) be the short (long) futures position taken by the representative producer, that is, the quantity of output sold (bought) for future delivery at the future (or forward) price p^f .

Here we shall assume that the good traded in the futures market, (i.e., specified as acceptable for delivery in the futures contract), is the same as the

³The rest of this chapter closely follows Danthine (1978).

⁴Looking forward, the demand for heating oil next winter is stochastic because the severity of the winter is impossible to predict in advance.

⁵The term *futures market* is normally reserved for a market for future delivery taking place in the context of an organized exchange. A *forward market* refers to private exchanges of similar contracts calling for the future delivery of a commodity or financial instrument. While knowledge of the credit worthiness and honesty of the counter-party is of essence in the case of forward contracts, a futures market is anonymous. The exchange is the relevant counter-party for the two sides in a contract. It protects itself and ensures that both parties' engagements will be fulfilled by demanding initial guarantee deposits as well as issuing daily *margin calls* to the party against whose position the price has moved. In a two-date setting, thus in the absence of interim price changes, the notion of margin calls is not relevant and it is not possible to distinguish futures from forwards.

commodity exchanged on the spot market. For this reason, arbitrageurs will ensure that, at date 1, the futures and the spot price will be exactly identical: In the language of futures markets, the *basis* is constantly equal to zero and there is thus no *basis risk*.

Under these conditions, the typical producer's cash flow y is

$$\tilde{y} = \tilde{p}g(x) - x + (p^f - \tilde{p})f$$

which can also be written as

$$\tilde{y} = \tilde{p}(g(x) - f) - x + p^f f.$$

It is seen that by setting $f = g(x)$, that is, by selling forward the totality of his production, the producers can eliminate all his risks. Although this need not be his optimal futures position, the feasibility of shedding all risks explains the separation result that follows (much in the spirit of the CAPM: diversifiable risk is not priced).

Let us assume that producers maximize the expected utility of their future cash flow where $U'(\cdot) > 0$ and $U''(\cdot) < 0$:

$$\max_{x \geq 0, f} EU(\tilde{y})$$

Differentiating with respect to x and f successively, and assuming an interior solution, we obtain the following two FOCs:

$$x : E[U_1(\tilde{y})\tilde{p}] = \frac{1}{g_1(x)} EU_1(\tilde{y}) \quad (16.1)$$

$$f : E[U_1(\tilde{y})\tilde{p}] = p^f EU_1(\tilde{y}) \quad (16.2)$$

which together imply:

$$p^f = \frac{1}{g_1(x)}. \quad (16.3)$$

Equation (16.3) is remarkable because it says that the optimal input level should be such that the marginal cost of production is set equal to the (known) futures price p^f , the latter replacing the expected spot price as the appropriate production signal. The futures price equals marginal cost condition is also worth noticing because it implies that, despite the uncertain context in which they operate, producers should not factor in a risk premium when computing their optimal production decision.

For us, a key implication of this result is that, since the supply level will directly depend on the futures price quoted at date 0, the equilibrium spot price at date 1 will be a function of the futures price realized one period earlier. Indeed, writing $x = x(p^f)$ and $g(x) = g(x(p^f))$ to highlight the implications of Equation (16.3) for the input and output levels, the supply-equals-demand condition for the date 1 spot market reads

$$Ng(x(p^f)) = a - cp + \tilde{\eta}$$

which implicitly defines the equilibrium (date 1) spot price as a function of the date 0 value taken by the futures price, or

$$\tilde{p} = p(p^f, \tilde{\eta}) . \quad (16.4)$$

It is clear from Equation (16.4) that the structure of our problem is such that the probability distribution on \tilde{p} cannot be spelled out independently of the value taken by p^f . Consequently, it would not be meaningful to assume expectations for \tilde{p} , on the part of producers or futures market speculators, which would not take account of this fundamental link between the two prices. This observation, which is a first step toward the definition of a rational expectation equilibrium, can be further developed by focusing now on the futures market.

Let us assume that, in addition to the N producers, n speculators take positions on the futures market. We define speculators by their exclusive involvement in the futures markets; in particular they have no position in the underlying commodity. Accordingly, their cash flows are simply:

$$\tilde{z}_i = (p^f - \tilde{p}) b_i$$

where b_i is the futures position (> 0 = short ; < 0 = long) taken by speculator i . Suppose for simplicity that their preferences are represented by a linear mean-variance utility function of their cash flows:

$$W(\tilde{z}_i) = E(\tilde{z}_i) - \frac{\chi}{2} var(\tilde{z}_i)$$

where χ represents the (Arrow-Pratt) Absolute Risk Aversion index of the representative speculator. We shall similarly specialize the utility function of producers. The assumption of a linear mean-variance utility representation is, in fact, equivalent to hypothesizing an exponential (CARA) utility function such as

$$W(\tilde{z}) = -\exp\left\{-\frac{\chi}{2}\tilde{z}\right\}$$

if the context is such that the argument of the function, z , is normally distributed. This hypothesis will be verified at the equilibrium of our model.

Under these hypotheses, it is easy to verify that the optimal futures position of speculator i is

$$b_i = \frac{p^f - E(\tilde{p}|p^f)}{\chi var(\tilde{p}|p^f)} \quad (16.5)$$

where the conditioning in the expectation and variance operators is made necessary by Equation (16.4). The form of Equation (16.5) is not surprising. It implies that the optimal futures position selected by a speculator will have the same sign as the expected difference between the futures price and the expected spot price, that is, a speculator will be short ($b > 0$) if and only if the futures price at which he sells is larger than the spot price at which he expects to be able to unload his position tomorrow. As to the size of his position, it will be proportional to the expected difference between the two prices, which is indicative of the size of the expected return, and inversely related to the perceived

riskiness of the speculation, measured by the product of the variance of the spot price with the Arrow-Pratt coefficient of risk aversion. More risk-averse speculators will assume smaller positions, everything else being the same.

Under a linear mean-variance specification of preferences, the producer's objective function becomes

$$\max_{x \geq 0, f} E(\tilde{p} | p^f) (g(x) - f - x) + p^f f - \frac{\xi}{2} (g(x) - f)^2 \text{var}(\tilde{p} | p^f)$$

where ξ is the absolute risk aversion measure for producers.

With this specification of the objective function, Equation (16.2), the FOC with respect to f , becomes

$$f = g(x(p^f)) + \frac{p^f - E(\tilde{p} | p^f)}{\xi \text{var}(\tilde{p} | p^f)} \equiv f(p^f) \quad (16.6)$$

which is the second part of the separation result alluded to previously. The optimal futures position of the representative producer consists in selling forward the totality of his production ($g(x)$) and then readjusting by a component that is simply the futures position taken by a speculator with the same degree of risk aversion. To see this, compare the last term in Equation (16.6) with Equation (16.5). A producer's actual futures position can be viewed as the sum of these two terms. He may under-hedge, that is, sell less than his future output at the futures price. This is so if he anticipates paying an insurance premium in the form of a sale price (p^f) lower than the spot price he expects to prevail tomorrow. But he could as well over-hedge and sell forward more than his total future output. That is, if he considers the current futures price to be a high enough price, he may be willing to speculate on it, selling high at the futures price what he hopes to buy low tomorrow on the spot market.

Putting together speculators' and producers' positions, the futures market clearing condition becomes:

$$\sum_{i=1}^n b_i + Nf = 0, \text{ or} \\ n \left\{ \frac{p^f - E(\tilde{p} | p^f)}{\chi \text{var}(\tilde{p} | p^f)} \right\} + N \left\{ \frac{p^f - E(\tilde{p} | p^f)}{\xi \text{var}(\tilde{p} | p^f)} \right\} + Ng(x(p^f)) = 0 \quad (16.7)$$

which must be solved for the equilibrium futures price p^f . Equation (16.7) makes clear that the equilibrium futures price p^f is dependent on the expectations held on the future spot price \tilde{p} ; we have previously emphasized the dependence on p^f of expectations about \tilde{p} . This apparently circular reasoning can be resolved under the rational expectations hypothesis, which consists of assuming that individuals have learned to understand the relationship summarized in Equation (16.4), that is,

$$E(\tilde{p} | p^f) = E[p(p^f, \tilde{\eta}) | p^f], \text{var}(\tilde{p} | p^f) = \text{var}[p(p^f, \tilde{\eta}) | p^f] \quad (16.8)$$

Definition 16.1:

In the context of this section, a Rational Expectations Equilibrium (REE) is

1. a futures price p^f solving Equation (16.7) given Equation (16.8), and the distributional assumption made on η , and
2. a spot price p solving Equation (16.4) given p^f and the realization of η .

The first part of the definition indicates that the futures price equilibrates the futures market at date 0 when agents rationally anticipate the effective condition under which the spot market will clear tomorrow and make use of the objective probability distribution on the stochastic parameter $\tilde{\eta}$. Given the supply of the commodity available tomorrow (itself a function of the equilibrium futures price quoted today), and given the particular value taken by $\tilde{\eta}$, (i.e., the final position of the demand curve), the second part specifies that the spot price clears the date 1 spot market.

16.4 Fully Revealing REE: An Example

Let us pursue this example one step further and assume that speculators have access to privileged information in the following sense: Before the futures exchange opens, speculator i , ($i = 1, \dots, n$) observes some unbiased approximation v_i to the future realization of the variable $\tilde{\eta}$. The signal v_i can be viewed as the future η itself plus an error of observation ω_i . The latter is specific to speculator i , but all speculators are similarly imprecise in the information they manage to gather. Thus,

$$v_i = \eta + \omega_i \text{ where the } \omega_i\text{'s are i.i.d. } N(0; \sigma_\omega^2)$$

across agents and across time periods.

This relationship can be interpreted as follows: η is a summary measure of the mood of consumers or of other conditions affecting demand. Speculators can obtain advanced information as to the particular value of this realization for the relevant period through, for instance, a survey of consumer's intentions or a detailed weather forecast (assuming the latter influences demand). These observations are not without errors, but (regarding these two periods as only one occasion of a multiperiod process where learning has been taking place), speculators are assumed to be sufficiently skilled to avoid systematic biases in their evaluations. In this model, this advance information is freely available to them.

Under these conditions, Equation (16.5) becomes

$$b_i = \frac{p^f - E(\tilde{p} | p^f; v_i)}{\chi \text{var}(\tilde{p} | p^f; v_i)} \equiv b(p^f; v_i),$$

where we make it explicit that both the expected value and the variance of the spot price are affected by the advance piece of information obtained by

speculator i . The Appendix details how these expectations can actually be computed, but this need not occupy us for the moment.

Formally, Equation (16.6) is unchanged, so that the futures market clearing condition can be written

$$Nf(p^f) + \sum_i^n b(p^f; v_i) = 0;$$

It is clear from this equation that the equilibrium futures price will be affected by the “elements” of information gathered by speculators. In fact, under appropriate regularity conditions, the market-clearing equation implicitly defines a function

$$p^f = l(v_1, v_2, \dots, v_n) \quad (16.9)$$

that formalizes this link and thus the information content of the equilibrium futures price.

All this implies that there is more than meets the eye in the conditioning on p^f of $E(\tilde{p}|p^f)$ and $var(\tilde{p}|p^f)$. So far the reasoning for this conditioning was given by Equation (16.4): A higher p^f stimulates supply from $g(x(p^f))$ and thus affects the equilibrium spot price. Now a higher p^f also indicates high v_i 's on average, thus transmitting information about the future realization of $\hat{\eta}$. The real implications of this link can be suggested by reference to Figure 16.1. In the absence of advance information, supply will be geared to the average demand conditions. \bar{Q} represents this average supply level, leading to a spot price \bar{p} under conditions of average demand ($\eta = 0$). If suppliers receive no advance warning of an abnormally high demand level, an above-average realization $\hat{\eta}$ requires a high price p_1 to balance supply and demand. If, on the other hand, speculators' advance information is transmitted to producers via the futures price, supply increases in anticipation of the high demand level and the price increase is mitigated.

We are now in a position to provide a precise answer to the question that has preoccupied us since Section 16.2: How much information is transmitted by the equilibrium price p^f ? It will not be a fully general answer: Our model has the nature of an example because it presumes specific functional forms. The result we will obtain certainly stands at one extreme on the spectrum of possible answers, but it can be considered as a useful benchmark. In what follows, we will construct, under the additional simplification $g(x) = \alpha x^{1/2}$, a consistent equilibrium in which the futures price is itself a summary of *all the information* there is to obtain, a summary which, in an operational sense, is fully equivalent to the complete list of signals obtained by all speculators. More precisely, we will show that the equilibrium futures price is an invertible (linear) function of $\sum v_j$ and that, indeed, it clears the futures market given that everyone realizes this property and bases his orders on the information he can thus extract. This result is important because $\sum v_j$ is a sufficient statistic for the entire vector (v_1, v_2, \dots, v_n) . This formal expression means that the sum contains as much relevant information for the problem at hand as the entire vector, in the sense that knowing the sum leads to placing the same market orders as knowing the

whole vector. Ours is thus a context where the answer to our question is: *All* the relevant information is aggregated in the equilibrium price and is revealed freely to market participants. The REE is thus *fully revealing*!

Let us proceed and make these precise assertions. Under the assumed technology, $q(x) = \alpha x^{1/2}$, Equations (16.3), (16.4), and (16.8) become, respectively,

$$\begin{aligned} g(x(p^f)) &= \frac{\alpha^2}{2} p^f \\ p(p^f, \tilde{\eta}) &= A - B p^f + \frac{1}{c} \tilde{\eta} \\ \text{with} \quad A &= \frac{a}{c}, \quad B = \frac{N}{c} \frac{\alpha^2}{2} \\ E(\tilde{p} | p^f) &= A - B p^f + \frac{1}{c} E(\tilde{\eta} | p^f) \\ \text{var}(\tilde{p} | p^f) &= \frac{1}{c^2} \text{var}(\tilde{\eta} | p^f) \end{aligned}$$

The informational structure is as follows. Considering the market as a whole, an experiment has been performed consisting of observing the values taken by n independent drawings of some random variable \bar{v} , where $\tilde{v} = \eta + \tilde{w}$ and \tilde{w} is $N(0, \sigma_w^2)$. The results are summarized in the vector $v = (v_1, v_2, \dots, v_n)$ or, as we shall demonstrate, in the sum of the v_j 's, $\sum v_j$, which is a *sufficient statistic* for $v = (v_1, v_2, \dots, v_n)$. The latter expression means that conditioning expectations on $\sum v_j$ or on $\sum v_j$ *and* the whole vector of v yields the same posterior distribution for $\tilde{\eta}$. In other words, the entire vector does not contain any information that is not already present in the sum. Formally, we have Definition 16.2.

Definition 16.2:

$(\frac{1}{n}) \sum v_j$ is a sufficient statistic for $v = (v_1, v_2, \dots, v_n)$ relative to the distribution $h(\eta)$ if and only if $h(\tilde{\eta} | \sum v_j, v) = h(\tilde{\eta} | \sum v_j)$.

Being a function of the observations [see Equation (16.9)], p^f is itself a statistic used by traders in calibrating their probabilities. The question is: How good a statistic can it be? How well can the futures price summarize the information available to the market? As promised, we now display an equilibrium where the price p^f is a sufficient statistic for the information available to the market; that is, it is invertible for the sufficient statistic $\sum v_j$. In that case, knowledge of p^f is equivalent to the knowledge of $\sum v_j$ and farmers' and speculators' expectations coincide. If the futures price has this revealing property, expectations held at equilibrium by all agents must be (see the Appendix for details):

$$E(\tilde{\eta} | p^f) = E(\tilde{\eta} | v_j, p^f) = E(\tilde{\eta} | \sum v_j) = \frac{\sigma_\eta^2}{n\sigma_\eta^2 + \sigma_w^2} \sum v_j \quad (16.10)$$

$$\text{var}(\tilde{\eta} | p^f) = \text{var}(\tilde{\eta} | v_j, p^f) = \frac{\sigma_w^2 \sigma_\eta^2}{n\sigma_\eta^2 + \sigma_w^2}. \quad (16.11)$$

Equations (16.10) and (16.11) make clear that conditioning on the futures price would, under our hypothesis, be equivalent to conditioning on $\sum v_j$, the latter being, of course, superior information relative to the single piece of individual information, v_i , initially obtained by speculator i . Using these expressions for the expectations in Equation (16.7), one can show after a few tedious manipulations that, as announced, the market-clearing futures price has the form

$$p^f = F + L \sum v_j \quad (16.12)$$

where

$$\begin{aligned} F &= \frac{(N\chi + n\xi) A}{(N\chi + n\xi)(B + 1) + N\alpha^2\xi\chi\frac{1}{c^2}\frac{\sigma_w^2\sigma_\eta^2}{n\sigma_\eta^2 + \sigma_w^2}} \text{ and} \\ L &= \frac{1}{c} \frac{\sigma_w^2\sigma_\eta^2}{n\sigma_\eta^2 + \sigma_w^2} \frac{F}{A}. \end{aligned}$$

Equation (16.12) shows the equilibrium price p^f to be proportional to $\sum v_j$ and thus a sufficient statistic as postulated. It satisfies our definition of an equilibrium. It is a market-clearing price, the result of speculators' and farmers' maximizing behavior, and it corresponds to an equilibrium state of expectations. That is, when Equation (16.12) is the hypothesized functional relationship between p^f and v , this relationship is indeed realized given that each agent then appropriately extracts the information $\sum v_j$ from the announcement of the equilibrium price.

16.5 The Efficient Market Hypothesis

The result obtained in Section 16.4 is without doubt extreme. It is interesting, however, as it stands as the paragon of the concept of market efficiency. Here is a formal and precise context in which the valuable pieces of information held by heterogeneously informed market participants are aggregated and freely transmitted to all via the trading process. This outcome is reminiscent of the statements made earlier in the century by the famous liberal economist F. von Hayek who celebrated the virtues of the market as an information aggregator [Hayek (1945)]. It must also correspond to what Fama (1970) intended when introducing the concept of strong form efficiency, defined as a situation where market prices fully reflect all publicly *and* privately held information.

The reader will recall that Fama (1970) also introduced the notions of *weak-form efficiency*, covering situations where market prices fully and instantaneously reflect the information included in historical prices, and of *semi-strong form efficiency* where prices, in addition, reflect all publicly available information (of whatever nature). A securities market equilibrium such as the one described in Chapter 9 under the heading of the CCAPM probably best captures what one can understand as semi-strong efficiency: Agents are rational

in the sense of being expected utility maximizers, they are homogeneously informed (so that all information is indeed publicly held), and they efficiently use all the relevant information when defining their asset holdings. In the CCAPM, no agent can systematically beat the market, a largely accepted hallmark of an efficient market equilibrium, provided beating the market is appropriately defined in terms of both risk and return.

The concept of Martingale, also used in Chapters 11 and 12, has long constituted another hallmark of market efficiency. It is useful here to provide a formal definition.

Definition 16.3

A stochastic process \tilde{x}_t is a Martingale with respect to an information set Φ_t if

$$E(\tilde{x}_{t+1} | \Phi_t) = \tilde{x}_t. \quad (16.13)$$

It is a short step from this notion of a Martingale to the assertion that one cannot beat the market, which is the case if the current price of a stock is the best predictor of its future price. The latter is likely to be the case if market participants indeed make full use of all available information: In that situation, future price changes can only be unpredictable. An equation like Equation (16.13) cannot be true exactly for stock prices as stock returns would then be zero on average. It is clear that what could be a Martingale under the previous intuitive reasoning would be a price series normalized to take account of dividends and a normal expected return for holding stock. To get an idea of what this would mean, let us refer to the price equilibrium Equation (9.2) of the CCAPM

$$U_1(Y_t)p_t = \delta E_t \{U_1(Y_{t+1})(p_{t+1} + Y_{t+1})\} \quad (16.14)$$

Making the assumption of risk neutrality, one obtains:

$$p_t = \delta E_t(p_{t+1} + Y_{t+1}) \quad (16.15)$$

If we entertain, for a moment, the possibility of a non-dividend paying stock, $Y_t \equiv 0$, then Equation (16.14) indeed implies that the normalized series $x_t = \delta^t p_t$ satisfies Equation (16.13) and is thus a Martingale. This normalization implies that the expected return on stockholding is constant and equal to the risk-free rate. In the case of a dividend-paying stock, a similar, but slightly more complicated, normalization yields the same result.

The main points of this discussion are (1) that a pure Martingale process requires adjusting the stock price series to take account of dividends and the existence of a positive normal return, and (2) that the Martingale property is a mark of market efficiency only under a strong hypothesis of risk neutrality that includes, as a corollary, the property that expected return to stockholding is constant. The large empirical literature on market efficiency has not always been able to take account appropriately of these qualifications. See Leroy (1989) for an in-depth survey of this issue.

Our model of the previous section is more ambitious, addressing as it does, the concept of strong form efficiency. Its merit is to underline what it takes for this extreme concept to be descriptive of reality, thus also helping to delineate its limits. Two of these limits deserve mentioning. The first one arises once one attempts, plausibly, to get rid of the hypothesis that speculators are able costlessly to obtain their elements of privileged information. If information is free, it is difficult to see why all speculators would not get all the relevant information, thus reverting to a model of homogeneous information. However, the spirit of our example is that resources are needed to collect information and that speculators are those market participants specializing in this costly search process. Yet why should speculator i expand resources to obtain private information v_i when the equilibrium price will freely reveal to him the sufficient statistic $\sum v_j$, which by itself is more informative than the information he could gather at a cost. The very fact that the equilibrium REE price is fully revealing implies that individual speculators have no use for their own piece of information, with the obvious corollary that they will not be prepared to spend a penny to obtain it. On the other hand, if speculators are not endowed with privileged information, there is no way the equilibrium price will be the celebrated information aggregator and transmitter. In turn, if the equilibrium price is not informative, it may well pay for speculators to obtain valuable private information. We are thus trapped in a vicious circle that results in the nonexistence of equilibrium, an outcome Grossman and Stiglitz (1980) have logically dubbed “the impossibility of informationally efficient markets.”

Another limitation of the conceptual setup of Section 16.4 resides in the fact that the hypotheses required for the equilibrium price to be fully revealing are numerous and particularly severe. The rational expectations hypothesis includes, as always, the assumption that market participants understand the environment in which they operate. This segment of the hypothesis is particularly demanding in the context of our model and it is crucial for agents to be able to extract sufficient statistics from the equilibrium futures price. By that we mean that, for individual agents to be in position to read all the information concealed in the equilibrium price, they need to know exactly the number of uninformed and informed agents and their respective degrees of risk aversion, which must be identical inside each agent class. The information held by the various speculators must have identical precision (i.e., an error term with the same variance), and none of the market participants can be motivated by liquidity considerations. All in all, these requirements are simply too strong to be plausibly met in real-life situations. Although the real-life complications may be partly compensated for by the fact that trading is done on a repeated, almost continuous basis, it is more reasonable to assume that the fully revealing equilibrium is the exception rather than the rule.

The more normal situation is certainly one where some, but not all, information is aggregated and transmitted by the equilibrium price. In such an equilibrium, the incentives to collect information remain, although if the price is too good a transmitter, they may be significantly reduced. The nonexistence-of-equilibrium problem uncovered by Grossman and Stiglitz is then more a curiosity

than a real source of worry. Equilibria with partial transmission of information have been described in the literature under the heading *noisy rational expectation equilibrium*. The apparatus is quite a bit messier than in the reference case discussed in Section 16.4 and we will not explore it further (see Hellwig (1980) for a first step in this direction). Suffice it to say that this class of models serves as the basis for the branch of financial economics known as *market microstructure* which strives to explain the specific forms and rules underlying asset trading in a competitive market environment. The reader is referred to O'Hara (1997) for a broad coverage of these topics.

References

- Akerlof, G. (1970), "The Market for Lemons: Qualitative Uncertainty and the Market Mechanism," *The Quarterly Journal of Economics*, 89, 488–500.
- Danthine, J.-P. (1978), "Information, Futures Prices and Stabilizing Speculation," *Journal of Economic Theory*, 17, 79–98.
- Fama, E. (1970), "Efficient Capital Markets: A Review of Theory and Empirical Work," *Journal of Finance* 25, 383–417.
- Greenwald, B., Stiglitz, J.E. (1993), "Financial Market Imperfections and Business Cycles," *The Quarterly Journal of Economics*, 108, 77–115.
- Grossman, S., Stiglitz, J.E. (1980), "On the Impossibility of Informationally Efficient Markets," *American Economic Review*, 70(3), 393–408.
- Hayek, F. H. (1954), "The Use of Knowledge in Society," *American Economic Review*, 61, 519–530.
- Hellwig, M. F. (1980), "On the Aggregation of Information in Competitive Markets," *Journal of Economic Theory*, 26, 279–312.
- Hirshleiffer, J., Riley, J.G. (1992), *The Analytics of Uncertainty and Information*, Cambridge University Press, Cambridge.
- LeRoy, S. F. (1989), "Efficient Capital Markets and Martingales," *Journal of Economic Literature* 27, 1583–1621.
- O'Hara, M. (1997), *Market Microstructure Theory*, Basil Blackwell, Malden, Mass.
- Stiglitz, J. E., Weiss, A. (1981), "Credit Rationing in Markets with Imperfect Information," *American Economic Review*, 71, 393–410.
- Townsend, R. (1979), "Optimal Contracts and Competitive Markets with Costly State Verification," *Journal of Economic Theory*, 21, 417–425.

Appendix: Bayesian Updating with the Normal Distribution

Theorem A 16.1:

If we assume \tilde{x} and \tilde{y} are two normally distributed vectors with

$$\begin{pmatrix} \tilde{x} \\ \tilde{y} \end{pmatrix} \sim N \left(\begin{pmatrix} \bar{x} \\ \bar{y} \end{pmatrix}, V \right),$$

with matrix of variances and covariances

$$V = \begin{pmatrix} V_{xx} & V_{xy} \\ V_{xy} & V_{yy} \end{pmatrix},$$

then the distribution of \tilde{x} conditional on the observation $\tilde{y} = y^0$ is normal with mean $\bar{x} + V_{xx}V_{yy}^{-1}(y^0 - \bar{y})$ and covariance matrix $V_{xx} - V_{xy}V_{yy}^{-1}V_{xy}$

Applications

Let $\tilde{v}_i = \tilde{\eta} + \tilde{\omega}_i$.

If $\begin{pmatrix} \tilde{\eta} \\ \tilde{\omega}_i \end{pmatrix} \sim N \left(\begin{pmatrix} 0 \\ 0 \end{pmatrix}, \begin{pmatrix} \sigma_\eta^2 & \sigma_\eta^2 \\ \sigma_\eta^2 & \sigma_\eta^2 + \sigma_\omega^2 \end{pmatrix} \right)$, then

$$\begin{aligned} E(\tilde{\eta} | v_i) &= 0 + \frac{\sigma_\eta^2}{\sigma_\eta^2 + \sigma_\omega^2} v_i \\ V(\tilde{\eta} | v_i) &= \sigma_\eta^2 - \frac{\sigma_\eta^4}{\sigma_\eta^2 + \sigma_\omega^2} = \frac{\sigma_\eta^2 \sigma_\omega^2}{\sigma_\eta^2 + \sigma_\omega^2} \end{aligned}$$

If $\begin{pmatrix} \tilde{\eta} \\ \sum \tilde{v}_i \end{pmatrix} \sim N \left(\begin{pmatrix} 0 \\ 0 \end{pmatrix}, \begin{pmatrix} \sigma_\eta^2 & n\sigma_\eta^2 \\ n\sigma_\eta^2 & n^2\sigma_\eta^2 + n\sigma_\omega^2 \end{pmatrix} \right)$, then

$$\begin{aligned} E(\tilde{\eta} | \sum v_i) &= 0 + \frac{n\sigma_\eta^2}{n^2\sigma_\eta^2 + n\sigma_\omega^2} \left(\sum v_i \right) = \frac{\sigma_\eta^2}{n\sigma_\eta^2 + \sigma_\omega^2} \sum v_i \\ var(\tilde{\eta} | \sum v_i) &= \sigma_\eta^2 - n\sigma_\eta^2 \frac{1}{n^2\sigma_\eta^2 + n\sigma_\omega^2} n\sigma_\eta^2 = \frac{\sigma_\eta^2 \sigma_\omega^2}{n\sigma_\eta^2 + \sigma_\omega^2} \end{aligned}$$

Review of Basic Options Concepts and Terminology

March 24, 2005

1 Introduction

The purchase of an options contract gives the buyer the right to buy (call options contract) or sell (put options contract) some other asset under pre-specified terms and circumstances. This *underlying* asset, as it is called, can in principle be anything with a well-defined price. For example, options on individual stocks, portfolios of stocks (i.e., indices such as the *S&P500*), futures contracts, bonds, and currencies are actively traded. Note that options contracts do not represent an obligation to buy or sell and, as such, must have a positive, or at worst zero, price.

“American” style options allow the right to buy or sell (the so-called “right of exercise”) at any time on or before a pre-specified future date (the “expiration” date). “European” options allow the right of exercise only at the pre-specified expiration date. Most of our discussion will be in the context of European call options. If the underlying asset does not provide any cash payments during the time to expiration (no dividends in the case of options on individual stocks), however, it can be shown that it is never wealth maximizing to exercise an American call option prior to expiration (its market price will at least equal and likely exceed its value if exercised). In this case, American and European call options are essentially the same, and are priced identically. The same statement is not true for puts.

In all applied options work, it is presumed that the introduction of options trading does not influence the price process of the underlying asset on which they are written. For a full general equilibrium in the presence of incomplete markets, however, this will not generally be the case.

2 Call and Put Options on Individual Stocks

1. European call options

- (a) Definition: A European call options contract gives the owner the right to buy a pre-specified number of shares of a pre-specified stock (the

underlying asset) at a pre-specified price (the “strike” or “exercise” price) on a pre-specified future date (the expiration date). American options allow exercise “on or before” the expiration date. A contract typically represents 100 options with the cumulative right to buy 100 shares.

- (b) Payoff diagram: It is customary to describe the payoff to an individual call option by its value at expiration as in Figure 1.

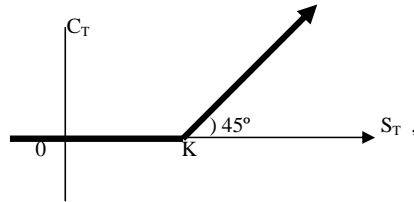


Figure 1: Payoff Diagram: European Call Option

In Figure 1, S_T denotes the possible values of the underlying stock at expiration date T , K the exercise price, and C_T the corresponding call value at expiration. Algebraically, we would write

$$C_T = \max \{0, S_T - K\}.$$

Figure 1 assumes the perspective of the buyer; the payoff to the seller (the so-called “writer” of the option) is exactly opposite to that of the buyer. See Figure 2.

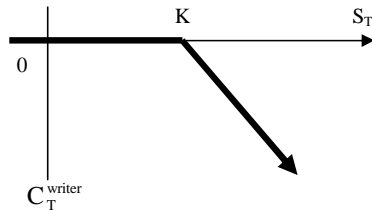


Figure 2: Payoff Diagram: European Call-Writer's Perspective

Note that options give rise to exactly offsetting wealth transfers between the buyer and the seller. The options related wealth positions of buyers and sellers must thus always sum to zero. As such we say that options are in zero net supply, and thus are not elements of “ M ”, the market portfolio of the classic CAPM.

- (c) Remarks: The purchaser of a call option is essentially buying the expected price appreciation of the underlying asset in excess of the exercise price. As we will make explicit in a later chapter, a call option

can be thought of as very highly leveraged position in the underlying stock - a property that makes it an ideal vehicle for speculation: for relatively little money (as the call option price will typically be much less than the underlying share's price) the buyer can acquire the upward potential.

There will, of course, be no options market without substantial diversity of expectations regarding the future price behavior of the underlying stock.

2. European Put Options

- (a) Definition: A European put options contract gives the buyer the right to sell a pre-specified number of shares of the underlying stock at a pre-specified price (the “exercise” or “strike” price) on a pre-specified future date (the expiration date). American puts allow for the sale on or before the expiration date. A typical contract represents 100 options with the cumulative right to sell 100 shares.
- (b) Payoff diagram: In the case of a put, the payoff at expiration to an individual option is represented in Figure 3.

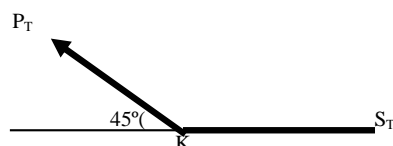


Figure 3: Payoff Diagram for a European Put

In Figure 3, P_T denotes the put's value at expiration; otherwise, the notation is the same as for calls. The algebraic equivalent to the payoff diagram is

$$P_T = \max\{0, K - S_T\}$$

The same comments about wealth transfers apply equally to the put as to the call; puts are thus also not included in the market portfolio M .

- (c) Remarks: Puts pay off when the underlying asset's price falls below the exercise price at expiration. This makes puts ideal financial instruments for “insuring” against price declines. Let us consider the payoff to the simplest “fundamental hedge” portfolio:

$$\left\{ \begin{array}{ll} 1 \text{ share of} & 1 \text{ put written on the} \\ \text{stock} & \text{stock with exercise price } K \end{array} \right\}.$$

Table 1: Payoff Table for Fundamental Hedge

Events	$S_T \leq K$	$S_T > K$
Stock	S_T	S_T
Put	$K - S_T$	0
Hedge Portfolio	K	S_T

To see how these two securities interact with one another, let us consider their net total value at expiration:

The diagrammatic equivalent is in Figure 4.

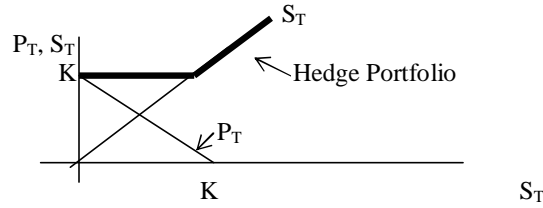


Figure 4: Payoff Diagram: Fundamental Hedge

The introduction of the put effectively bounds the share price to fall no lower than K . Such insurance costs money, of course, and its price is the price of the put.

Puts and calls are fundamentally different securities: calls pay off when the underlying asset's price at expiration exceeds K ; puts pay off when its price falls short of K . Although the payoff patterns of puts and calls are individually simple, virtually any payoff pattern can be replicated by a properly constructed portfolio of these instruments.

3 The Black-Scholes Formula for a European Call Option.

1. What it presumes: The probability distribution on the possible payoffs to call ownership will depend upon the underlying stock's price process. The Black-Scholes formula gives the price of a European call under the following assumptions:
 - (a) the underlying stock pays no dividends over the time to expiration;
 - (b) the risk free rate of interest is constant over the time to expiration;
 - (c) the continuously compounded rate of return on the underlying stock is governed by a geometric Brownian motion with constant mean and variance over the time to expiration.

This model of rate of return evolution essentially presumes that the rate of return on the underlying stock – its rate of price appreciation since there are no dividends – over any small interval of time $\Delta t \in [0, T]$ is given by

$$r_{t, t+\Delta t} = \frac{\Delta S_{t, t+\Delta t}}{S_t} = \hat{\mu}\Delta t + \hat{\sigma}\tilde{\varepsilon}\sqrt{\Delta t}, \quad (1)$$

where $\tilde{\varepsilon}$ denotes the standard normal distribution and $\hat{\mu}$, $\hat{\sigma}$ are, respectively, the annualized continuously compounded mean return and the standard deviation of the continuously compounded return on the stock. Under this abstraction the rate of return over any small interval of time Δt is distributed $N(\hat{\mu}\Delta t, \hat{\sigma}\sqrt{\Delta t})$; furthermore, these returns are independently distributed through time. Recall (Chapter 3) that these are the two most basic statistical properties of stock returns. More precisely, Equation (1) describes the discrete time approximation to geometric Brownian motion.

True Geometric Brownian motion presumes continuous trading, and its attendant continuous compounding of returns. Of course continuous trading presumes an uncountably large number of “trades” in any finite interval of time, which is impossible. It should be thought of as a very useful mathematical abstraction.

Under continuous trading the expression analogous to Equation (1) is

$$\frac{dS}{S} = \hat{\mu}dt + \hat{\sigma}\tilde{\varepsilon}\sqrt{dt}. \quad (2)$$

Much more will be said about this price process in the web-complement entitled “*An Intuitive Overview of Continuous Time Finance*”

2. The formula: The Black-Scholes formula is given by

$$C_T(S, K) = SN(d_1) - e^{\hat{r}_f T} KN(d_2)$$

where

$$\begin{aligned} d_1 &= \frac{\ln\left(\frac{S}{K}\right) + \left(\hat{r}_f - \frac{1}{2}\hat{\sigma}^2\right)T}{\hat{\sigma}\sqrt{T}} \\ d_2 &= d_1 - \hat{\sigma}\sqrt{T}. \end{aligned}$$

In this formula:

S = the price of the stock “today” (at the time the call valuation is being undertaken);

K = the exercise price;

T = the time to expiration, measured in years;

\hat{r}_f = the estimated continuously compounded annual risk free rate;

$\hat{\sigma}$ = the estimated standard deviation of the continuously compounded rate of return on the underlying asset annualized; and

$N(\cdot)$ is the standard normal distribution.

In any practical problem, of course, $\hat{\sigma}$ must be estimated. The risk free rate is usually unambiguous as normally there is a T-bill coming due on approximately the same date as the options contracts expire (U.S. markets).

3. An example

Suppose

$S = \$68$

$K = \$60$

$T = 88 \text{ days} = \frac{88}{365} = .241 \text{ years}$

$r_f = 6\%$ (not continuously compounded)

$\hat{\sigma} = .40$

The r_f inserted into the formula is that rate which, when continuously compounded, is equivalent to the actual 6% annual rate; this must satisfy

$$e^{\hat{r}_f} = 1.06, \text{ or } \hat{r}_f = \ln(1.06) = .058.$$

Thus,

$$\begin{aligned} d_1 &= \frac{\ln \frac{68}{60} + (.058 + \frac{1}{2}(.4)^2(.241))}{(.40)\sqrt{.241}} = .806 \\ d_2 &= .806 - (.40)\sqrt{.241} = .610 \\ N(d_1) &= N(.806) \approx .79 \\ N(d_2) &= N(.610) \approx .729 \\ C &= \$.68(.79) - e^{-(.058)(.241)}(\$60)(.729) \\ &= \$10.60 \end{aligned}$$

4. Estimating σ .

We gain intuition about the Black-Scholes model if we understand how its inputs are obtained, and the only input with any real ambiguity is σ . Here we present a straightforward approach to its estimation based on the security's historical price series.

Since volatility is an unstable attribute of a stock, by convention it is viewed as unreliable to go more than 180 days into the past for the choice of historical period. Furthermore, since we are trying to estimate the

μ, σ of a continuously compounded return process, the interval of measurement should be, in principle, as small as possible. For most practical applications, daily data is the best we can obtain.

The procedure is as follows:

i) Select the number of chosen historical observations and index them $i = 0, 1, 2, \dots, n$ with observation 0 most distant into the past and observation n most recent. This gives us $n + 1$ price observations. From this we will obtain n daily return observations.

ii) Compute the equivalent continuously compounded ROR on the underlying asset over the time intervals implied by the selection of the data (i.e., if we choose to use daily data we compute the continuously compounded daily rate):

$$r_i = \ln \left(\frac{S_i}{S_{i-1}} \right)$$

This is the equivalent continuously compounded ROR from the end of period $i - 1$ to the end of period i .

Why is this the correct calculation? Suppose $S_i = 110$, $S_{i-1} = 100$; we want that continuously compounded return x to be such that

$$\begin{aligned} S_{i-1}e^x &= S_i, \text{ or } 100e^x = 110 \\ e^x &= (110/100) \\ x &= \ln\left(\frac{110}{100}\right) = .0953. \end{aligned}$$

This is the continuously compounded rate that will increase the price from \$100 to \$110.

iii) Compute the sample mean

$$\hat{\mu} = \frac{1}{n} \left(\sum_{i=1}^n r_i \right)$$

Remark: If the time intervals are all adjacent; i.e., if we have not omitted any observations, then

$$\begin{aligned} \hat{\mu} &= \frac{1}{n} \sum_{i=1}^n r_i = \frac{1}{n} \left[\ln \left(\frac{S_1}{S_0} \right) + \ln \left(\frac{S_2}{S_1} \right) + \dots + \ln \left(\frac{S_n}{S_{n-1}} \right) \right] \\ &= \frac{1}{n} \left[\ln \left(\frac{S_1}{S_0} \cdot \frac{S_2}{S_1} \dots \frac{S_n}{S_{n-1}} \right) \right] \\ &= \frac{1}{n} \ln \left(\frac{S_n}{S_0} \right) \end{aligned}$$

Note that if we omit some calendar observations – perhaps due to, say, merger rumors at the time which are no longer relevant – this shortcut fails.

iv) Estimate σ

$$\hat{\sigma} = \sqrt{\frac{1}{n-1} \left(\sum_{i=1}^n (r_i - \hat{\mu})^2 \right)} , \text{ or}$$

$$\hat{\sigma} = \sqrt{\frac{1}{n-1} \sum_{i=1}^n r_i^2 - \frac{1}{n(n-1)} \left(\sum_{i=1}^n r_i \right)^2}$$

Example: Consider the (daily) data in Table 2.

Table 2		
<u>Period</u>	<u>Closing price</u>	$\ln \left(\frac{S_i}{S_{i-1}} \right)$
i=0	\$26	$r_1 = \ln \left(\frac{26.50}{26} \right) = .0190482$
i=1	\$26.50	$r_2 = \ln \left(\frac{26.25}{26.50} \right) = -.009479$
i=2	\$26.25	$r_3 = \ln \left(\frac{26.25}{26.25} \right) = 0$
i=3	\$26.25	$r_4 = \ln \left(\frac{26.50}{26.25} \right) = .0094787$
i=4	\$26.50	

In this case $\hat{\mu} = \frac{1}{4} \ln \left(\frac{26.50}{26} \right) = 1/4; \ln(1.0192308) = .004762$

Using the above formula,

$$\sum_{i=1}^4 r_i^2 = (.0190482)^2 + (-.009479)^2 + (.0094787)^2 = .0005472$$

$$\left(\sum_{i=1}^4 r_i \right)^2 = (.0190482 - .009479 + .0094787)^2 = (.01905)^2 = .0003628$$

$$\hat{\sigma} = \sqrt{\frac{1}{3}(.005472) - \frac{1}{12}(.003628)}$$

$$= \sqrt{.0001809 - .0000302} = \sqrt{.0001507} = .0123$$

$$\hat{\sigma}^2 = (.0123)^2 = .0001507$$

v) Annualize the estimate of σ

We will assume 250 trading days per year. Our estimate for the continuously compounded annual return is thus:

$$\hat{\sigma}_{annual}^2 = 250 \underbrace{(.0001507)}_{\sigma_{daily}^2} = .0377$$

Remark: Why do we do this? Why can we multiply our estimate by 250 to scale things up? We can do this because of our geometric Brownian

motion assumption that returns are independently distributed. This is detailed as follows:

Our objective: an estimate for $\text{var} \left(\ln \left(\frac{S_{T=1yr}}{S_0} \right) \right)$ given 250 trading days,

$$\begin{aligned}
 \text{var} \left(\ln \left(\frac{S_{T=1yr}}{S_0} \right) \right) &= \text{var} \left(\ln \left(\frac{S_{day250}}{S_0} \right) \right) \\
 &= \text{var} \left(\ln \left(\frac{S_{day1}}{S_0} \cdot \frac{S_{day2}}{S_{day1}} \cdots \frac{S_{day250}}{S_{day249}} \right) \right) \\
 &= \text{var} \left(\ln \left(\frac{S_{day1}}{S_0} \right) + \ln \left(\frac{S_{day2}}{S_{day1}} \right) + \dots + \ln \left(\frac{S_{day250}}{S_{day249}} \right) \right) \\
 &= \text{var} \left(\ln \left(\frac{S_{day1}}{S_0} \right) \right) + \text{var} \left(\ln \left(\frac{S_{day2}}{S_{day1}} \right) \right) + \dots + \text{var} \left(\ln \left(\frac{S_{day250}}{S_{day249}} \right) \right)
 \end{aligned}$$

The latter equivalence is true because returns are uncorrelated from day to day under the geometric Brownian motion assumption.

Furthermore, the daily return distribution is presumed to be the same for every day, and thus the daily variance is the same for every day under geometric Brownian motion.

Thus, $\text{var} \left(\ln \left(\frac{S_{T=1yr}}{S_0} \right) \right) = 250 \sigma_{daily}^2$.

We have obtained an estimate for σ_{daily}^2 which we will write as $\hat{\sigma}_{daily}^2$. To convert this to an annual variance, we must thus multiply by 250.

Hence $\hat{\sigma}_{annual}^2 = 250 \cdot \sigma_{daily}^2 = .0377$, as noted.

If a weekly $\hat{\sigma}^2$ were obtained, it would be multiplied by 52.

4 The Black-Scholes Formula for an Index.

Recall that the Black-Scholes formula assumed that the underlying stock did not pay any dividends, and if this is not the case an adjustment must be made. A natural way to consider adapting the Black-Scholes formula to the dividend situation is to replace the underlying stock's price S in the formula by $S - PV(EDIVs)$, where $PV(EDIVs)$ is the present value, relative to $t = 0$, the date at which the calculation is being undertaken, of all dividends expected to be paid over the time to the option's expiration. In cases where the dividend is highly uncertain this calculation could be problematic. We want to make such an adjustment because the payment of a dividend reduces the underlying stock's value by the amount of the dividend and thus reduces the value (*ceteris paribus*) of the option written on it. Options are not "dividend protected," as it is said.

For an index, such as the S&P₅₀₀, the dividend yield on the index portfolio can be viewed as continuous, and the steady payment of this dividend will have

a continuous tendency to reduce the index value. Let d denote the dividend yield on the index. In a manner exactly analogous to the single stock dividend treatment noted above, the corresponding Black-Scholes formula is

$$\begin{aligned}
 C &= Se^{-dT}N(d_1) - Ke^{-r_f T}N(d_2) \\
 \text{where } d_1 &= \frac{\ln S/K + \left(r_f - d + \sigma^2/2\right)T}{\sigma\sqrt{T}} \\
 d_2 &= d_1 - \sigma\sqrt{T} .
 \end{aligned}$$

An Intuitive Overview of Continuous Time Finance

June 1, 2005

1 Introduction

If we think of stock prices as arising from the equilibration of traders demands and supplies, then the binomial model is implicitly one in which security trading occurs at discrete time intervals, however short, and this is factually what actually happens. It will be mathematically convenient, however, to abstract from this intuitive setting and hypothesize that trading takes place “continuously.” This is consistent with the notion of continuous compounding. But it is not fully realistic: it implies that an uncountable number of individual transactions may transpire in any interval of time, however small, which is physically impossible.

Continuous time finance is principally concerned with techniques for the pricing of derivative securities under the fiction of continuous trading. These techniques frequently allow closed form solutions to be obtained – at the price of working in a context that is less intuitive than discrete time. In this Appendix we hope to convey some idea as to how this is done.

We will need first to develop a continuous time model of a stock’s price evolution through time. Such a model must respect the basic statistical regularities which are known to characterize, empirically, equity returns:

- (i) stock prices are lognormally distributed, which means that returns (continuously compounded) are normally distributed;
- (ii) for short time horizons stock returns are independently and identically distributed over non-overlapping time intervals.

After we have faithfully represented these equity regularities in a continuous time setting, we will move on to a consideration of derivatives pricing. In doing so we aim to give some idea how the principles of risk neutral valuation carry over to this specialized setting. The discussion aims at intuition; no attempt is made to be mathematically complete.

In all cases this intuition has its origins in the discrete time context. This leads to a discussion of random walks.

2 Random Walks and Brownian Motion

Consider a time horizon composed of N adjacent time intervals each of duration Δt , and indexed by $t_0, t_1, t_2, \dots, t_N$; that is,

$$t_i - t_{i-1} = \Delta t, \quad i = 1, 2, \dots, N.$$

We define a discrete time stochastic process on this succession of time indices by

$$\begin{aligned} x(t_0) &= 0 \\ x(t_{j+1}) &= x(t_j) + \tilde{\varepsilon}(t_j)\sqrt{\Delta t}, \quad j = 0, 1, 2, \dots, N-1, \end{aligned}$$

where, for all j , $\tilde{\varepsilon}(t_j) \sim N(0, 1)$. It is further assumed that the random factors $\tilde{\varepsilon}(t_j)$ are independent of one another; i.e.,

$$E(\tilde{\varepsilon}(t_j)\tilde{\varepsilon}(t_i)) = 0, \quad i \neq j.$$

This is a specific example of a random walk, specific in the sense that the uncertain disturbance term follows a normal distribution¹.

We are interested to understand the behavior of a random walk over extended time periods. More precisely, we want to characterize the statistical properties of the difference

$$x(t_k) - x(t_j) \text{ for any } j < k.$$

Clearly,

$$\tilde{x}(t_k) - x(t_j) = \sum_{i=j}^{k-1} \tilde{\varepsilon}(t_i)\sqrt{\Delta t}.$$

Since the random disturbances $\tilde{\varepsilon}(t_i)$ all have mean zero,

$$E(\tilde{x}(t_k) - x(t_j)) = 0.$$

Furthermore,

$$\begin{aligned} \text{var}(x(t_k) - x(t_j)) &= E\left(\sum_{i=j}^{k-1} \tilde{\varepsilon}(t_i)\sqrt{\Delta t}\right)^2 \\ &= E\left(\sum_{i=j}^{k-1} [\tilde{\varepsilon}(t_i)]^2 \Delta t\right) \text{ (by independence)} \\ &= \sum_{i=j}^{k-1} (1)\Delta t = (k-j)\Delta t, \text{ since} \\ E[\tilde{\varepsilon}(t_i)]^2 &= 1. \end{aligned}$$

¹In particular, a very simple random walk could be of the form $x(t_{j+1}) = x(t_j) + n(t_j)$, where for all $j = 0, 1, 2, \dots$

$$n(t_j) = \begin{cases} +1, & \text{if a coin is flipped and a head appears} \\ -1, & \text{if a coin is flipped and a tail appears} \end{cases}.$$

At each time interval $x(t_j)$ either increases or diminishes by one depending on the outcome of the coin toss. Suppose we think of $x(t_0) \equiv 0$ as representing the center of the sidewalk where an intoxicated person staggers one step to the right or to the left of the center in a manner that is consistent with independent coin flips (heads implies to the right). This example is the source of the name “random walk.”

If we identify

$$x_{t_j} = \ln q_{t_j}^e,$$

where $q_{t_j}^e$ is the price of the stock at time t_j , then this simple random walk model becomes a candidate for our model of stock price evolution beginning from $t = 0$: At each node t_j , the logarithm of the stock's price is distributed normally, with mean $\ln q_{t_0}^e$ and variance $j\Delta t$.

Since the discrete time random walk is so respectful of the empirical realities of stock prices, it is natural to seek its counterpart for continuous time. This is referred to as a "Brownian Motion" (or a Wiener process), and it represents the limit of the discrete time random walk as we pass to continuous time; i.e., as $\Delta t \mapsto 0$. It is represented symbolically by

$$dz = \tilde{\varepsilon}(t)\sqrt{dt},$$

where $\tilde{\varepsilon}(t) \sim N(0, 1)$, and for any times t, t' where $t \neq t'$, and $E(\tilde{\varepsilon}(t')\tilde{\varepsilon}(t)) = 0$. We used the word "symbolically" not only because the term dz does not represent a differential in the terminology of ordinary calculus but also because we make no attempt here to describe how such a limit is taken. Following what is commonplace notation in the literature we will also not write a \sim over z even though it represents a random quantity.

More formally, a stochastic process $z(t)$ defined on $[0, T]$ is a Brownian motion provided the following three properties are satisfied:

- (i) for any $t_1 < t_2$, $z(t_2) - z(t_1)$ is normally distributed with mean zero and variance $t_2 - t_1$;
- (ii) for any $0 \leq t_1 < t_2 \leq t_3 < t_4$, $z(t_4) - z(t_3)$ is statistically independent of $z(t_2) - z(t_1)$; and
- (iii) $z(t_0) \equiv 0$ with probability one.

A Brownian motion is a very unusual stochastic process, and we can only give a hint about what is actually transpiring as it evolves. Three of its properties are considered below:

1. First, a Brownian motion is a continuous process. If we were able to trace out a sample path $z(t)$ of a Brownian motion, we would not see any jumps².
2. However, this sample path is not at all "smooth" and is, in fact, as "jagged as can be" which we formalize by saying that it is nowhere differential. A function must be essentially smooth if it is to be differentiable. That is, if we magnify a segment of its time path enough, it will appear approximately linear. This latter "smoothness" is totally absent with a Brownian motion.
3. Lastly a Brownian motion is of "unbounded variation." This is perhaps the least intuitive of its properties. By this is intended the idea that if we could take one of those mileage wheels which are drawn along a route on a map to assess the overall distance (each revolution of the wheel corresponding to a fixed number of kilometers) and apply it to the sample path of a Brownian motion,

²At times such as the announcement of a take over bid, stock prices exhibit jumps. We will not consider such "jump processes," although considerable current research effort is being devoted to studying them, and to the pricing of derivatives written on them.

then no matter how small the time interval, the mileage wheel would record “an infinite distance” (if it ever got to the end of the path!)

One way of visualizing such a process is to imagine a rough sketch of a particular sample path where we connect its position at a sequence of discrete time intervals by straight lines. Figure 1 proposes one such path.



Suppose that we were next to enlarge the segment between time intervals t_1 and t_2 . We would find something on the order of Figure 2.



Continue this process of taking a segment, enlarging it, taking another sub-segment of that segment, enlarging it etc., etc. (in Figure 2 we would next enlarge the segment from t_3 to t_4). Under a typical differentiable function of bounded variation, we would eventually be enlarging such a small segment that it would appear as a straight line. With a Brownian motion, however, this will never happen. No matter how much we enlarge even a segment that corresponds to an arbitrarily short time interval, the same “sawtooth” pattern will appear, and there will be many, many “teeth”.

A Brownian motion process represents a very special case of a continuous

process with independent increments. For such processes, the standard deviation per unit of time becomes unbounded as the interval becomes smaller and smaller:

$$\lim_{\Delta t \rightarrow 0} \frac{\sigma \sqrt{\Delta t}}{\Delta t} = \lim_{\Delta t \rightarrow 0} \frac{\sigma}{\sqrt{\Delta t}} = \infty.$$

No matter how small the time period, proportionately, a *lot* of variation remains. This constitutes our translation of the abstraction of a discrete time random walk to a context of continuous trading³.

3 More General Continuous Time Processes

A Brownian motion will be the principal building block of our description of the continuous time evolution of a stock's price – it will be the “engine” or “source” of the uncertainty. To it is often added a deterministic component which is intended to capture the “average” behavior through time of the process. Together we have something of the form

$$dx(t) = a dt + b \tilde{\varepsilon}(t) \sqrt{dt} = a dt + b dz, \quad (1)$$

where the first component is the deterministic one and a is referred to as the drift term.

This is an example of a generalized Brownian motion or, to use more common terminology, a generalized Wiener process. If there were no uncertainty $x(t)$ would evolve deterministically; if we integrate

$$dx(t) = a dt, \text{ we obtain}$$

$$x(t) = x(0) + at$$

The solution to (1) is thus of the form

$$x(t) = x(0) + at + bz(t), \quad (2)$$

where the properties of $z(t)$ were articulated earlier (recall properties (1), (2), and (3) of the definition). These imply that:

$$\begin{aligned} E(x(t)) &= x(0) + at, \\ \text{var}(x(t)) &= b^2 t, \text{ and} \\ s.d.(x(t)) &= b\sqrt{t}. \end{aligned}$$

Equation (2) may be further generalized to allow the coefficients to depend upon the time and the current level of the process:

$$dx(t) = a(x(t), t) dt + b(x(t), t) dz. \quad (3)$$

³The name Brownian motion comes from a 19th century physicist Brown, who studied the behavior of dust particles floating on the surface of water. Under a microscope dust particles are seen to move randomly about in a manner similar to the sawtooth pattern above except that the motion can be in any 360° direction. The interpretation of the phenomena is that the dust particles experience the effect of random collisions by moving water molecules.

In this latter form, it is referred to as an Ito process after one of the earliest and most important developers of this field. An important issue in the literature - but one we will ignore - is to determine the conditions on $a(x(t), t)$ and $b(x(t), t)$ in order for equation (3) to have a solution. Equations (1) and (3) are generically referred to as stochastic differential equations.

Given this background, we now return to the original objective of modeling the behavior of a stock's price process.

4 A Continuous-Time Model of Stock Price Behavior

Let us now restrict our attention only to those stocks which pay no dividends, so that stock returns are exclusively determined by price changes. Our basic discrete time model formulation is:

$$\ln q^e(t + \Delta t) - \ln q^e(t) = \mu\Delta t + \sigma\tilde{\varepsilon}\sqrt{\Delta t}. \quad (4)$$

Notice that the stochastic process is imposed on differences in the logarithm of the stock's price. Equation (4) thus asserts that the continuously compounded return to the ownership of the stock over the time period t to $t + \Delta t$ is distributed normally with mean $\mu\Delta t$ and variance $\sigma^2\Delta t$.

This is clearly a lognormal model:

$$\ln(q^e(t + \Delta t)) \sim N\left(\ln q^e(t) + \mu\Delta t, \sigma\tilde{\varepsilon}\sqrt{\Delta t}\right).$$

It is a more general formulation than a pure random walk as it admits the possibility that the mean increase in the logarithm of the price is positive. The continuous time analogue of (4) is

$$d\ln q^e(t) = \mu dt + \sigma dz. \quad (5)$$

Following (2), it has the solution

$$\ln q^e(t) = \ln q^e(0) + \mu t + \sigma z(t), \quad (6)$$

where

$$\begin{aligned} E \ln q^e(t) &= \ln q^e(0) + \mu t, \text{ and} \\ \text{var } q^e(t) &= \sigma^2 t \end{aligned}$$

Since $\ln q^e(t)$ on average grows linearly with t (so that, on average, $q^e(t)$ will grow exponentially), Equations (5) and (6) are, together, referred to as a geometric Brownian motion (GBM). It is clearly a lognormal process: $\ln q^e(t) \sim N(\ln q^e(0) + \mu t, \sigma\sqrt{t})$. The webcomplement entitled "Review of Basic Options Concepts and Terminology" illustrates how the parameters μ and σ can be estimated under the maintained assumption that time is measured in years.

While Equation (6) is a complete description of the evolution of the logarithm of a stock's price, we are rather interested in the evolution of the price

itself. Passing from a continuous time process on $\ln q^e(t)$ to one on $q^e(t)$ is not a trivial matter, however, and we need some additional background to make the conversion correctly. This is considered in the next few paragraphs.

The essence of lognormality is the idea that if a random variable \tilde{y} is distributed normally, then the random variable $\tilde{w} = e^{\tilde{y}}$ is distributed lognormally. Suppose, in particular, that $\tilde{y} \sim N(\mu_y, \sigma_y)$. A natural question is: how are μ_w and σ_w related to μ_y and σ_y when $\tilde{w} = e^{\tilde{y}}$? We first note that

$$\mu_w \neq e^{\mu_y}, \text{ and } \sigma_w \neq e^{\sigma_y}.$$

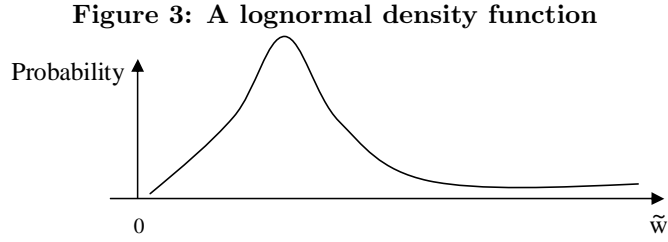
Rather, it can be shown that

$$\mu_w = e^{\mu_y + 1/2\sigma_y^2} \quad (7)$$

and

$$\sigma_w = e^{\mu_y + 1/2\sigma_y^2} \left(e^{\sigma_y^2} - 1 \right)^{1/2}. \quad (8)$$

These formulae are not obvious, but we can at least shed some light on (7). Why the variance of \tilde{y} should have an impact on the mean of \tilde{w} . To see why this is so, let us remind ourselves of the shape of the lognormal probability density function as found in Figure 3



Suppose there is an increase in variance. Since this distribution is pinched off to the left at zero, a higher variance can only imply (within the same class of distributions) that probability is principally shifted to higher values of \tilde{w} . But this will have the simultaneous effect of increasing the mean of \tilde{w} . The variance of \tilde{y} and the mean of \tilde{y} cannot be specified independently. The mean and standard deviation of the lognormal variable \tilde{w} are thus each related to both the mean and variance of \tilde{y} as per the relationships in Equations (7) and (8).

These results allow us to express the mean and standard deviation of $q^e(t)$ (by analogy, \tilde{w}) in relation to $\ln q^e(t) + \mu t$ and $\sigma^2 t$ (by analogy, the mean and variance of \tilde{y}) via Equations (5) and (6):

$$Eq^e(t) = e^{\ln q^e(0) + \mu t + \frac{1}{2}\sigma^2 t} = q^e(0)e^{\mu t + \frac{1}{2}\sigma^2 t} \quad (9)$$

$$\begin{aligned} \text{s.d.} q^e(t) &= e^{\ln q^e(0) + \mu t + \frac{1}{2}\sigma^2 t} \left(e^{\sigma^2 t} - 1 \right)^{\frac{1}{2}} \\ &= q^e(0)e^{\mu t + \frac{1}{2}\sigma^2 t} \left(e^{\sigma^2 t} - 1 \right)^{\frac{1}{2}}. \end{aligned} \quad (10)$$

We are now in a position, at least at an intuitive level, to pass from a stochastic differential equation describing the behavior of $\ln q^e(t)$ to one that governs the behavior of $q^e(t)$. If $\ln q^e(t)$ is governed by Equation (5), then

$$\frac{dq^e(t)}{q^e(t)} = \left(\mu + \frac{1}{2}\sigma^2 \right) dt + \sigma dz(t) \quad (11)$$

where $\frac{dq^e(t)}{q^e(t)}$ can be interpreted as the instantaneous (stochastic) rate of price change. Rewriting Equation (11) slightly differently yields

$$dq^e(t) = \left(\mu + \frac{1}{2}\sigma^2 \right) q^e(t)dt + \sigma q^e(t)dz(t) \quad (12)$$

which informs us that the stochastic differential equation governing the stocks price represents an Ito process since the coefficients of dt and $dz(t)$ are both time dependent.

We would also expect that if $q^e(t)$ were governed by

$$dq^e(t) = \mu q^e(t)dt + \sigma q^e(t)dz(t), \text{ then} \quad (13)$$

$$d \ln q^e(t) = \left(\mu - \frac{1}{2}\sigma^2 \right) dt + \sigma dz(t). \quad (14)$$

Equations (13) and (14) are fundamental to what follows.

5 Simulation and Call Pricing

5.1 Ito Processes

Ito Processes and their constituents, most especially the Brownian motion, are difficult to grasp at this abstract level and it will assist our intuition to describe how we might simulate a discrete time approximation to them.

Suppose we have estimated $\hat{\mu}$ and $\hat{\sigma}$ for a stock's price process as suggested in the webcomplement "Review of Options..". Recall that these estimates are derived from daily price data properly scaled up to reflect the fact that in this literature it is customary to measure time in years. We have two potential stochastic differential equations to guide us – Equations (13) and (14) – and each has a discrete time approximate counterpart.

(i) Discrete Time Counterpart to Equation (13)

If we approximate the stochastic differential $dq^e(t)$ by the change in the stock's price over a short interval of time Δt we have,

$$\begin{aligned} q^e(t + \Delta t) - q^e(t) &= \hat{\mu}q^e(t)\Delta t + \hat{\sigma}q^e(t)\tilde{\varepsilon}(t)\sqrt{\Delta t}, \text{ or} \\ q^e(t + \Delta t) &= q^e(t) \left[1 + \hat{\mu}\Delta t + \hat{\sigma}\tilde{\varepsilon}(t)\sqrt{\Delta t} \right] \end{aligned} \quad (15)$$

There is a problem with this representation, however, because for any $q^e(t)$, the price next "period," $q^e(t + \Delta t)$, is normally distributed (recall that $\tilde{\varepsilon}(t) \sim N(0, 1)$) rather than lognormal as a correct match to the data requires. In particular, there is the unfortunate possibility that the price could go negative, although for small time intervals Δt , this is exceedingly unlikely.

(ii) Discrete Time Counterpart to Equation (14)

Approximating $d \ln q^e(t)$ by successive log values of the price over small time intervals Δt yields

$$\begin{aligned} \ln q^e(t + \Delta t) - \ln q^e(t) &= \left(\hat{\mu} - \frac{1}{2}\hat{\sigma}^2 \right) \Delta t + \hat{\sigma}\tilde{\varepsilon}(t)\sqrt{\Delta t}, \text{ or} \\ \ln q^e(t + \Delta t) &= \ln q^e(t) + \left(\hat{\mu} - \frac{1}{2}\hat{\sigma}^2 \right) \Delta t + \hat{\sigma}\tilde{\varepsilon}\sqrt{\Delta t}. \end{aligned} \quad (16)$$

Here it is the logarithm of the price in period $t + \Delta t$ that is normally distributed, as required, and for this reason we'll limit ourselves to (16) and its successors. For simulation purposes, it is convenient to express equation (16) as

$$q^e(t + \Delta t) = q^e(t)e^{(\hat{\mu} - \frac{1}{2}\hat{\sigma}^2)\Delta t + \hat{\sigma}\tilde{\varepsilon}(t)\sqrt{\Delta t}}. \quad (17)$$

It is easy to generate a possible sample path of price realizations for (17). First select an interval of time Δt , and the number of successive time periods of interest (this will be the length of the sample path), say N . Using a random number generator, next generate N successive draws from the standard normal distribution. By construction, these draws are independent and thus successive rates of return $\left(\frac{q^e(t + \Delta t)}{q^e(t)} - 1 \right)$ will be statistically independent of one another.

Let this series of N draws be represented by $\{\varepsilon_j\}_{j=1}^N$. The corresponding sample path (or "time series") of prices is thus created as per Equation (18)

$$q^e(t_{j+1}) = q^e(t_j)e^{(\hat{\mu} - \frac{1}{2}\hat{\sigma}^2)\Delta t + \hat{\sigma}\varepsilon_j\sqrt{\Delta t}}, \quad (18)$$

where $t_{j+1} = t_j + \Delta t$.

This is not the price path that would be used for derivatives pricing, however.

5.2 The binomial model

Under the binomial model, call valuation is undertaken in a context where the probabilities have been changed in such a way so that all assets, including the underlying stock, earn the risk free rate. The simulation-based counterpart to this transformation is to replace $\hat{\mu}$ by $\ln(1 + r_f)$ in Equations (17) and (18):

$$q^e(t + \Delta t) = q^e(t)e^{(\ln(1 + r_f) - \frac{1}{2}\hat{\sigma}^2)\Delta t + \hat{\sigma}\tilde{\varepsilon}(t)\sqrt{\Delta t}}, \quad (19)$$

where r_f is the one year risk free rate (not continuously compounded) and $\ln(1 + r_f)$ is its continuously compounded counterpart.

How would we proceed to price a call in this simulation context? Since the value of the call at expiration is exclusively determined by the value of the underlying asset at that time, we first need a representative number of possible “risk neutral prices” for the underlying asset at expiration. The entire risk neutral sample path - as per equation (18) - is not required. By “representative” we mean enough prices so that their collective distribution is approximately lognormal. Suppose it was resolved to create J sample prices (to be even reasonably accurate $J \geq 1000$) at expiration, T years from now. Given random draws $\{\varepsilon_k\}_{k=1}^J$ from $N(0, 1)$, the corresponding underlying stock price realizations are $\{q_k^e(T)\}_{k=1}^J$ as given by

$$q_k^e(T) = q^e(0)e^{(\ln(1+r_f) - \frac{1}{2}\sigma^2)T + \sigma\varepsilon_k\sqrt{\Delta T}} \quad (20)$$

For each of these prices, the corresponding call value at expiration is

$$C_k^T = \max \{0, q_k^e(T) - E\}, \quad k = 1, 2, \dots, J.$$

The average expected payoff across all these possibilities is

$$C_{Avg}^T = \frac{1}{J} \sum_{k=1}^J C_k^T.$$

Since under risk neutral valuation the expected payoff of any derivative asset in the span of the underlying stock and a risk free bond is discounted back at the risk free rate, our estimate of the calls value today (when the stock's price is $q^e(0)$) is

$$C^0 = e^{-\ln(1+r_f)T} C_{Avg}^T. \quad (21)$$

In the case of an Asian option or some other path dependent option, a large number of sample paths must be generated since the exercise price of the option (and thus its value at expiration) is dependent upon the entire sample path of underlying asset prices leading to it.

Monte Carlo simulation is, as the above method is called, not the only pricing technique where the underlying idea is related to the notion of risk neutral valuation. There are ways that stochastic differential equations can be solved directly.

6 Solving Stochastic Differential Equations: A First Approach

Monte Carlo simulation employs the notion of risk neutral valuation but it does not provide closed form solutions for derivatives prices, such as the Black Scholes

formula in the case of calls ⁴. How are such closed form expressions obtained? In what follows we provide a non-technical outline of the first of two available methods. The context will once again be European call valuation where the underlying stock pays no dividends.

The idea is to obtain a partial differential equation whose solution, given the appropriate boundary condition is the price of the call. This approach is due to Black and Scholes (1973) and, in a more general context, Merton (1973). The latter author's arguments will guide our discussion here.

In the same spirit as the replicating portfolio approach mentioned in Section 4, Merton (1973) noticed that the payoff to a call can be represented in continuous time by a portfolio of the underlying stock and a risk free bond whose quantities are continuously adjusted. Given the stochastic differential equation which governs the stock's price (13) and another non stochastic differential equation governing the bond's price evolution, it becomes possible to construct the stochastic differential equation governing the value of the replicating portfolio. This latter transformation is accomplished via an important theorem which is referred to in the literature as Ito's Lemma. Using results from the stochastic calculus, this expression can be shown to imply that the value of the replicating portfolio must satisfy a particular partial differential equation. Together with the appropriate boundary condition (e.g., that $C(T) = \max\{q^e(T) - E, 0\}$), this partial differential equation has a known solution – the Black Scholes formula.

In what follows we begin first with a brief overview of this first approach; this is accomplished in three steps.

6.1 The Behavior of Stochastic Differentials.

In order to motivate what follows, we need to get a better idea of what the object $dz(t)$ means. It is clearly a random variable of some sort. We first explore its moments. Formally, $dz(t)$ is

$$\lim_{\Delta t \rightarrow 0} z(t + \Delta t) - z(t), \quad (22)$$

where we will not attempt to be precise as to how the limit is taken. We are reminded, however, that

$$\begin{aligned} E[z(t + \Delta t) - z(t)] &= 0, \text{ and} \\ \text{var}[z(t + \Delta t) - z(t)] &= (\sqrt{\Delta t})^2 = \Delta t, \text{ for all } \Delta t. \end{aligned}$$

It is not entirely surprising, therefore that

$$E(dz(t)) \equiv \lim_{\Delta t \rightarrow 0} E[z(t + \Delta t) - z(t)] = 0, \text{ and} \quad (23)$$

$$\text{var}(dz(t)) = \lim_{\Delta t \rightarrow 0} E[(z(t + \Delta t) - z(t))^2] = dt. \quad (24)$$

⁴The estimate obtained using Monte Carlo simulation will coincide with the Black Scholes value to a high degree of precision, however, if the number of simulated underlying stock prices is large ($\geq 10,000$) and the parameters r_f , E , σ , T are identical.

The object $dz(t)$ may thus be viewed as denoting an infinitesimal random variable with zero mean and variance dt (very small, but we are in a world of infinitesimals).

There are several other useful relationships:

$$E(dz(t) dz(t)) \equiv \text{var}(dz(t)) = dt \quad (25)$$

$$\begin{aligned} \text{var}(dz(t) dz(t)) &= \lim_{\Delta t \rightarrow 0} E \left[(z(t + \Delta t) - z(t))^4 - (\Delta t)^2 \right] \\ &\approx 0 \end{aligned} \quad (26)$$

$$E(dz(t) dt) = \lim_{\Delta t \rightarrow 0} E[(z(t + \Delta t) - z(t)) \Delta t] = 0 \quad (27)$$

$$\text{var}(dz(t) dt) = \lim_{\Delta t \rightarrow 0} E \left[(z(t + \Delta t) - z(t))^2 (\Delta t)^2 \right] \approx 0 \quad (28)$$

Equation (28) and (26) imply, respectively, that (25) and (??) are not only satisfied in expectation but with equality. Expression (25) is, in particular, quite surprising, as it argues that the square of a Brownian motion random process is effectively deterministic.

These results are frequently summarized as in Table 2 where $(dt)^2$ is negligible in the sense that it is very much smaller than dt and we may treat it as zero.

Table 2: The Product of Stochastic Differentials

	dz	dt
dz	dt	0
dt	0	0

The power of these results is apparent if we explore their implications for the computation of a quantity such as $(dq^e(t))^2$:

$$\begin{aligned} (dq^e(t))^2 &= (\mu dt + \sigma dz(t))^2 \\ &= \mu^2 (dt)^2 + 2\mu \sigma dt dz(t) + \sigma^2 (dz(t))^2 \\ &= \sigma^2 dt, \end{aligned}$$

since, by the results in Table 2, $(dt)(dt) = 0$ and $dt dz(t) = 0$.

The object $dq^e(t)$ thus behaves in the manner of a random walk in that its variance is proportional to the length of the time interval.

We will use these results in the context of Ito's lemma.

6.2 Ito's Lemma

A statement of this fundamental result is outlined below.

Theorem (Ito's Lemma).

Consider an Ito process $dx(t)$ of form $dx(t) = a(x(t), t)dt + b(x(t), t)dz(t)$, where $dz(t)$ is a Brownian motion, and consider a process $y(t) = F(x(t), t)$. Under quite general conditions $y(t)$ satisfies the stochastic differential equation

$$dy(t) = \frac{\partial F}{\partial x} dx(t) + \frac{\partial F}{\partial t} dt + \frac{1}{2} \frac{\partial^2 F}{\partial x^2} (dx(t))^2. \quad (29)$$

The presence of the right most term (which would be absent in a standard differential equation) is due to the unique properties of a stochastic differential equation. Taking advantage of results (in Table 2) let us specialize Equation (29) to the standard Ito process, where for notational simplicity, we suppress the dependence of coefficients $a(\cdot)$ and $b(\cdot)$ on $x(t)$ and t :

$$\begin{aligned} dy(t) &= \frac{\partial F}{\partial x} (adt + b dz(t)) + \frac{\partial F}{\partial t} dt + \frac{1}{2} \frac{\partial^2 F}{\partial x^2} (adt + b dz(t))^2 \\ &= \frac{\partial F}{\partial x} adt + \frac{\partial F}{\partial x} b dz(t) + \frac{\partial F}{\partial t} dt + \frac{1}{2} \frac{\partial^2 F}{\partial x^2} (a^2(dt)^2 + abdt dz(t) + b^2(dz(t))^2) \end{aligned}$$

Note that $(dt)^2 = 0$, $dt dz(t) = 0$, and $(dz(t))^2 = dt$.

Making these substitutions and collecting terms gives

$$dy(t) = \left(\frac{\partial F}{\partial x} a + \frac{\partial F}{\partial t} + \frac{1}{2} \frac{\partial^2 F}{\partial x^2} b^2 \right) dt + \frac{\partial F}{\partial x} b dz(t). \quad (30)$$

As a simple application, let us take as given

$$dq^e(t) = \mu q^e(t)dt + \sigma q^e(t)dz(t),$$

and attempt to derive the relationship for $d \ln q^e(t)$.

Here we have $a(q^e(t), t) \equiv \mu q^e(t)$, $b(q^e(t), t) \equiv \sigma q^e(t)$, and

$$\frac{\partial F}{\partial q^e(t)} = \frac{1}{q^e(t)}, \text{ and } \frac{\partial^2 F}{\partial q^e(t)^2} = -\frac{1}{q^e(t)^2}.$$

Lastly $\frac{\partial F}{\partial t} = 0$.

Substituting these results into Equation (30) yields.

$$\begin{aligned} d \ln q^e(t) &= \left[\frac{1}{q^e(t)} \mu q^e(t) + 0 + \frac{1}{2} (-1) \left(\frac{1}{q^e(t)} \right)^2 (\sigma q^e(t))^2 \right] dt + \frac{1}{q^e(t)} \sigma q^e(t) dz(t) \\ &= \left(\mu - \frac{1}{2} \sigma^2 \right) dt + \sigma dz(t), \end{aligned}$$

as was observed earlier.

This is the background.

6.3 The Black Scholes Formula

Merton (1973) requires four assumptions:

1. There are no market imperfections (perfect competition), transactions costs, taxes, short sales constraints or any other impediment to the continuous trading of securities.
2. There is unlimited riskless borrowing and lending at the constant risk free rate. If $q^b(t)$ is the period t price of a discount bond, then $q^b(t)$ is governed by the differential equation

$$\begin{aligned} dq^b(t) &= r_f q^b(t) dt, \text{ or} \\ B(t) &= B(0) e^{r_f t}; \end{aligned}$$

3. The underlying stock's price dynamics are given by a geometric Brownian motion of the form

$$\begin{aligned} dq^e(t) &= \mu q^e(t) dt + \sigma q^e(t) dz(t), \\ q^e(0) &> 0; \end{aligned}$$

4. There are no arbitrage opportunities across the financial markets in which the call, the underlying stock and the discount bond are traded.

Attention is restricted to call pricing formulae which are functions only of the stock's price currently and the time (so, e.g., the possibility of past stock price dependence is ignored); that is,

$$C = C(q^0(t), t).$$

By a straightforward application of Ito's lemma the call's price dynamics must be given by

$$dC = \left[\mu q^e(t) \frac{\partial C}{\partial q^e(t)} + \frac{\partial C}{\partial t} + \frac{\sigma^2}{2} \frac{\partial^2 C}{\partial q^e(t)^2} \right] dt + \sigma q^e(t) \frac{\partial C}{\partial q^e(t)} dz(t),$$

which is of limited help since the form of $C(q^e(t), t)$ is precisely what is not known. The partials with respect to $q^e(t)$ and t of $C(q^e(t), t)$ must be somehow circumvented.

Following the replicating portfolio approach, Merton (1973) defines the value of the call in terms of the self financing continuously adjustable portfolio P composed of $\Delta(q^e(t), t)$ shares and $N(q^e(t), t)$ risk free discount bonds:

$$V(q^e(t), t) = \Delta(q^e(t), t) q^e(t) + N(q^e(t), t) q^b(t). \quad (31)$$

By a straightforward application of Ito's lemma, the value of the portfolio must evolve according to (suppressing functional dependence in order to reduce the burdensome notation):

$$dV = \Delta dq^e + N dq^b + d\Delta q^e + dN q^b + (d\Delta) dq^e \quad (32)$$

Since $V(\cdot)$ is assumed to be self-financing, any change in its value can only be due to changes in the values of the constituent assets and not in the numbers of them. Thus it must be that

$$dV = \Delta dq^e + N dq^b, \quad (33)$$

which implies that the remaining terms in Equation (33) are identically zero:

$$d\Delta q^e + dN q^b + (d\Delta) dq^e \equiv 0. \quad (34)$$

But both $\Delta(\cdot)$ and $N(\cdot)$ are functions of $q^e(t)$, and t and thus Ito's lemma can be applied to represent their evolution in terms of $dz(t)$ and dt . Using the relationships of Table 2, and collecting terms, both those preceding $dz(t)$ and those preceding dt must individually be zero.

Together these relationships imply that the value of the portfolio must satisfy the following partial differential equation:

$$\frac{1}{2} \sigma^2 q^e V_{q^e q^e} + r_f q^e V_{q^e} + V_t = r_f V, \quad (35)$$

which has as its solution the Black Scholes formula when coupled with the terminal condition $V(q^e(T), T) = \max[0, q^e(T) - E]$.

7 A Second Approach: Martingale Methods

This method originated in the work of Harrison and Kreps (1979). It is popular as a methodology because it frequently allows for simpler computations than in the PDE approach. The underlying mathematics, however, is very complex and beyond the scope of this book. In order to convey a sense of what is going on, we present a brief heuristic argument that relies on the binomial abstraction.

Recall that in the binomial model, we undertook our pricing in a tree context where the underlying asset's price process had been modified. In particular, the true probabilities of the "up" and "down" state were replaced by the corresponding risk neutral probabilities. All assets (including the underlying stock) displayed an expected return equal to the risk free rate in the transformed setting.

Under geometric Brownian motion, the underlying price process is represented by an Ito stochastic differential equation of the form

$$dq^e(t) = \mu q^e(t) dt + \sigma q^e(t) dz(t). \quad (36)$$

In order to transform this price process into a risk neutral setting, two changes must be made.

1. The expression μ defines the mean and it must be replaced by r_f ; Only with this substitution will the mean return on the underlying stock become r_f . Note that r_f denotes the corresponding continuously compounded risk free rate.

2. The standard Brownian motion process must be modified. In particular, we replace dz by dz^* , where the two processes are related via the transformation:

$$dz^*(t) = dz(t) + \left(\frac{\mu - r_f}{\sigma} \right) dt$$

The transformed price process is thus

$$dq^e(t) = r_f q^e(t) dt + \sigma q^e(t) dz^*(t). \quad (37)$$

By Equation (14) the corresponding process on $\ln q^e(t)$ is

$$d \ln q^e(t) = \left(r_f - \frac{1}{2} \sigma^2 \right) dt + \sigma dz^*(t) \quad (38)$$

Let T denote the expiration date of a simple European call option. In the same spirit as the binomial model the price of a call must be the present value of its expected payoff at expiration under the transformed process.

Equation (38) informs us that in the transformed economy,

$$\ln \left(\frac{q^e(t)}{q^e(0)} \right) \sim N \left(\left(r_f - \frac{1}{2} \sigma^2 \right) T, \sigma^2 T \right). \quad (39)$$

Since, in the transformed economy,

$$\text{prob}(q^e(t) \geq E) = \text{prob}(\ln q^e(t) \geq \ln E),$$

we can compute the call's value using the probability density implied by Equation (39):

$$C = e^{-r_f T} \int_{\ln E}^{\infty} (e^s - E) f(s) ds,$$

where $f(s)$ is the probability density on the \ln of the stock's price.

Making the appropriate substitutions yields

$$C = e^{-r_f T} \left(\frac{1}{\sqrt{2\pi\sigma^2 T}} \right) \int_{\ln K}^{\infty} (e^s - K) e^{\frac{-(s - \ln q^e(t) - r_f T + \frac{\sigma^2 T}{2})}{2\sigma^2 T}} ds \quad (40)$$

which, when the integration is performed yields the Black Scholes Formula.

8 Applications

We make reference to a number that have been considered earlier in the text.

8.1 The Consumption-Savings Problem.

This is a classic economic problem and we considered it fairly thoroughly in Chapter 5. Without the requisite math background there is not a lot we can say about the continuous time analogue than to set up the problem, but even that first step will be helpful.

Suppose the risky portfolio (“ M ”) is governed by the following price process
 $dq^M(t) = q^M(t)[\mu_M dt + \sigma_M dz(t)]$,
 $q^M(0)$ given, and the risk free asset by
 $dq^b(t) = r q^b(t) dt$, $q^b(0)$ given.

If an investor has initial wealth $Y(0)$, and chooses to invest the proportion $w(t)$ (possibly continuously varying) in the risky portfolio, then his wealth $Y(t)$ will evolve according to

$$dY(t) = Y(t)[w(t)(\mu_M - r_t) + r_f]dt + Y(t)[\pi(t)\sigma dz(t)] - c(t)dt, \quad (41)$$

where $c(t)$ is his consumption path. With objective function

$$\max_{c(t), w(t)} E \int_0^T e^{-\gamma t} U(c(t)) dt, \quad (42)$$

the investors ‘problem’ is one of maximizing Equation (42) subject to Equation (41) and initial conditions on wealth, and the constraint that $Y(t) \geq 0$ for all t .

A classic result allows us to transform this problem to one that it turns out can be solved much more easily:

$$\begin{aligned} \max_{c(t), w(t)} E \int_0^T e^{-\gamma t} u(c(t)) dt \\ \text{s.t. } PV_0(c(t)) = E^* \int_0^T e^{-r_f t} c(t) dt \leq Y(0) \end{aligned} \quad (43)$$

where E^* is the transformed risk neutral measure under which the investor’s wealth follows a Brownian motion process.

In what we have presented above, all the notation is directly analogous to that of Chapter 5: $U(\cdot)$ is the investors utility of (instantaneous) consumption, γ his (instantaneous) subjective discount rate, and T his time horizon,

8.2 An Application to Portfolio Analysis.

Here we hope to give a hint of how to extend the portfolio analysis of Chapters 5 or 6 to a setting where trading is (hypothetically) continuous and individual security returns follow geometric Brownian motions.

Let there be $i = 1, 2, \dots, N$ equity securities, each of whose return is governed by the process

$$\frac{dq_i^e(t)}{q_i^e(t)} = \mu_i dt + \sigma_i dz(t) \quad (44)$$

where $\sigma_i > 0$. These processes may also be correlated with one another in a manner that we can represent precisely. Conducting a portfolio analysis in this setting has been found to have two principal advantages. First, it provides new insights concerning the implications of diversification for long run portfolio returns, and, second, it allows for an easier solution to certain classes of problems. We will note these advantages with the implicit understanding that the derived portfolio rules must be viewed as guides for practical applications. Literally interpreted, they will imply, for example, continuous portfolio rebalancing – at an unbounded total expense, if the cost of doing each rebalancing is positive – which is absurd. In practice one would rather employ them weekly or perhaps daily.

The stated objective will be to maximize the expected rate of appreciation of a portfolio's value; equivalently, to maximize its expected terminal value which is the terminal wealth of the investor who owns it. Most portfolio managers would be familiar with this goal.

To get an idea of what this simplest criterion implies, and to make it more plausible in our setting, we first consider the discrete time equivalent (and, by implication) the discrete time approximation to GBM.

8.3 Digression to Discrete Time

Suppose a CRRA investor has initial wealth $Y(0)$ at time $t = 0$ and is considering investing in any or all of a set of non-dividend paying stocks whose returns are iid. Since the rate of expected appreciation of the portfolio is its expected rate of return, and since the return distributions of the available assets are iid, the investors' optional portfolio proportions will be invariant to the level of his wealth, and the distribution of his portfolio's returns will itself be iid. At the conclusion of his planning horizon, T periods from the present, the investors' wealth will be

$$Y_T = Y_0 \prod_{s=1}^T \tilde{R}_s^P, \quad (45)$$

where \tilde{R}_s^P denotes the (iid,) gross portfolio return in period s . It follows that

$$\begin{aligned} \ln \left(\frac{Y_T}{Y_0} \right) &= \sum_{s=1}^T \ln \tilde{R}_s^P, \text{ and} \\ \ln \left(\frac{Y_T}{Y_0} \right)^{\frac{1}{T}} &= \left(\frac{1}{T} \right) \sum_{s=1}^T \ln \tilde{R}_s^P. \end{aligned} \quad (46)$$

Note that whenever we introduce the \ln we effectively assume continuous compounding within the time period. As the number of periods in the time horizon grows without bound, $T \mapsto \infty$, by the Law of Large Numbers,

$$\left(\frac{Y_T}{Y_0}\right)^{\frac{1}{T}} \mapsto e^{E \ln \tilde{R}^P}, \text{ or} \quad (47)$$

$$Y_T \mapsto Y_0 e^{TE \ln \tilde{R}^P} \quad (48)$$

Consider an investor with a many period time horizon who wishes to maximize her expected terminal wealth under continuous compounding. The relationship in Equation (48) informs her that

(1) it is sufficient, under the aforementioned assumptions, for her to choose portfolio proportions which maximize $E \ln \tilde{R}^P$, the expected logarithm of the one period return, and

(2) by doing so the average growth in rate of her wealth will approach a deterministic limit.

Before returning to the continuous time setting let us also entertain a brief classic example, one in which an investor must decide what fractions of his wealth to assign to a highly risky stock and to a risk free asset (actually, the risk free asset is equivalent to keeping money in a shoebox under the bed). For an amount $Y(0)$ invested in either asset, the respective returns are found in Figure 4.

Figure 4: Two alternative Investment Returns



Let w represent the proportion in the stock, and notice that the expected gross return to either asset under continuous compounding is zero:

$$\text{Stock} : ER^e = \frac{1}{2} \ln(2) + \frac{1}{2} \ln\left(\frac{1}{2}\right) = 0$$

$$\text{Shoebox} : ER^{sb} = \frac{1}{2} \ln(1) + \frac{1}{2} \ln(1) = 0$$

With each asset paying the same expected return, and the stock being wildly risky, at first appearance the shoebox would seem to be the investment of choice. But according to Equation (48) the investor ought to allocate his wealth among the two assets so as to maximize the expected \ln of the portfolio's one-period gross return:

$$\max E \ln \tilde{R}^P = \max_w \left\{ \frac{1}{2} \ln(2w + (1-w)) + \frac{1}{2} \ln\left(\frac{1}{2}w + (1-w)\right) \right\}$$

A straight forward application of the calculus yields $w = 3/4$, with the consequent portfolio returns in each state as show in Figure 5

Figure 5 Optimal Portfolio Returns in Each State

$$\ln(1+w) = \ln(1.75) = 0.5596, \text{ prob.} = \frac{1}{2}$$

$$\ln(1 - \frac{1}{2} \ln w) = \ln(0.625) = -0.47, \text{ prob.} = \frac{1}{2}$$

As a result, $E \ln R^{\bar{P}} = 0.0448$ with an effective risk free period return (for a very long time horizon) of 4.5% ($e^{0.448} = 1.045$).

This result is surprising and the intuition is not obvious. Briefly, the optimal proportions of $w = 3/4$ and $1 - w = 1/4$ reflect the fact that by always keeping a fixed fraction of wealth in the risk free asset, the worst trajectories can be avoided. By “frequent” trading, although each asset has an expected return of zero, a combination will yield an expected return that is strictly positive, and over a long time horizon effectively riskless. Frequent trading expands market opportunities.

8.4 Return to Continuous Time

The previous setup applies directly to a continuous time setting as all of the fundamental assumptions are satisfied. In particular, there are a very large number of periods (an uncountable number, in fact) and the returns to the various securities are iid through time. Let us make the added generalization that the individual asset returns are correlated through their Brownian motion components. By an application of Ito’s lemma, we may write

$$\text{cov}(dz_i, dz_j) = E(dz_i(t)dz_j(t)) = \sigma_{ij}dt,$$

where σ_j denotes the (i, j) entry of the (instantaneous) variance – covariance matrix.

As has been our custom, denote the portfolio’s proportions for the N assets by w_1, \dots, w_N and let the superscript P denote the portfolio itself. As in earlier chapters, the process on the portfolio’s instantaneous rate of return, $\frac{dY^P(t)}{Y^P(t)}$ is the weighted average of the instantaneous constituent asset returns (as given in Equation (44)):

$$\begin{aligned} \frac{dY^P(t)}{Y^P(t)} &= \sum_{i=1}^N w_i \frac{dq_i^e(t)}{q_i^e(t)} = \sum_{i=1}^N w_i (\mu_i dt + dz_i(t)) \\ &= \left(\sum_{i=1}^N w_i \mu_i \right) dt + \sum_{i=1}^N w_i dz_i(t), \end{aligned} \quad (49)$$

where the variance of the stochastic term is given by

$$\begin{aligned} E \left(\sum_{i=1}^N w_i dz_i(t) \right)^2 &= E \left\{ \left(\sum_{i=1}^N w_i dz_i(t) \right) \left(\sum_{j=1}^N w_j dz_j(t) \right) \right\} \\ &= \left(\sum_{i=1}^N \sum_{j=1}^N w_i w_j \sigma_{ij} \right) dt \end{aligned}$$

Equation (49) describes the process on the portfolio’s rate of return and we see that it implies that the portfolio’s value, at any future time horizon T will

be lognormally distributed; furthermore, an uncountable infinity of periods will have passed. By analogy (and formally) our discrete time reflections suggest that an investor should in this context also choose portfolio proportion so as to maximize the mean growth rate, V_P , of the portfolio as given by

$$E \left\{ \ln \frac{Y^P(t)}{Y(0)} \right\} = T_{vp}$$

Since the portfolio's value itself follows a Brownian motion (with drift $\sum_{i=1}^N w_i \mu_i$ and disturbance $\sum_{i=1}^N w_i dz_i$),

$$E \left[\ln \frac{Y^P(t)}{Y(0)} \right] = \left(\sum_{i=1}^N w_i \mu_i \right) T - \frac{1}{2} \left(\sum_{i=1}^N \sum_{j=1}^N w_i w_j \sigma_{ij} \right) T, \text{ thus} \quad (50)$$

$$\nu_P = \left(\frac{1}{T} \right) E \left[\ln \frac{Y^P(t)}{Y(0)} \right] = \sum_{i=1}^N w_i \mu_i - \frac{1}{2} \sum_{i=1}^N \sum_{j=1}^N w_i w_j \sigma_{ij}. \quad (51)$$

The investor should choose portfolio proportions to maximize this latter quantity.

Without belaboring this development much further, it behooves us to recognize the message implicit in (51). This can be accomplished most straightforwardly in a context of an equally weighted portfolio where each of the N assets is independently distributed of one another ($\sigma_{ij} = 0$ for $i \neq j$), and all have the same mean and variance ($(\mu_i \sigma_i) = (\mu, \sigma)$, $i = 1, 2, \dots, N$).

In this case (51) reduces to

$$\nu_P = \mu - \left(\frac{1}{2N} \right) \sigma^2, \quad (52)$$

with the direct implication that the more identical stocks the investor adds to the portfolio the greater the mean instantaneous return. In this sense it is useful to search for many similarly volatile stocks whose returns are independent of one another: by combining them in a portfolio where we continually (frequently) rebalance to maintain equal proportions, not only will portfolio variance decline ($\frac{1}{2N} \sigma^2$), as in the discrete time case, but the mean return will rise (which is *not* the case in discrete time!).

8.5 The Consumption CAPM in Continuous Time

Our final application concerns the consumption CAPM of Chapter 9, and the question we address is this: What is the equilibrium asset price behavior in a Mehra-Prescott asset pricing context when the growth rate in consumption follows a GBM? Specializing preferences to be of the customary forms $U(c) = (c^{1-\gamma}/1-\gamma)$, pricing relationship (9.4) reduces to:

$$\begin{aligned}
P_t &= E_t \left\{ Y_t \sum_{j=1}^{\infty} \beta^j x_{t+j}^{1-\gamma} \right\} \\
&= Y_t \sum_{j=1}^{\infty} \beta^j E_t \left\{ x_{t+j}^{1-\gamma} \right\}
\end{aligned}$$

where x_{t+j} is the growth rate in output (equivalently, consumption in the Mehra-Prescott economy) from period j to period $j+1$.

We hypothesize that the growth rate x follows a GBM of the form

$$dx = \mu x dt + \sigma x dz,$$

where we interpret x_{t+j} as the discrete time realisation of $x(t)$ at time $t+j$.

One result from statistics is needed. Suppose w is lognormally distributed which we write $\tilde{w} \sim L(\xi, \eta)$ where $\xi = E \ln \tilde{w}$ and $\eta^2 = \text{var} \ln \tilde{w}$. Then for any real number q ,

$$E \{ w^q \} = e^{q\xi + \frac{1}{2} q^2 \eta^2}$$

By the process on the growth rate just assumed, $x(t) \sim L((\mu - \frac{1}{2}\sigma^2)t, \sigma\sqrt{t})$ so that at time $t+j$, $x_{t+j} \sim L((\mu - \frac{1}{2}\sigma^2)j, \sigma\sqrt{j})$. By this result,

$$\begin{aligned}
E \left\{ x_{t+j}^{1-\gamma} \right\} &= e^{(1-\gamma)(\mu - \frac{1}{2}\sigma^2)j + \frac{1}{2}(1-\gamma)^2\sigma^2 j} \\
&= e^{(1-\gamma)(\mu - \frac{1}{2}\gamma\sigma^2)j},
\end{aligned}$$

and thus,

$$\begin{aligned}
P_t &= Y_t \sum_{j=1}^{\infty} \beta^j e^{(1-\gamma)(\mu - \frac{1}{2}\gamma\sigma^2)j} \\
&= Y_t \sum_{j=1}^{\infty} \left(\beta e^{(1-\gamma)(\mu - \frac{1}{2}\gamma\sigma^2)} \right)^j
\end{aligned}$$

which is well defined (the sum has a finite value) if $\beta e^{(1-\gamma)(\mu - \frac{1}{2}\gamma\sigma^2)} < 1$, which we will assume to be the case. Then

$$P_t = Y_t \frac{\beta e^{(1-\gamma)(\mu - \frac{1}{2}\gamma\sigma^2)}}{1 - \beta e^{(1-\gamma)(\mu - \frac{1}{2}\gamma\sigma^2)}}$$

This is an illustration of the fact that working in continuous time often allows convenient closed form solutions. Our remark are taken from Mehra and Sah (2001).

9 Final Comments

There is much more to be said. There are many more extensions of the CCAPM style models to a continuous time setting. Another issue is the sense in which a continuous time price process (e.g. Equation (13)) can be viewed as an equilibrium price process in the sense of that concept as presented in this book. This remains a focus of research.

Continuous time is clearly different from discrete time, but does it use (as a derivatives pricing tool) enrich our economic understanding of the large financial and macroeconomics reality? That is less clear.

Intermediate Financial Theory
Danthine and Donaldson

Solutions to Exercises

Chapter 1

- 1.1. U is a utility function, i.e., $U(x) > U(y) \Leftrightarrow x \succ y$
 $f(\cdot)$ is an increasing monotone transformation, $f(a) > f(b) \Leftrightarrow a > b$;
 then $f(U(x)) > f(U(y)) \Leftrightarrow U(x) > U(y) \Leftrightarrow x \succ y$
- 1.2. Utility function $U(c_1, c_2)$:
 FOC: $U_1/U_2 = p_1/p_2$
 Let $f = f(U(\cdot))$ be a monotone transformation.
 Apply the chain rule for derivatives:
 FOC: $f_1/f_2 = f' U_1 / f' U_2 = p_1/p_2$ (prime denotes derivation).
 Economic interpretation: f and U represent the same preferences, they must lead to the same choices.
- 1.3. When an agent has very little of one given good, he is willing to give up a big quantity of another good to obtain a bit more of the first.
 MRS is constant when the utility function is linear additive (that is, the indifference curve is also linear):

$$U(c_1, c_2) = \alpha c_1 + \beta c_2$$

$$MRS = \frac{\alpha}{\beta}$$

Not very interesting; for example, the optimal choice over 2 goods for a consumer is always to consume one good only (if the slope of the budget line is different from the MRS) or an indefinite quantity of the 2 goods (if the slopes are equal).

Convex preferences can exhibit indifference curves with flat spots, strictly convex preferences cannot. The utility function is not strictly quasi-concave here.

Pareto set: 2 cases.

- Indifference curves for the agents have the same slope: Pareto set is the entire box;
- Indifference curves do not have the same slope: Pareto set is the lower side and the right side of the box, or the upper side and the left side, depending on which MRS is higher.

- 1.4. a. $U_1 = 6^{0.5} 4^{0.5} = 4.90$
 $U_2 = 14^{0.5} 16^{0.5} = 14.97$
 $MRS_j = \frac{\partial U_j / \partial c_1^j}{\partial U_j / \partial c_2^j} = \frac{\alpha c_2^j}{(1 - \alpha) c_1^j}$
 with $\alpha = 0.5$, $MRS_j = \frac{c_2^j}{c_1^j}$
 $MRS_1 = \frac{4}{6} = 0.67$

$$MRS_2 = \frac{16}{14} = 1.14$$

$MRS_1 \neq MRS_2$, not Pareto Optimal; it is possible to reallocate the goods and make one agent (at least) better off without hurting the other.

b. $PS = \{ c_1^j = c_2^j, j=1,2 : c_1^i + c_2^i = 20, i=1,2 \}$, the Pareto set is a straight line (diagonal from lower-left to upper-right corner).

c. The problem of the agents is

$$\text{Max } U^j \text{ s.t. } p_1 e_1^j + e_2^j = p_1 c_1^j + c_2^j.$$

The Lagrangian and the FOC's are given by

$$L^j = (c_1^j)^{1/2} (c_2^j)^{1/2} + y(p_1 e_1^j + e_2^j - p_1 c_1^j - c_2^j)$$

$$\frac{\partial L^j}{\partial c_1^j} = \frac{1}{2} \left(\frac{c_2^j}{c_1^j} \right)^{1/2} - y p_1 = 0$$

$$\frac{\partial L^j}{\partial c_2^j} = \frac{1}{2} \left(\frac{c_1^j}{c_2^j} \right)^{1/2} - y = 0$$

$$\frac{\partial L^j}{\partial y} = p_1 e_1^j + e_2^j - p_1 c_1^j - c_2^j = 0$$

Rearranging the FOC's leads to $p_1 = \frac{c_2^j}{c_1^j}$. Now we insert this ratio into the budget constraints of agent

1 $p_1 6 + 4 - 2p_1 c_1^1 = 0$ and after rearranging we get $c_1^1 = 3 + \frac{2}{p_1}$. This expression can be interpreted as

a demand function. The remaining demand functions can be obtained using the same steps.

$$c_2^1 = 3p_1 + 2$$

$$c_1^2 = 7 + \frac{8}{p_1}$$

$$c_2^2 = 7p_1 + 8$$

To determine market equilibrium, we use the market clearing condition $c_1^1 + c_1^2 = 20, c_2^1 + c_2^2 = 20$.

Finally we find $p_1 = 1$ and $c_1^1 = c_2^1 = 5, c_1^2 = c_2^2 = 15$.

The after-trade MRS and utility levels are:

$$U_1 = 5^{0.5} 5^{0.5} = 5$$

$$U_2 = 15^{0.5} 15^{0.5} = 15$$

$$MRS_1 = \frac{5}{5} = 1$$

$$MRS_2 = \frac{15}{15} = 1$$

Both agents have increased their utility level and their after-trade MRS is equalized.

d. $U_j(c_1^j, c_2^j) = \ln((c_1^j)^\alpha \cdot (c_2^j)^{1-\alpha}) = \alpha \ln c_1^j + (1-\alpha) \ln c_2^j$,

$$MRS_j = \frac{\partial U_j / \partial c_1^j}{\partial U_j / \partial c_2^j} = \frac{\alpha c_2^j}{(1-\alpha)c_1^j}$$

Same condition as that obtained in a). This is not a surprise since the new utility function is a monotone transformation (logarithm) of the utility function used originally.

$$U_1 = \ln(6^{0.5} 4^{0.5}) = 1.59$$

$$U_2 = \ln(14^{0.5} 16^{0.5}) = 2.71$$

MRS's are identical to those obtained in a), but utility levels are not. The agents will make the same maximizing choice with both utility functions, and the utility level has no real meaning, beyond the statement that for a given individual a higher utility level is better.

e. Since the maximizing conditions are the same as those obtained in a)-c) and the budget constraints are not altered, we know that the equilibrium allocations will be the same too (so is the price ratio). The after-trade MRS and utility levels are:

$$U_1 = \ln(5^{0.5} 5^{0.5}) = 1.61$$

$$U_2 = \ln(15^{0.5} 15^{0.5}) = 2.71$$

$$MRS_1 = \frac{5}{5} = 1$$

$$MRS_2 = \frac{15}{15} = 1$$

- 1.5. Recall that in equilibrium there should not be excess demand or excess supply for any good in the economy. If there is, then prices change accordingly to restore the equilibrium. The figure shows excess demand for good 2 and excess supply for good 1, a situation which requires p_2 to increase and p_1 to decrease to restore market clearing. This means that p_1/p_2 should decrease and the budget line should move counter-clockwise.

Chapter 3

3.1. Mathematical interpretation:

We can use Jensen's inequality, which states that if $f(\cdot)$ is concave, then

$$E(f(X)) \leq f(E(X))$$

Indeed, we have that

$$E(f(X)) = f(E(X)) \Leftrightarrow f'' = 0$$

As a result, when $f(\cdot)$ is not linear, the ranking of lotteries with the expected utility criterion might be altered.

Economic interpretation:

Under uncertainty, the important quantities are risk aversion coefficients, which depend on the first and second order derivatives. If we apply a non-linear transformation, these quantities are altered.

Indeed, $R_A(f(U(\cdot))) = R_A(U(\cdot)) \Leftrightarrow f$ is linear.

a. $L = (B, M, 0.50) = 0.50 \times U(B) + 0.50 \times U(M) = 55 > U(P) = 50$. Lottery L is preferred to the "sure lottery" P.

b. $f(U(X)) = a + b \times U(X)$

$$L_f = (B, M, 0.50)_f = 0.50 \times (a + bU(B)) + 0.50 \times (a + bU(M)) = a + b55 > f(U(P)) = a + bU(P) = a + b50.$$

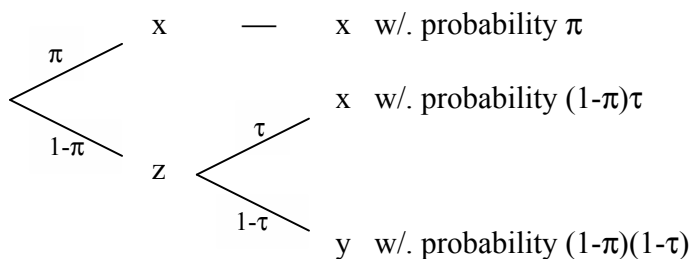
Again, L is preferred to P under transformation f.

$$g(U(X)) = \ln U(X)$$

$$L_g = (B, M, 0.50)_g = 0.50 \times \ln U(100) + 0.50 \times \ln U(10) = 3.46 < g(U(P)) = \ln U(50) = 3.91. P \text{ is preferred to } L \text{ under transformation } g.$$

3.2. Lotteries:

We show that $(x, z, \pi) = (x, y, \pi + (1-\pi)\tau)$ if $z = (x, y, \tau)$.



The total probabilities of the possible states are

$$\pi(x) = \pi + (1-\pi)\tau$$

$$\pi(y) = (1-\pi)(1-\tau)$$

Of course, $\pi(x) + \pi(y) = \pi + (1-\pi)\tau + (1-\pi)(1-\tau) = 1$. Hence we obtain lottery $(x, y, \pi + (1-\pi)\tau)$.

Could the two lotteries (x, z, π) and $(x, y, \pi + (1-\pi)\tau)$ with $z = (x, y, \tau)$ be viewed as non-equivalent? Yes, in a non-expected utility world where there is a preference for gambling. Yes, also, in a world where non-rational agents might be confused by the different contexts in which they are requested to make choices. While the situation represented by the two lotteries is too simple to make this plausible here, the behavioral finance literature building on the work of Kahneman and Tversky (see references in the text) point out that in more realistic experimental situations similar ‘confusions’ are frequent.

3.3 U is concave. By definition, for a concave function $f(\cdot)$

$$f(\lambda a + (1-\lambda)b) \geq \lambda f(a) + (1-\lambda)f(b), \lambda \in [0,1]$$

Use the definition with $f = U$, $a = c_1$, $b = c_2$, $\lambda = 1/2$

$$U\left(\frac{1}{2}c_1 + \frac{1}{2}c_2\right) \geq \frac{1}{2}U(c_1) + \frac{1}{2}U(c_2)$$

$$U(\bar{c}) \geq \frac{1}{2}U(c_1) + \frac{1}{2}U(c_2)$$

$$2U(\bar{c}) \geq U(c_1) + U(c_2)$$

$$V(\bar{c}, \bar{c}) \geq V(c_1, c_2)$$

Chapter 4

4.1 Risk Aversion: (Answers to a), b), c), and d) are given together here)

<p>(1) $U(Y) = -\frac{1}{Y}$ $U'(Y) = \frac{1}{Y^2} > 0$ $U''(Y) = -\frac{2}{Y^3} < 0$ $R_A = \frac{2}{Y}$ $R_R = 2$ $\frac{\partial R_A}{\partial Y} = -\frac{2}{Y^2} < 0$</p>	<p>(2) $U(Y) = \ln Y$ $U'(Y) = \frac{1}{Y} > 0$ $U''(Y) = -\frac{1}{Y^2} < 0$ $R_A = \frac{1}{Y}$ $R_R = 1$ $\frac{\partial R_A}{\partial Y} = -\frac{1}{Y^2} < 0$</p>	<p>(3) $U(Y) = -Y^{-\gamma}$ $U'(Y) = \gamma Y^{-\gamma-1} > 0 \Leftrightarrow \gamma > 0$ $U''(Y) = -\gamma(\gamma+1)Y^{-\gamma-2} < 0$ $R_A = \frac{\gamma+1}{Y}$ $R_R = \gamma+1$ $\frac{\partial R_A}{\partial Y} = -\frac{\gamma+1}{Y^2} < 0$</p>
<p>(4) $U(Y) = -\exp(-\gamma Y)$ $U'(Y) = \gamma \exp(-\gamma Y) > 0 \Leftrightarrow \gamma > 0$ $U''(Y) = -\gamma^2 \exp(-\gamma Y) < 0$ $R_A = \gamma$ $R_R = \gamma Y$ $\frac{\partial R_R}{\partial Y} = \gamma > 0$</p>	<p>(5) $U(Y) = \frac{Y^\gamma}{\gamma}$ $U'(Y) = Y^{\gamma-1}$ $U''(Y) = (\gamma-1)Y^{\gamma-2} < 0 \Leftrightarrow \gamma < 1$ $R_A = \frac{1-\gamma}{Y}$ $R_R = 1-\gamma$ $\frac{\partial R_A}{\partial Y} = \frac{\gamma-1}{Y^2} < 0$</p>	
<p>(6) $U(Y) = \alpha Y - \beta Y^2, \alpha > 0, \beta > 0$ $U'(Y) = \alpha - 2\beta Y > 0 \Leftrightarrow Y < \frac{\alpha}{2\beta}$ $U''(Y) = -2\beta < 0$ $R_A = \frac{2\beta}{\alpha - 2\beta Y} > 0$ $R_R = \frac{2\beta}{\alpha - 2\beta Y} Y > 0$ $\frac{\partial R_A}{\partial Y} = \frac{4\beta^2}{(\alpha - 2\beta Y)^2} > 0$ $\frac{\partial R_R}{\partial Y} = \frac{\partial}{\partial Y}(R_A Y) = \frac{\partial R_A}{\partial Y} Y + R_A > 0$</p>		

Note that γ controls for the degree of risk aversion. We check it with the derivative of R_A and R_R w.r.t. γ .

$$\begin{array}{lll}
U(Y) = -Y^{-\gamma} & \frac{\partial R_A}{\partial \gamma} = \frac{1}{Y} & \frac{\partial R_R}{\partial \gamma} = 1 \\
U(Y) = -\exp(-\gamma Y) & \frac{\partial R_A}{\partial \gamma} = 1 & \frac{\partial R_R}{\partial \gamma} = Y \\
U(Y) = \frac{Y^\gamma}{\gamma} & \frac{\partial R_A}{\partial \gamma} = -\frac{1}{Y} & \frac{\partial R_R}{\partial \gamma} = -1
\end{array}$$

In the last utility function above, we should better use $\gamma \equiv 1-\theta$, so that

$$U(Y) = \frac{Y^\gamma}{\gamma} \equiv \frac{Y^{1-\theta}}{1-\theta}, \text{ and } \frac{\partial R_A}{\partial \theta} = \frac{1}{Y}, \frac{\partial R_R}{\partial \theta} = 1 \text{ (look at } R_R = 1-\gamma = \theta \text{). After this change, every}$$

derivative w.r.t. θ is positive. If we increase θ , we increase the level of risk aversion (both absolute and relative).

4.2. Certainty equivalent .

The problem to be solved is: find x such that $\pi_1 U(Y_1^1) + \pi_2 U(Y_2^1) \equiv U(x)$

where Y_i^1 denotes outcome of lottery L_1 in state i and π_i denotes the probability of state i .

If U is bijective, it can be "inverted", so the solution is

$$x = U^{-1}\{\pi_1 U(Y_1^1) + \pi_2 U(Y_2^1)\}$$

where U^{-1} is the inverse function of U .

$$(1) \quad U(Y) = -\frac{1}{Y}, U^{-1}(Y) = -\frac{1}{Y}$$

$$x = -\left(\frac{1}{2}\left(-\frac{1}{50000} - \frac{1}{10000}\right)\right)^{-1} = 16666.67$$

$$(2) \quad U(Y) = \ln Y, U^{-1}(Y) = \exp(Y)$$

$$x = \exp\left(\frac{1}{2}\left(\ln \frac{1}{50000} + \ln \frac{1}{10000}\right)\right) = 22360.68$$

$$(3) \quad U(Y) = \frac{Y^\gamma}{\gamma}, U^{-1}(Y) = (\gamma Y)^{1/\gamma}$$

$$x = \left(\gamma \left(\frac{1}{2}\left(\frac{Y_1^\gamma}{\gamma} + \frac{Y_2^\gamma}{\gamma}\right)\right)\right)^{1/\gamma}$$

with $\gamma = .25$, $x = 24232.88$

with $\gamma = .75$, $x = 28125.91$

Increasing γ leads to an increase in the value of x . This is because $1-\gamma$ (and not γ !) is the coefficient of relative risk aversion for this utility function. Therefore, increasing γ decreases the level of risk aversion, and the certainty equivalent is higher, i.e. the risk premium is lower.

4.3. Risk premium .

The problem to be solved (indifference between insurance and no insurance) is

$$EU(Y) = \sum_i \ln Y_i \pi(i) = \ln(100,000 - P)$$

where P is the insurance premium, Y_i is the worth in state i and $\pi(i)$ is the probability of state i .

The solution to the problem is

$$P = 100,000 - \exp(EU(Y))$$

The solutions under the three scenarios are

$$\text{Scenario A: } P = 13312.04$$

$$\text{Scenario B: } P = 13910.83$$

$$\text{Scenario C: } P = 22739.27$$

Starting from scenario A, in scenario B and C we have transferred 1 percent from state 3 to another state (to state 2 in scenario B and to state 1 in scenario C). However, the outcome is very different: The premium is slightly bigger in scenario B, while it is a lot higher in C. This could have been expected because the logarithmic utility function is very curved at low values, and flattens out rapidly, i.e. $\ln 1$ is very different from $\ln 100000$, but $\ln 50000$ is only slightly different from $\ln 100000$. Also, logarithmic utility function is DARA.

4.4. Simply take the expected utility (initial wealth is normalized to 1)

$$E(U(1 + \tilde{r}_A)) \equiv E(U(1 + \tilde{r}_B + \tilde{\xi})) \geq E(U(1 + \tilde{r}_B))$$

and apply Theorem 3.2. All individuals with increasing utility functions prefer A to B.

4.5 a. Let \tilde{x}_A and \tilde{x}_B be two probability distributions. The notion that \tilde{x}_A FSD \tilde{x}_B is the idea that \tilde{x}_A assigns greater probability weight to higher outcome values; equivalently, it assigns lower outcome values a lower probability relative to \tilde{x}_B . Notice that there is no concern for “relative riskiness” in this comparison: the outcomes under \tilde{x}_A could be made more 'spread out' in the region of higher values than \tilde{x}_B .

b. The notion that \tilde{x}_A SSD \tilde{x}_B is the sense that \tilde{x}_B is related to \tilde{x}_A via a “pure increase in risk”. This is the sense of \tilde{x}_B being defined from \tilde{x}_A via a mean preserving spread. \tilde{x}_B is just \tilde{x}_A where the values have been spread out. Of course, any risk averse agent would prefer \tilde{x}_A .

c. Only two moments of a distribution are relevant for comparison: the mean and the variance. Agents like the former and dislike the latter. Thus, given two distributions with the same mean, the one with the higher variance is less desirable; similarly, given two distributions with the same variance, the one with the greater mean return is preferred.

d. (i) Compare first under mean variance criterion.

$$E\tilde{x}_A = \frac{1}{4}(2) + \frac{1}{2}(4) + \frac{1}{4}(9) = 4.75$$

$$E\tilde{x}_B = \frac{1}{3}(1 + 6 + 8) = 5$$

$$\sigma_A^2 = \frac{1}{4}(2 - 4.75)^2 + \frac{1}{2}(4 - 4.75)^2 + \frac{1}{4}(9 - 4.75)^2 = 6.6875$$

$$\begin{aligned}\sigma_B^2 &= \frac{1}{3}(1-5)^2 + \frac{1}{3}(6-5)^2 + \frac{1}{3}(8-5)^2 \\ &= 26/3 = 8\frac{2}{3}.\end{aligned}$$

So, $E\tilde{x}_A < E\tilde{x}_B$, $\sigma_A^2 < \sigma_B^2$

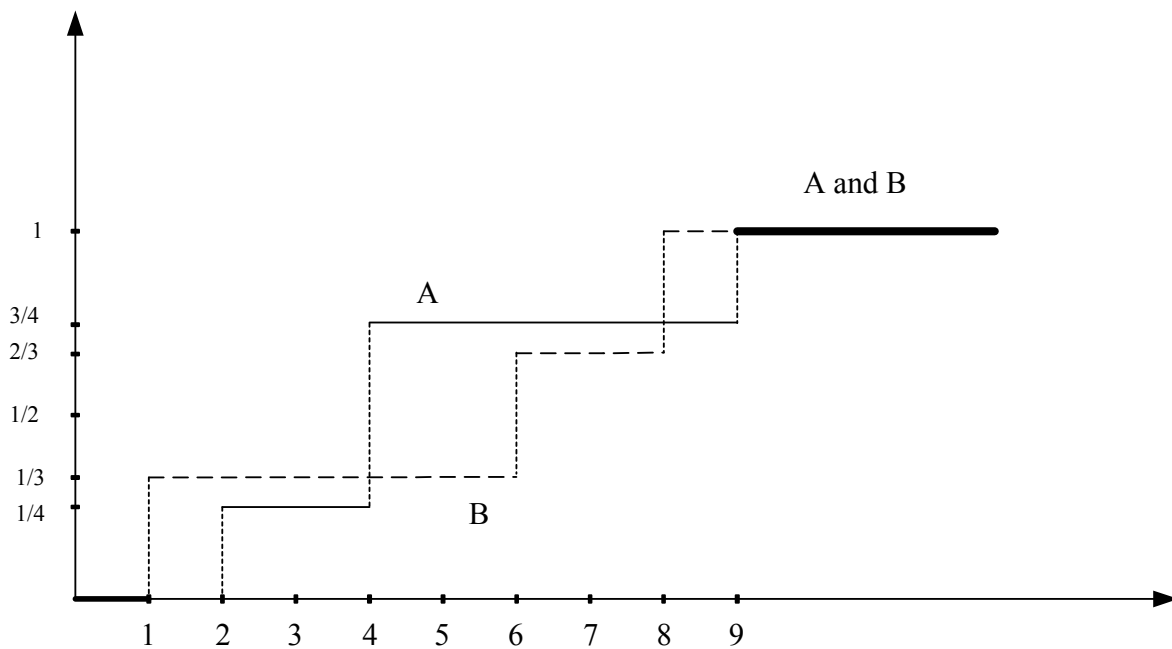
Thus \tilde{x}_A dominates \tilde{x}_B under mean variance.

(ii) Now let's compare them under FSD.

Let $F(\tilde{x}_A)$ be denoted

$F(\tilde{x}_B)$ be denoted

and let us graph $F(\tilde{x}_A)$ and $F(\tilde{x}_B)$



It does not appear that either \tilde{x}_A FSD \tilde{x}_B or \tilde{x}_B FSD \tilde{x}_A ; either dominates the other in the FSD sense.

Thus, mean-variance dominance does not imply FSD.

(iii)

x	$\int_0^x f_B(t)dt$	$\int_0^x F_B(t)dt$	$\int_0^x f_A(t)dt$	$\int_0^x F_A(t)dt$	$\int_0^x [F_B(t) - F_A(t)]dt$
0	0	0	0	0	0
1	1/3	1/3	0	0	1/3
2	1/3	2/3	1/4	1/4	5/12
3	1/3	1	1/4	1/2	1/2
4	1/3	1 1/3	3/4	5/4	1/12
5	1/3	1 2/3	3/4	2	-1/3

There is no SSD as the sign of $\int_0^x [F_B(t) - F_A(t)]dt$ is not monotone.

4.6 a. The certainty equivalent is defined by the equation :

$$U(CE_Z) = EU(\tilde{Z}), \text{ since } Y_0 \equiv 0, \text{ and } \tilde{Z} = (16, 4; \frac{1}{2})$$

$$(CE_Z)^{\frac{1}{2}} = \frac{1}{2}(16)^{\frac{1}{2}} + \frac{1}{2}(4)^{\frac{1}{2}} = \frac{1}{2}(4) + \frac{1}{2}(2) = 3$$

$$\text{Thus } CE_Z = 9$$

b. The insurance policy guarantees the expected payoff :

$$E\tilde{Z} = \frac{1}{2}(16) + \frac{1}{2}(4) = 8 + 2 = 10$$

$$\Pi, \text{ the premium, satisfies } \Pi = E\tilde{Z} - CE_Z = 1.$$

c. The insurance would pay -6 in the high state, +6 in the low state. The most the agent would be willing to pay is 1.

$$U(CE_Z) = 10^{\frac{1}{2}} = \pi'(16)^{\frac{1}{2}} + (1 - \pi')(4)^{\frac{1}{2}}, \pi' = .58$$

The probability premium is .08.

d. Now consider the gamble $(36, 16, \frac{1}{2}) = Z'$

$$U(CE_{Z'}) = (CE_{Z'})^{\frac{1}{2}} = \frac{1}{2}(36)^{\frac{1}{2}} + \frac{1}{2}(16)^{\frac{1}{2}}$$

$$(CE_{Z'})^{\frac{1}{2}} = \frac{1}{2}(6) + \frac{1}{2}(4) = 5$$

$$CE_{Z'} = 25$$

$$\Pi_{Z'} = 1 \text{ (as before)}$$

$$\pi'' \text{ solves : } (26)^{\frac{1}{2}} = \pi''(36)^{\frac{1}{2}} + (1 - \pi'')(16)^{\frac{1}{2}}$$

$$5.10 = \pi''6 + (1 - \pi'')4 = 2\pi'' + 4$$

$$2\pi' = 1.1, \pi' = \frac{1.1}{2} = .55$$

Thus the probability premium is $.55 - .50 = .05$

The probability premium has fallen because the agent is wealthier in the case and is operating on a less risk averse portion of his utility curve. As a result, the premium, as measured in probability terms, is less.

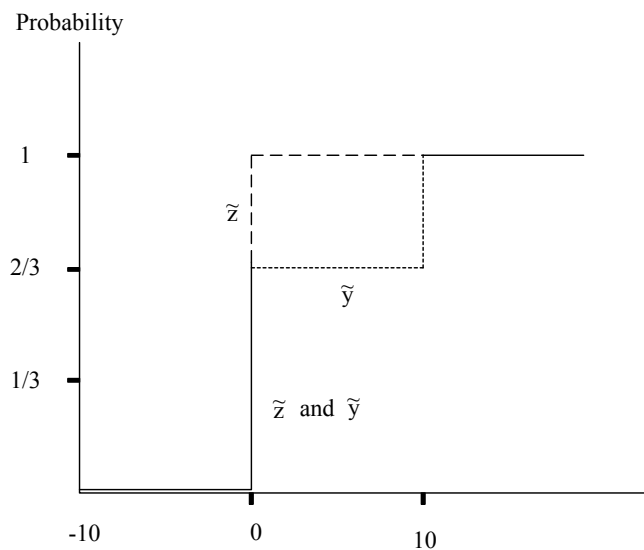
4.7 No. Reworking the data of Table 3.3 shows that it is not always the case that

$$\int_0^x [F_4(t) - F_3(t)] dt > 0.$$

Specifically, for $x = 5$ to 7 , $F_3(x) > F_4(x)$. Graphically this corresponds to the fact that in Fig 3.6 area B of the graph is now bigger than area A.

4.8 a. State by state dominance : no.

b. FSD : yes. See graph



These two notions are not equivalent.

Chapter 5

5.1. For full investment in the risky asset the first order condition has to satisfy the following:

$$E[U'(Y_0(1 + \tilde{r}))(\tilde{r} - r_f)] \geq 0$$

Now expanding $U'(Y_0(1 + \tilde{r}))$ around $Y_0(1 + r_f)$, we get, after some manipulations and ignoring higher terms :

$$\begin{aligned} & E[U'(Y_0(1 + \tilde{r}))(\tilde{r} - r_f)] \\ &= U'[Y_0(1 + r_f)]E(\tilde{r} - r_f) + U''[Y_0(1 + r_f)]E(\tilde{r} - r_f)^2 Y_0 \geq 0 \end{aligned}$$

Hence,

$E(\tilde{r} - r_f) \geq R_A[Y_0(1 + r_f)]E(\tilde{r} - r_f)^2 Y_0$ which is the smallest risk premium required for full investment in the risky asset.

5.2. a.
$$\begin{aligned} R(\pi_1) &= (1 - a)2 + a = 2 - a \\ R(\pi_2) &= (1 - a)2 + 2a = 2 \\ R(\pi_3) &= (1 - a)2 + 3a = 2 + a \end{aligned}$$

b.

$$EU = \pi_1 U(2 - a) + \pi_2 U(2) + \pi_3 U(2 + a)$$

$$\frac{\partial EU}{\partial a} = -\pi_1 U'(2 - a) + \pi_3 U'(2 + a) = 0$$

$$a = 0 \Leftrightarrow \underset{>0}{U'(2)}[\pi_3 - \pi_1] = 0 \Leftrightarrow \pi_3 = \pi_1 \Leftrightarrow E(z) = 2$$

c.

$$\begin{aligned} \text{Define } W(a) &= E(U(Y_0(1 + r_f) + a(\tilde{r} - r_f))) \\ &= E(U(2 + a(z - 2))) \\ W'(a) &= E(U'(2 + a(z - 2))(z - 2)) = 0 \end{aligned}$$

d.

$$\bullet U(Y) = 1 - \exp(-bY)$$

$$W(a) = \pi_1 [1 - \exp(-b(2-a))] + \pi_2 [1 - \exp(-b(2))] + \pi_3 [1 - \exp(-b(2+a))]$$

$$W'(a) = -\pi_1 b \exp(-b(2-a)) + \pi_3 b \exp(-b(2+a)) = 0$$

$$\ln(\pi_3 b)(-b(2+a)) = \ln(\pi_1 b)(-b(2-a))$$

$$a = \frac{\ln(\pi_1) - \ln(\pi_3)}{2b}$$

$$\bullet U(Y) = \frac{1}{1-\gamma} Y^{1-\gamma}$$

$$W(a) = \pi_1 \left[\frac{1}{1-\gamma} (2-a)^{1-\gamma} \right] + \pi_2 \left[\frac{1}{1-\gamma} (2)^{1-\gamma} \right] + \pi_3 \left[\frac{1}{1-\gamma} (2+a)^{1-\gamma} \right]$$

$$W'(a) = -\pi_1 (2-a)^\gamma + \pi_3 (2+a)^\gamma = 0$$

$$a = 2 \frac{\pi_3^{1/\gamma} - \pi_1^{1/\gamma}}{\pi_1^{1/\gamma} + \pi_3^{1/\gamma}}$$

Assuming $\pi_3 > \pi_1, b > 0, 0 < \gamma < 1$ we have in either cases $\frac{\partial a}{\partial Y} > 0$.

e.

$$\bullet U(Y) = 1 - \exp(-bY)$$

$$R_A = b$$

$$\bullet U(Y) = \frac{1}{1-\gamma} Y^{1-\gamma}$$

$$R_A = \gamma / Y$$

5.3

$$\begin{aligned} \text{a. } Y &= (1 + \tilde{r})a + (Y_0 - a)(1 + r_f) \\ &= Y_0(1 + r_f) + a(\tilde{r} - r_f) \end{aligned}$$

$$\text{b. } \max_a EU(Y_0(1 + r_f) + a(\tilde{r} - r_f))$$

$$\text{F.O.C.: } E\{U'(Y_0(1 + r_f) + a^*(\tilde{r} - r_f))(\tilde{r} - r_f)\} = 0$$

$$E\{U''(Y_0(1 + r_f) + a^*(\tilde{r} - r_f))(\tilde{r} - r_f)^2\} < 0$$

Since the second derivative is negative we are at a maximum with respect to a at $a = a^*$, the optimum.

c. We first want to formally obtain $\frac{da^*}{dY_0}$. Take the total differential of the F.O.C. with respect to Y_0 .

Since $E\{U'(Y_0(1+r_f) + a(\tilde{r} - r_f))(\tilde{r} - r_f)\} = 0$,

$$E\left\{U''(Y_0(1+r_f) + a(\tilde{r} - r_f))(\tilde{r} - r_f)\left[(1+r_f) + \frac{da}{dY_0}(\tilde{r} - r_f)\right]\right\} = 0$$

$$- \frac{E\{U''(Y_0(1+r_f) + a(\tilde{r} - r_f))(\tilde{r} - r_f)(1+r_f)\}}{E\{U''(Y_0(1+r_f) + a(\tilde{r} - r_f))(\tilde{r} - r_f)^2\}} = \frac{da}{dY_0}$$

The denominator is < 0 ; in conjunction with the negative preceeding the term we see that the sign of $\frac{da}{dY_0}$ is entirely dependent on $E\{U''(Y_0(1+r_f) + a(\tilde{r} - r_f))(\tilde{r} - r_f)(1+r_f)\}$ the numerator.

Since $(1+r_f) > 0$ it is in fact dependent on $EU''(Y_0(1+r_f) + a(\tilde{r} - r_f))(\tilde{r} - r_f)$. We want to show this latter expression is positive and hence $\frac{da}{dY_0} > 0$, as our intuition would suggest.

d. $R'_A(Y) < 0$ is declining absolute risk aversion: as an investor becomes wealthier he should be willing to place more of his wealth (though not necessarily proportionately more) in the risky asset, if he displays $R'_A(Y) < 0$.

e. We will divide the set of risky realizations into two sets: $\tilde{r} \geq r_f$ and $r_f > \tilde{r}$. We maintain the assumption $R'_A(Y) < 0$.

Case 1: $\tilde{r} \geq r_f$; Notice that

$$Y_0(1+r_f) + a(\tilde{r} - r_f) > Y_0(1+r_f).$$

Then, $R_A(Y_0(1+r_f) + a(\tilde{r} - r_f)) \leq R_A(Y_0(1+r_f))$.

Case 2: if $\tilde{r} < r_f$

$$Y_0(1+r_f) + a(\tilde{r} - r_f) < Y_0(1+r_f).$$

Thus $R_A(Y_0(1+r_f) + a(\tilde{r} - r_f)) > R_A(Y_0(1+r_f))$

f. Now we will use these observations.

Case 1: $\tilde{r} \geq r_f$

$$- \frac{U''(Y_0(1+r_f) + a(\tilde{r} - r_f))}{U'(Y_0(1+r_f) + a(\tilde{r} - r_f))} = R_A(Y_0(1+r_f) + a(\tilde{r} - r_f))$$

$$\leq R_A(Y_0(1+r_f))$$

Hence

$$U''(Y_0(1+r_f) + a(\tilde{r} - r_f)) \geq -R_A(Y_0(1+r_f))U'(Y_0(1+r_f) + a(\tilde{r} - r_f))$$

Since $\tilde{r} - r_f \geq 0$ for this case,

$$(i) \quad U''(Y_0(1+r_f) + a(\tilde{r} - r_f))(\tilde{r} - r_f) \geq -R_A(Y_0(1+r_f))U'(Y_0(1+r_f) + a(\tilde{r} - r_f))(\tilde{r} - r_f)$$

Case 2: $\tilde{r} < r_f$

$$\begin{aligned} -\frac{U''(Y_0(1+r_f) + a(\tilde{r} - r_f))}{U'(Y_0(1+r_f) + a(\tilde{r} - r_f))} &= R_A(Y_0(1+r_f) + a(\tilde{r} - r_f)) \\ &\geq R_A(Y_0(1+r_f)). \end{aligned}$$

Thus,

$$U''(Y_0(1+r_f) + a(\tilde{r} - r_f)) \leq -R_A(Y_0(1+r_f))U'(Y_0(1+r_f) + a(\tilde{r} - r_f)).$$

Since for this case $\tilde{r} - r_f < 0$,

$$(ii) \quad U''(Y_0(1+r_f) + a(\tilde{r} - r_f))(\tilde{r} - r_f) \geq -R_A(Y_0(1+r_f))U'(Y_0(1+r_f) + a(\tilde{r} - r_f))(\tilde{r} - r_f)$$

Since (i) and (ii) encompass “all the probability,” we can write:

$$\begin{aligned} E\{U''(Y_0(1+r_f) + a(\tilde{r} - r_f))(\tilde{r} - r_f)\} \\ > -R_A(Y_0(1+r_f)) E\{U'(Y_0(1+r_f) + a(\tilde{r} - r_f))(\tilde{r} - r_f)\} \end{aligned}$$

Since by the FOC for a ,

$$E\{U'(Y_0(1+r_f) + a(\tilde{r} - r_f))(\tilde{r} - r_f)\} = 0$$

we know $E\{U''(Y_0(1+r_f) + a(\tilde{r} - r_f))(\tilde{r} - r_f)\} \geq 0$

and $\frac{da}{dY_0} \geq 0$.

Key assumption: declining absolute risk aversion.

5.4. Let \tilde{R}_A and \tilde{R}_B represent the gross returns on the two assets ;

$$\begin{aligned} E(U(a\tilde{R}_A + (1-a)\tilde{R}_B)) &> E(aU(\tilde{R}_A) + (1-a)U(\tilde{R}_B)) = aE(U(\tilde{R}_A)) + (1-a)E(U(\tilde{R}_B)) \\ &= E(U(\tilde{R}_A)) = E(U(\tilde{R}_B)) \end{aligned}$$

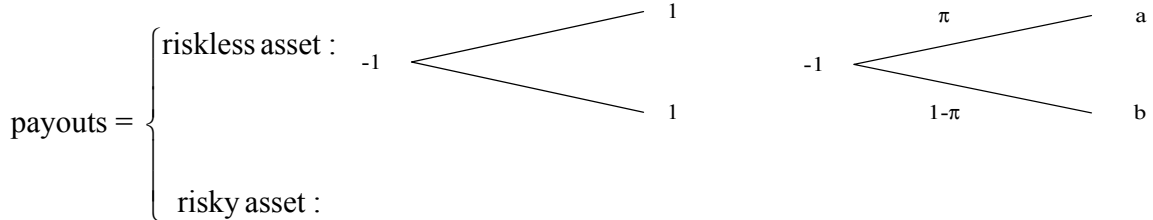
since the utility function is concave and gross asset returns are identically distributed. This result means that the investor is going to invest in both securities – it is never optimal in this situation to invest only in one of the two assets. He thus chooses to diversify. Moreover if the investor cares only about the first two moments he will invest equal amounts in the assets to minimize variance. To show this note that the expected return on the portfolio is constant independently of the chosen allocation. A mean-variance investor will thus choose a to minimize the variance of

the portfolio. The latter is

$$a^2 \text{var } \tilde{R}_A + (1-a)^2 \text{var } \tilde{R}_B = \text{var } \tilde{R}_A (2a^2 + 1 - 2a) = \text{var } \tilde{R}_A (1 + 2(a^2 - a)).$$

Minimizing the portfolio variance thus amounts to minimizing the quantity $(a^2 - a)$ which requires selecting $a=1/2$.

5.5 $Y_0 \equiv 1$



Since the price of both assets is 1, the number of units purchased equals the value of the purchase.

Let x_1 = units (value) of risk free asset demanded

x_2 = units (value) of risky asset demanded

Since $Y_0 \equiv 1$, $x_1 + x_2 = 1$ and $x_2 = 1 - x_1$

a. Necessary condition for $x_1 > 0$:

" $r_f > a$ " or, in this case $a < 1$

Necessary condition for the demand for the risky asset to be positive :

$$\pi a + (1 - \pi)b > 1.$$

b. We would expect $\frac{dx_1}{da} < 0$

The optimization problem is

$$\max \pi U(x_1 \cdot 1 + (1 - x_1)a) + (1 - \pi)U(x_1 \cdot 1 + (1 - x_1)b)$$

$$\text{s.t. } x_1 + x_2 \leq 1 \text{ or } 0 \leq x_1 \leq 1$$

The F.O.C. for an interior solution is

$$\pi U'(x_1 + (1 - x_1)a)(1 - a) + (1 - \pi)U'(x_1 + (1 - x_1)b)(1 - b) = 0. \text{ This equation implicitly defines } x_1 = x_1(a) ; \text{ taking the total differential yields :}$$

$$\begin{aligned} \pi U'(x_1 + (1 - x_1)a)(-1) + \pi U''(x_1 + (1 - x_1)a)(1 - a) \left[(1 - a) \frac{dx_1}{da} + (1 - x_1) \right] \\ + (1 - \pi)U''(x_1 + (1 - x_1)b)(1 - b)^2 \frac{dx_1}{da} = 0 \end{aligned}$$

$$\frac{dx_1}{da} = \frac{\pi U'(x_1 + (1 - x_1)a) - \pi(1 - a)(1 - x_1)U''(x_1 + (1 - x_1)a)}{\pi U''(x_1 + (1 - x_1)a)(1 - a)^2 + (1 - \pi)U''(x_1 + (1 - x_1)b)(1 - b)^2} < 0$$

as the numerator is positive and the denominator negative.

We would expect that an increase in the probability of the unfavorable risky asset state a , would tend to decrease the demand for the risky asset. The “world” is becoming riskier.

Thus $\frac{dx_1}{d\pi} > 0$.

Chapter 6

- 6.1 A high β does not say anything about the level of diversification. The β speaks about the co-variations between the returns on an asset or a portfolio and the returns on the market portfolio. But it does not say anything about the degree of diversification of a portfolio. A portfolio with a high β may or may not be well diversified and it may or may not be far away from the efficient frontier.
- 6.2 Let $\sigma_1 = \sigma_2 = \sigma_3 = \sigma$ be the total risk common to every asset. For an equally weighted portfolio:

$$\sigma_p^2 = \sum_{i=1}^3 w_i^2 \sigma_i^2 = 3\left(\frac{1}{9}\right)\sigma^2; \quad \sigma_p = \sqrt{\frac{1}{3}} \sigma$$

The fraction of asset i 's risk that it contributes to a portfolio is given by ρ_{ip} . Without loss of generality consider asset 1.

$$\begin{aligned} \rho_{ip} &= \frac{\text{cov}(r_1, \sum_{i=1}^3 w_i r_i)}{\sigma_1 \sigma_p} \\ &= \frac{\sum_{i=1}^3 w_i \text{cov}(r_1, r_i)}{\sigma_1 \sigma_p} \\ &= \frac{\frac{1}{3} \text{cov}(r_1, r_1)}{\sigma_1 \sigma_p} \quad (\text{since } \text{cov}(r_i, r_j) = 0 \text{ if } i \neq j) \\ &= \frac{\frac{1}{3} \sigma^2}{\sigma \sqrt{\frac{1}{3}} \sigma} = \frac{\sqrt{\frac{1}{3}}}{\sqrt{3}} = \frac{1}{\sqrt{3}} = 0.57735 \end{aligned}$$

$$(\text{check: } \sigma_p = \sum_{i=1}^3 (\rho_{ip} \sigma_i) w_i = \sum_{i=1}^3 \frac{1}{3} \left(\frac{1}{\sqrt{3}} \sigma\right) = \frac{1}{\sqrt{3}} \sigma)$$

Thus, the fraction 0.57735 of the asset's risk is contributed to the portfolio and the fraction $1 - 0.57735 = 0.42265$ is diversified away.

- 6.3 The CML describes the risk/return tradeoff for efficient portfolios, while the SML describes the same tradeoff for arbitrary assets (that are in "M").

For arbitrary assets that have the same statistical characteristics as M (i.e. that are perfectly positively correlated with M), the two should give identical expected return estimates - and they do:

$$\text{SML: } E\tilde{r}_j = r_f + \frac{\rho_{jM}\sigma_j}{\sigma_M} [E\tilde{r}_M - r_f], \text{ for any asset } j.$$

If “j” is perfectly positively correlated with M,

$$\begin{aligned} E\tilde{r}_j &= r_f + \frac{\sigma_M\sigma_j}{\sigma_M} \left[\frac{Er_M - r_f}{\sigma_M} \right] \\ &= r_f + \sigma_j \left[\frac{Er_M - r_f}{\sigma_M} \right], \text{ which is the equation of the capital market} \\ &\text{line for an efficient portfolio "j".} \end{aligned}$$

6.4

a. No! It is incorrect. The CAPM tells you to equate the expected return on the loan equal to r_f . Since there is some probability of default, you must set the rate higher than r_f in order to insure an expected return equal to r_f .

b. Let r_L = rate on the loans. You want

$$\begin{aligned} 1.1 &= 0.95(1+r_L) + 0.05(1) \\ r_L &= 0.105263 \text{ or } 10.5263\% \end{aligned}$$

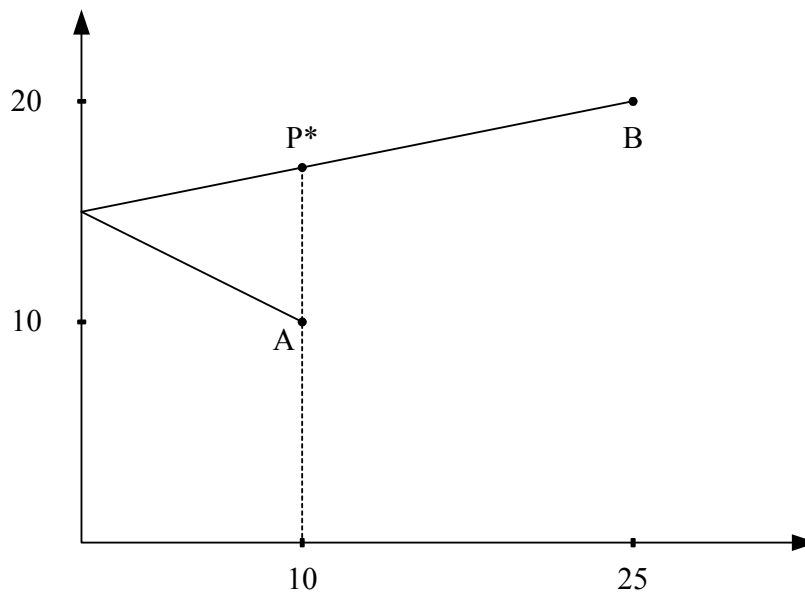
c. Again, let r_L = the rate on the loans:

$$1.1 = 0.95(1+r_L) + 0.05(0.8) \quad (\text{if there is a default, only 80\% of the value of the loan is recoverable}).$$

$$\begin{aligned} 1.1 &= 0.95(1+r_L) + 0.04 \\ 1.06 &= 0.95 + 0.95 r_L \\ r_L &= 0.11/0.95 = 0.1157 \text{ or } 11.57\% \end{aligned}$$

6.5 This problem ‘starkly’ illustrates the gains to diversification. There are two ways to solve it.

a. Method 1 : use the hint



Since $\rho_{AB} = -1$

$$\begin{aligned}\sigma_P &= 10\% = w_B \sigma_B - (1 - w_B) \sigma_A \\ &= w_B (25\%) - (1 - w_B) (10\%) \\ 10\% &= w_B (25\%) - 10\% + 10\% w_B \\ w_B &= \frac{20\%}{35\%} = \frac{4}{7}; w_A = \frac{3}{7}.\end{aligned}$$

$$Er_{P^*} = \frac{3}{7}(10\%) + \frac{4}{7}(20\%) = 15.71\%, \text{ a } 5.71\% \text{ gain.}$$

b. Method 2 : first solve for the minimum risk portfolio.

$$\begin{aligned}\sigma_P &= 0 = w_A \sigma_A - (1 - w_A) \sigma_B \\ 0 &= w_A (10\%) - (1 - w_A) (25\%) \\ w_A &= \frac{5}{7}; w_B = \frac{2}{7}.\end{aligned}$$

$$Er_{\text{min risk portfolio}} = \frac{5}{7}(10\%) + \frac{2}{7}(20\%) = 12.85\%$$

$$\text{Slope of line joining B + min risk portfolio is } \frac{20\% - 12.85\%}{25\%}$$

$$\text{Thus, } Er_{P^*} = 12.85\% + \left(\frac{20\% - 12.85\%}{25\%} \right) (10\%) = 15.71\%, \text{ a } 5.71\% \text{ gain.}$$

6.6 a. The basis for answering this question is the following :

Let C, D be two assets with $\sigma_C < \sigma_D$. Let us consider adding some of 'D' to 'C'. Under what circumstances will the risk go down ? Let $w_C = 1 - w_D$.

$$\sigma_P^2 = (1 - w_D) \sigma_C^2 + w_D^2 \sigma_D^2 + 2w_D(1 - w_D) \text{cov}(\tilde{\epsilon}_C, \tilde{\epsilon}_D)$$

Then

$$\left. \frac{\partial \sigma_p^2}{\partial w_D} \right|_{w_D=0} = 2[\text{cov}(\tilde{r}_C, \tilde{r}_D) - \sigma_C^2]$$

If $\text{cov}(\tilde{r}_C, \tilde{r}_D) - \sigma_C^2 < 0$, if we add 'D' to 'C' the portfolio's risk will decline below that of σ_C . Equivalently

$$\text{cov}(\tilde{r}_C, \tilde{r}_D) < \sigma_C^2$$

$$\rho_{CD} \sigma_C \sigma_D < \sigma_C^2, \text{ or } \rho_{CD} < \frac{\sigma_C}{\sigma_D}$$

or, for the case at hand,

$$\rho_{\text{your port, Australian index}} < \frac{\sigma_{\text{your port.}}}{\sigma_{\text{Aust. index}}}.$$

b. and c.

So we have to compute these data from the sample statistics that we are given.

$$\hat{E}r_{\text{your}} = \frac{1}{6}\{.54 + .24 - .06 + .24 - .06 + .54\} = .24$$

$$\hat{E}r_{\text{Aus.}} = \frac{1}{6}\{.50 + .10 - .10 + .60 - .20 + .80\} = .2833$$

$$\begin{aligned} \hat{\sigma}_{\text{your}}^2 &= \frac{1}{5}\{(.54 - .24)^2 + (.24 - .24)^2 + (-.06 - .24)^2 + (.24 - .24)^2 \\ &\quad + (-.06 - .24)^2 + (.54 - .24)^2\} \\ &= \frac{1}{5}\{.09 + .09 + .09 + .09\} = .072 \end{aligned}$$

$$\begin{aligned} \hat{\sigma}_{\text{Aus.}}^2 &= \frac{1}{5}\{(.50 - .2833)^2 + (-.10 - .2833)^2 + (.10 - .2833)^2 + (.60 - .2833)^2 \\ &\quad + (.20 - .2833)^2 + (.80 - .2833)^2\} \\ &= \frac{1}{5}\{.04696 + .14692 + .03360 + .10030 + .23358 + .2670\} \\ &= \frac{1}{5}\{.82836\} \\ &= .165672 \end{aligned}$$

$$\begin{aligned} \text{cov}(r_{\text{your}}, r_{\text{Aus.}}) &= \frac{1}{5}\{(.54 - .24)(.50 - .2833) + (.24 - .24)(-.10 - .2833) + (-.06 - .24) \\ &\quad (.10 - .2833) + (.24 - .24)(.60 - .2833) + (-.06 - .24)(.20 - .2833) \\ &\quad + (.54 - .24)(.80 - .2833)\} \\ &= \frac{1}{5}\{.06501 + .05499 + .14499 + .15501\} \\ &= .084 \end{aligned}$$

now check the above inequality :

$$\rho_{\text{your, Aus.}} = \frac{\text{cov}(r_{\text{your}}, r_{\text{Aus.}})}{\sigma_{\text{your}} \sigma_{\text{Aus.}}} = \frac{.084}{\sqrt{.072} \sqrt{.166}} = .77$$

Clearly $\rho = .77 > \frac{\sigma_{\text{your}}}{\sigma_{\text{Aus.}}} = \frac{.2683}{.4070} = .6592$

So the answer are :

b. no

c. no. These are, in fact, two ways of asking the same questions.

d. Whenever you are asked a question like this, it is in reference to a regression ; in this case

$$\begin{aligned}\tilde{r}_{\text{Aus}} &= \hat{\alpha}_{\text{Aus}} + \hat{\beta}_{\text{Aus}} r_{\text{your}} + \tilde{\epsilon}_{\text{Aus}} \\ \Rightarrow \sigma_{\text{Aus}}^2 &= (\hat{\beta}_{\text{Aus}})^2 \sigma_{r_{\text{your}}}^2 + \sigma_{\epsilon_{\text{Aus}}}^2\end{aligned}$$

The fraction of Australian's variation explained by variations in your portfolio's return is

$$\begin{aligned}R^2 &= \frac{(\hat{\beta}_{\text{Aus}})^2 \sigma_{\text{your}}^2}{\sigma_{\text{Aus}}^2} = \frac{\left(\frac{\rho_{\text{Aus,your}} \sigma_{\text{Aus}} \sigma_{\text{your}}}{\sigma_{\text{your}}^2} \right)^2 \sigma_{\text{your}}^2}{\sigma_{\text{Aus}}^2} \\ &= (\rho_{\text{Aus,your}})^2 \\ &= (.77)^2 = .59\end{aligned}$$

e. No ! How could it be ? Adding Australian stocks doesn't reduce risk. Even if it did, you don't know that your portfolio is identical to the true M.

Chapter 7

- 7.1. Write the SML equation to make the market risk premium appear, then multiply by $\frac{\sigma_M}{\sigma_M}$,

$$\begin{aligned} E(r_j) &= r_f + (E(r_M) - r_f)\beta_j = r_f + (E(r_M) - r_f) \frac{\sigma_{jM}}{\sigma_M^2} \\ &= r_f + \frac{(E(r_M) - r_f)}{\sigma_M} \frac{\sigma_{jM}\sigma_M}{\sigma_M^2} \end{aligned}$$

Rewrite the last term

$$\frac{\sigma_{jM}\sigma_M}{\sigma_M^2} = \frac{\sigma_j\sigma_M\rho_{jM}\sigma_M}{\sigma_M^2} = \sigma_j\rho_{jM}$$

Then we get

$$E(r_j) = r_f + \frac{(E(r_M) - r_f)}{\sigma_M} (\sigma_j\rho_{jM})$$

and the conclusion follows since $0 \leq |\rho_{jM}| \leq 1$.

- 7.2. Intuitively, the CML in the ‘more risk averse economy’ should be steeper, in view of its risk/return trade-off interpretation. This is true in particular because one would expect the risk free rate to be lower, as the demand for the risk free asset should be higher, and the return on the optimal risky portfolio to be higher, as the more risk averse investors require a higher compensation for the risk they bear. Note, however, that the Markovitz model is not framed to answer such a question explicitly. It builds on ‘given’ expected returns that are assumed to be ‘equilibrium’. If we imagine, as in this question, a change in the primitives of the economy, we have to turn to our intuition to guess how these given returns would differ in the alternative set of circumstances. The model does not help us with this reasoning. For such (fundamental) questions, a general equilibrium setting will prove superior.
- 7.3. The frontier of the economy where asset returns are more correlated and where diversification opportunities are thus lower is contained inside the efficient frontier of the economy where assets are less correlated. If the risk free rate was constant, this would guarantee that the slope of the CML would be lower in the economy, and the reward for risk taking lower as well. This appears counter-intuitive – with less diversification opportunities those who take risks should get a higher reward –, an observation which suggests that the risk free rate should be lower in the higher correlation economy. Refer to our remarks in the solution to 6.2 : the CAPM model, by its nature, does not explicitly help us answer such a question.
- 7.4. If investors hold homogeneous expectations concerning asset returns, mean returns on risky assets -per dollar invested- will be the same. Otherwise they would face different efficient frontiers and most likely would invest different proportions in risky assets. Moreover, the marginal rate of substitution between risk and return would depend on the level of wealth.

- 7.5. Using standard notations and applying the formulas, we get
 $A = 3.77, B = 5.85, C = 2.65, D = 1.31$

$$g = \begin{bmatrix} 1.529 \\ -0.059 \\ -0.471 \end{bmatrix} \quad h = \begin{bmatrix} -0.618 \\ 0.235 \\ 0.382 \end{bmatrix}$$

$$E(r_{MVP}) = 1.42$$

$$w_{MVP} = \begin{bmatrix} 0.652 \\ 0.275 \\ 0.072 \end{bmatrix}$$

$$E(r_{ZCP_3}) = 1.3028$$

$$w_{ZCP_3} = \begin{bmatrix} 0.725 \\ 0.248 \\ 0.028 \end{bmatrix}$$

- 7.6 a. The agent's problem is (agent i):

$$\max_{(x_1^i, x_2^i, \dots, x_j^i)} E \left\{ U_i \left[(Y_0^i - \sum_j x_j^i)(1 + r_f) + \sum_j x_j^i(1 + \tilde{r}_j) \right] \right\}$$

The F.O.C. wrt asset j is :

$$\begin{aligned} E \{ U_i'(\tilde{Y}_i) (- (1 + r_f) + (1 + \tilde{r}_j)) \} &= 0, \text{ or} \\ E \{ U_i'(\tilde{Y}_i) (\tilde{r}_j - r_f) \} &= 0 \end{aligned} \quad (1)$$

- b. We apply the relationship

$E_{xy} = \text{cov}(x, y) + E_x E_y$ to equation (1).

$$\begin{aligned} 0 &= E \{ U_i'(\tilde{Y}_i) (\tilde{r}_j - r_f) \} = E \{ U_i'(\tilde{Y}_i) \} E(\tilde{r}_j - r_f) + \text{cov}(U_i'(\tilde{Y}_i), \tilde{r}_j - r_f) \\ &= E \{ U_i'(\tilde{Y}_i) \} E(\tilde{r}_j - r_f) + \text{cov}(U_i'(\tilde{Y}_i), \tilde{r}_j) \end{aligned}$$

$$\text{Thus, } E \{ U_i'(\tilde{Y}_i) \} E(\tilde{r}_j - r_f) = -\text{cov}(U_i'(\tilde{Y}_i), \tilde{r}_j) \quad (2)$$

- c. We make use of the relationship

$\text{cov}(g(\tilde{x}), \tilde{y}) = E(g'(\tilde{x})) \text{cov}(\tilde{x}, \tilde{y})$, where we identify $g(\tilde{Y}_i) = U_i'(\tilde{Y}_i)$.

Apply this result to the R.H.S. of equation (2) yields :

$$E \{ U_i'(\tilde{Y}_i) \} E(\tilde{r}_j - r_f) = -E(U_i''(\tilde{Y}_i)) \text{cov}(\tilde{Y}_i, \tilde{r}_j). \quad (3)$$

- d. We can rewrite equation (3) as :

$$E(\tilde{r}_j - r_f) = \frac{-E\{U_i''(\tilde{Y}_i)\}}{E\{U_i'(\tilde{Y}_i)\}} \text{cov}(\tilde{Y}_i, \tilde{r}_j)$$

Let us denote $R_{A_i} = \frac{-E\{U_i''(\tilde{Y}_i)\}}{E\{U_i'(\tilde{Y}_i)\}}$ as it is reminiscent (but not equal to) the Absolute Risk

Aversion measure.

The above equation can be rewritten as

$$E(\tilde{r}_j - r_f) = \left(\frac{1}{\frac{1}{R_{A_i}}} \right) \text{cov}(\tilde{Y}_i, \tilde{r}_j), \text{ or}$$

$$\left(\frac{1}{R_{A_i}} \right) E(\tilde{r}_j - r_f) = \text{cov}(\tilde{Y}_i, \tilde{r}_j),$$

Summing over all agents i gives

$$\sum_{i=1}^I \left(\frac{1}{R_{A_i}} \right) E(\tilde{r}_j - r_f) = \sum_{i=1}^I \text{cov}(\tilde{Y}_i, \tilde{r}_j), \text{ or}$$

$$E(\tilde{r}_j - r_f) \left(\sum_{i=1}^I \left(\frac{1}{R_{A_i}} \right) \right) = \text{cov} \left(\sum_{i=1}^I \tilde{Y}_i, \tilde{r}_j \right)$$

Let us identify $\sum_{i=1}^I \tilde{Y}_i \equiv Y_{MO}(1 + \tilde{r}_M)$. Then we have :

$$\begin{aligned} E(\tilde{r}_j - r_f) \left(\sum_{i=1}^I \left(\frac{1}{R_{A_i}} \right) \right) &= \text{cov}(Y_{MO}(1 + \tilde{r}_M), \tilde{r}_j) \\ &= Y_{MO} \text{cov}(\tilde{r}_M, \tilde{r}_j) \end{aligned}$$

Thus,

$$E(\tilde{r}_j - r_f) = \frac{Y_{MO}}{\left(\sum_{i=1}^I \left(\frac{1}{R_{A_i}} \right) \right)} \text{cov}(\tilde{r}_M, \tilde{r}_j) \quad (4)$$

e. Let w_j be the proportion of economy wide wealth invested in asset j .

Then, for all j

$$w_j E(\tilde{r}_j - r_f) = \frac{Y_{MO}}{\left(\sum_{i=1}^I \left(\frac{1}{R_{A_i}} \right) \right)} w_j \text{cov}(\tilde{r}_M, \tilde{r}_j)$$

Thus,

$$\sum_{j=1}^J w_j E(\tilde{r}_j - r_f) = \frac{Y_{MO}}{\left(\sum_{i=1}^I \left(\frac{1}{R_{A_i}} \right) \right)} \sum_{j=1}^J w_j \text{cov}(\tilde{r}_M, \tilde{r}_j).$$

It follows that

$$E\left(\sum_{j=1}^J w_j \tilde{r}_j - r_f\right) = \frac{Y_{MO}}{\left(\sum_{i=1}^I \left(\frac{1}{R_{A_i}}\right)\right)} \text{cov}\left(\tilde{r}_M, \sum_{j=1}^J w_j \tilde{r}_j\right)$$

By construction,

$$\sum_{j=1}^J w_j \tilde{r}_j = \tilde{r}_M.$$

Then

$$\begin{aligned} E\tilde{r}_M - r_f &= \frac{Y_{MO}}{\left(\sum_{i=1}^I \left(\frac{1}{R_{A_i}}\right)\right)} \text{cov}(\tilde{r}_M, \tilde{r}_M) \\ E\tilde{r}_M - r_f &= \frac{Y_{MO}}{\left(\sum_{i=1}^I \left(\frac{1}{R_{A_i}}\right)\right)} \text{var}(\tilde{r}_M). \end{aligned} \quad (5)$$

f. (4) states that

$$E(\tilde{r}_j - r_f) = \frac{Y_{MO}}{\left(\sum_{i=1}^I \left(\frac{1}{R_{A_i}}\right)\right)} \text{cov}(\tilde{r}_M, \tilde{r}_j)$$

From (5) $\frac{Y_{MO}}{\left(\sum_{i=1}^I \left(\frac{1}{R_{A_i}}\right)\right)} = \frac{E(\tilde{r}_M - r_f)}{\text{var}(\tilde{r}_M)}$; substituting this latter expression into (4) gives :

$$E(\tilde{r}_j - r_f) = \frac{\text{cov}(\tilde{r}_M, \tilde{r}_j)}{\text{var}(\tilde{r}_M)} E(\tilde{r}_M - r_f), \text{ the traditional CAPM.}$$

Chapter 8

8.1. a.

$$q_i = \frac{1}{(1 + r_i)^i}$$

$$q_1 = 0.91$$

$$q_2 = 0.8224$$

$$q_3 = 0.7424$$

These are in fact the corresponding risk-free discount bond prices.

b. The matrix is the same at each date. The n-period A-D matrix is then $[A - D]^n$. If we are in state i today, we look at line i, written $[A - D]_i^n$, and we sum the corresponding A-D prices to obtain a sure payoff of one unit in each future state. Since it is assumed we are in state 1

$$[A - D]_1 = 0.28 + 0.33 + 0.30 = 0.91 = q_1$$

$$[A - D]_1^2 = 0.8224 = q_2$$

$$[A - D]_1^3 = 0.7424 = q_3$$

8.2 The price of an A-D security is the (subjective) probability weighted MRS in the corresponding state. It is determined by three considerations: the discount factor which is imbedded in the MU of future consumption, the state probability and the relative scarcities reflected in the intertemporal marginal rate of substitution, that is, in the ratio of the future MU to the present MU. The latter is affected by the expected consumption/ endowment in the future state and by the shape of the agents' utility functions (their rates of risk aversion).

8.3. We determine a term structure for each initial state.

To-day's state is 1:

$$1 + r_1^1 = \frac{1}{0.53 + 0.43} = 1.0417$$

$$(1 + r_1^2)^2 = 1.0800$$

$$1 + r_1^2 = 1.0392$$

$$(1 + r_1^3)^3 = 1.1193$$

$$1 + r_1^3 = 1.0383$$

To-day's state is 2 :

$$1 + r_2^1 = 1.0310$$

$$(1 + r_2^2)^2 = 1.0679$$

$$1 + r_2^2 = 1.0334$$

$$(1 + r_2^3)^3 = 1.1067$$

$$1 + r_2^3 = 1.0344$$

8.4. We determine the price of A-D securities for each date, starting with bond 1 for date 1, $q_1 = 96/100 = 0.96$. Then we use the method of pricing intermediate cash flows with A-D prices to

price bonds of longer maturity, for example the price of bond 2 is such that

$$900 = 100 \times q_1 + 1100 \times q_2 \equiv \frac{100}{1+r_1} + \frac{1100}{(1+r_2)^2}$$

which gives $q_2 = \frac{900 - 100 \times q_1}{1100} = 0.7309$. Similarly

$$q_3 = 0.5331$$

$$q_4 = 0.3194$$

$$q_5 = 0.01608$$

- 8.5. a. Given preferences and endowments, it is clear that the allocation $\{(4, 2, 2), (4, 2, 2)\}$ is PO and feasible. In general, there is an infinity of PO allocations.

b. Yes, but only if one of the following securities is traded

$$s_1 = \begin{matrix} -1 \\ 1 \end{matrix} \quad \text{or} \quad s_2 = \begin{matrix} 1 \\ -1 \end{matrix}$$

For example, Agent 1 would sell s_1 , and Agent 2 would buy it. In general one security is not sufficient to complete the markets when there are two future states.

c. Agents will be happy to store the commodity for two reasons : consumption smoothing – they are pleased to transfer consumption from period 1 to period 2-, and in addition by shifting to tomorrow some of the current consumption they are able to reduce somewhat (but not fully) the endowment risk they face. For these two reasons, storing will enable them to increase their utility level.

d. Remember aggregate uncertainty means that the total quantity available at date 2 is not the same for all the states. If one agent is risk-neutral, he will however be willing to bear all the risks. Provided enough trading instruments exist, the consumption of the risk-averse agent can thus be completely smoothed out and this constitutes a Pareto Optimum.

- 8.6 a.
1. Because of the variance term diminishing utility, consumption should be equated across states for each agent.
 2. There are many Pareto optima. For example, the allocations below are both Pareto optimal :

	t=0	t=1	
		$\theta = 1$	$\theta = 2$
<u>Allocation 1</u>			
Agent 1	4	3	3
Agent 2	4	3	3
<u>Allocation 2</u>			
Agent 1	5	4	4
Agent 2	3	2	2

The set of Pareto optima satisfies:

$$\{(c_0^1, c_1^1, c_2^1), (c_0^2, c_1^2, c_2^2)\} : c_1^1 = c_2^1 \text{ (and thus } c_2^1 = c_2^2), c_1^2 + c_1^1 = 6; c_0^1 + c_0^2 = 8\}$$

3. Yes. Given $E(c)$ in the second period, $\text{var } c$ is minimized.

b.

1. The Pareto optima satisfy

$$\max_{c_0^1, c_1^1, c_2^1} \left\{ c_0^1 + \frac{1}{4} \ln c_1^1 + \frac{3}{4} \ln c_2^1 + \lambda \left[8 - c_0^1 + \frac{1}{4} \ln(6 - c_1^1) + \frac{3}{4} \ln(6 - c_2^1) \right] \right\}$$

The F.O.C.'s are:

$$\text{i) } c_0^1 : 1 - \lambda = 0$$

$$\text{ii) } c_1^1 : \frac{1}{4} \left(\frac{1}{c_1^1} \right) + \lambda \left(\frac{1}{4} \right) \left(\frac{1}{6 - c_1^1} \right) (-1) = 0$$

$$\text{iii) } c_2^1 : \frac{3}{4} \left(\frac{1}{c_2^1} \right) + \lambda \left(\frac{3}{4} \right) \left(\frac{1}{6 - c_2^1} \right) (-1) = 0$$

$$\text{From (ii) } \frac{1}{c_1^1} = \lambda \left(\frac{1}{6 - c_1^1} \right) \Rightarrow c_1^1 = \frac{6}{1 + \lambda}.$$

$$\text{From (iii) } \frac{1}{c_2^1} = \lambda \left(\frac{1}{6 - c_2^1} \right) \Rightarrow c_2^1 = \frac{6}{1 + \lambda}$$

A Pareto optimum clearly requires $c_1^1 = c_2^1$, and thus $c_1^2 = c_2^2$;

$$\text{If } \lambda > 1, \quad c_0^1 = 0, c_0^2 = 8$$

$$\text{If } \lambda = 1, \quad c_0^1 + c_0^2 = 8$$

$$\text{If } \lambda < 1, \quad c_0^1 = 8, c_0^2 = 0$$

The Pareto optimal allocation here and for the first part of the problem are the same. Both agents are risk averse and we would expect them to try to standardize period 1 consumption.

2. Agents' problems can be written

$$\text{Agent 1: } \max_{Q_1^1, Q_2^1} (4 - P_1 Q_1^1 - P_2 Q_2^1) + \frac{1}{4} \ln(1 + Q_1^1) + \frac{3}{4} \ln(5 + Q_2^1)$$

$$\text{Agent 2: } \max_{Q_1^2, Q_2^2} (4 - P_1 Q_1^2 - P_2 Q_2^2) + \frac{1}{4} \ln(5 + Q_1^2) + \frac{3}{4} \ln(1 + Q_2^2)$$

Market clearing conditions:

$$Q_1^1 + Q_1^2 = 0$$

$$Q_2^1 + Q_2^2 = 0$$

(both securities are in zero net supply).

The F.O.C.'s are:

$$\text{Agent 1: } Q_1^1 : P_1 = \frac{1}{4} \left(\frac{1}{1 + Q_1^1} \right)$$

$$Q_2^1 : P_2 = \frac{3}{4} \left(\frac{1}{5 + Q_2^1} \right)$$

$$\text{Agent 2: } Q_1^2 : P_1 = \frac{1}{4} \left(\frac{1}{5 + Q_1^2} \right)$$

$$Q_2^2 : P_2 = \frac{3}{4} \left(\frac{1}{1 + Q_2^2} \right)$$

These F.O.C.'s, together with market clearing imply, as expected:

$$\frac{1}{1 + Q_1^1} = \frac{1}{5 + Q_1^2} \Rightarrow Q_1^1 = 2; Q_1^2 = -2.$$

$$\frac{1}{5 + Q_1^1} = \frac{1}{1 + Q_2^2} \Rightarrow Q_2^1 = -2; Q_2^2 = +2.$$

$$\text{Thus, } P_1 = \frac{1}{4} \left(\frac{1}{1 + Q_1^1} \right) = \frac{1}{4} \left(\frac{1}{3} \right) = \frac{1}{12}$$

$$P_2 = \frac{3}{4} \left(\frac{1}{5 + Q_1^1} \right) = \frac{3}{4} \left(\frac{1}{3} \right) = \frac{3}{12}$$

Allocations at Equilibrium:

	t=0	t=1	
		$\theta = 1$	$\theta = 2$
Agent 1:	$4 - \frac{1}{12}(2) - \frac{3}{12}(-2) = 4\frac{1}{3}$	3	3
Agent 2:	$4 - \frac{1}{12}(-2) - \frac{3}{12}(2) = 3\frac{2}{3}$	3	3

This is a Pareto Optima, consumption is stabilized in t=1. However, since agent 1 had more consumption in the more likely state, he is paid in terms of t=0 consumption for agreeing to the exchanges. Agent 2 transfers t=0 wealth to him.

3. Now only (1,0) is traded. The C.E. will not be Pareto optimal as the market is incomplete. The C.E. is as follows:

$$\text{Agent 1: } \max_{Q_1^1} (4 - P_1 Q_1^1) + \frac{1}{4} \ln(1 + Q_1^1) + \frac{3}{4} \ln(5)$$

$$\text{Agent 2: } \max_{Q_1^2} (4 - P_1 Q_1^2) + \frac{1}{4} \ln(5 + Q_1^2) + \frac{3}{4} \ln(1)$$

The F.O.C.'s are :

$$\text{Agent 1: } P_1 = \frac{1}{4} \left(\frac{1}{1 + Q_1^1} \right)$$

$$\text{Agent 2: } P_1 = \frac{1}{4} \left(\frac{1}{5 + Q_1^2} \right)$$

Thus

$$\frac{1}{1 + Q_1^1} = \frac{1}{5 + Q_1^2} \Rightarrow Q_1^1 = +2; Q_1^2 = -2. \quad P_1 = \frac{1}{4} \left(\frac{1}{3} \right) = \frac{1}{12}.$$

Allocation

	t=0	t=1	
		$\theta = 1$	$\theta = 2$
Agent 1:	$4 - \frac{1}{6} = 3\frac{5}{6}$	3	5
Agent 2:	$4 + \frac{1}{6} = 4\frac{1}{6}$	3	1

Consumption is stabilized in state θ_1 : effectively agent 1 buys consumption insurance from agent 2.

8.7 The Pareto optima satisfy:

$$\max_{c_0^1, c_1^1, c_2^1} .25c_0^1 + .5\left[\frac{1}{2}\ln c_1^1 + \frac{1}{2}\ln c_2^1\right] + \lambda\left[6 - c_0^1 + \frac{1}{2}\ln(6 - c_1^1) + \frac{1}{2}\ln(6 - c_2^1)\right]$$

The F.O.C.'s are

$$c_0^1 : .25 - \lambda = 0$$

$$c_1^1 : .5\left(\frac{1}{2}\right)\left(\frac{1}{c_1^1}\right) + \lambda\left(\frac{1}{2}\right)\left(\frac{1}{6 - c_1^1}\right)(-1) = 0$$

$$c_2^1 : .5\left(\frac{1}{2}\right)\left(\frac{1}{c_2^1}\right) + \lambda\left(\frac{1}{2}\right)\left(\frac{1}{6 - c_2^1}\right)(-1) = 0$$

$$\frac{1}{2}\left(\frac{1}{c_1^1}\right) = \lambda\left(\frac{1}{6 - c_1^1}\right) \Rightarrow 6 - c_1^1 = 2c_1^1\lambda, \quad c_1^1 = \frac{6}{1 + 2\lambda}.$$

$$\frac{1}{2}\left(\frac{1}{c_2^1}\right) = \lambda\left(\frac{1}{6 - c_2^1}\right) \Rightarrow 6 - c_2^1 = 2c_2^1\lambda, \quad c_2^1 = \frac{6}{1 + 2\lambda}$$

Thus $c_1^1 = c_2^1$, and therefore $c_1^2 = c_2^2$;

If there is no aggregate risk and the agents preferences are the same state by state, then a Pareto optimum will require perfect risk sharing. This example has these features.

The Pareto optimum is clearly not unique. The set of Pareto optima can be described by:

For all $\lambda \geq 0$

	t=0	t=1	
		$\theta = 1$	$\theta = 2$
Agent 1:	$\begin{cases} 6 & \text{if } \lambda < .25 \\ 0 & \text{if } \lambda > .25 \end{cases}$	$\frac{6}{1 + 2\lambda}$	$\frac{6}{1 + 2\lambda}$
Agent 2:	$\begin{cases} 0 & \text{if } \lambda < .25 \\ 6 & \text{if } \lambda > .25 \end{cases}$	$6\left(1 - \frac{1}{1 + 2\lambda}\right)$	$6\left(1 - \frac{1}{1 + 2\lambda}\right)$
	$\lambda = .25$, indeterminate		

In the second case (state 2 endowment = 5 for agent 1, 3 for agent 2), there will be a Pareto optimum but it will be impossible to achieve perfect risk sharing as there is aggregate risk.

b. The agents' problems are:

$$\text{Agent 1: } \max_{Q^1, R^1} .25(2 - P_Q Q^1 - P_R R^1) + .5 \left\{ \frac{1}{2} \ln(2 + Q^1) + \frac{1}{2} \ln(4 + R^1) \right\}$$

$$\text{Agent 2: } \max_{Q^2, R^2} (4 - P_Q Q^2 - P_R R^2) + \frac{1}{2} \ln(4 + Q^2) + \frac{1}{2} \ln(2 + R^2)$$

Where, in equilibrium, $Q^1 + Q^2 = 0$ (market clearing). Both securities are in zero net supply.
 $R^1 + R^2 = 0$

The F.O.C.'s are

Agent 1:

$$Q^1: .25 P_Q = .5 \left(\frac{1}{2} \right) \left(\frac{1}{2 + Q^1} \right) \Leftrightarrow P_Q = \frac{1}{2 + Q^1}$$

$$R^1: .25 P_R = .5 \left(\frac{1}{2} \right) \left(\frac{1}{4 + R^1} \right) \Leftrightarrow P_R = \frac{1}{4 + R^1}$$

Agent 2:

$$Q^2: P_Q = \frac{1}{2} \left(\frac{1}{4 + Q^2} \right)$$

$$R^2: P_R = \frac{1}{2} \left(\frac{1}{2 + R^2} \right)$$

This implies:

$$\frac{1}{2 + Q^1} = \frac{1}{2} \left(\frac{1}{4 + Q^2} \right) = \frac{1}{2} \left(\frac{1}{4 - Q^1} \right)$$

$$Q^1 = 2; Q^2 = -2$$

$$\frac{1}{4 + R^1} = \frac{1}{2} \left(\frac{1}{2 + R^2} \right) = \frac{1}{2} \left(\frac{1}{2 - R^1} \right)$$

$$2(2 - R^1) = 4 + R^1, R^1 = 0, R^2 = 0$$

$$\text{As a result, } P_Q = \frac{1}{2 + Q^1} = \frac{1}{2 + 2} = \frac{1}{4}.$$

$$P_R = \left(\frac{1}{4 + R^1} \right) = \frac{1}{4}$$

The implied allocations are thus:

	t=0	t=1	
		θ_1	θ_2
Agent 1:	2-1/2=1.5	4	4
Agent 2:	4+1/2=4.5	2	2

c. Let us assume the firm can introduce 1 unit of either security. Either way, the problems of the agents and their F.O.C.'s are not affected. What is affected are the market clearing conditions:

If 1 unit of Q is introduced	If 1 unit of R is introduced
$Q^1 + Q^2 = 1$	$Q^1 + Q^2 = 0$
$R^1 + R^2 = 0$	$R^1 + R^2 = 1$

Let's value the securities in either case.

If one unit of Q is introduced:

The F.O.C.'s become

Agent 1: $P_Q = \frac{1}{2 + Q^1}$

$$P_R = \frac{1}{4 + R^1}$$

Agent 2: $P_Q = \left(\frac{1}{4 + Q^2} \right) \frac{1}{2} = \frac{1}{2} \left(\frac{1}{4 + 1 - Q^1} \right) = \frac{1}{2(5 - Q^2)}$

$$P_R = \frac{1}{2} \left(\frac{1}{2 + R^2} \right) = \frac{1}{2} \left(\frac{1}{2 - R^1} \right)$$

The equation involving R are unchanged. Thus $P_R = 1/4$, $R^1 = 0$, $R^2 = 0$

For the security Q we need to solve:

$$\frac{1}{2 + Q^1} = \frac{1}{2(5 - Q^2)}, \Rightarrow 10 - 2Q^1 = 2 + Q^1$$

$$Q^1 = \frac{8}{3}; Q^2 = 1 - Q^1 = 1 - \frac{8}{3} = -\frac{5}{3}$$

$$\text{Thus } P_Q = \frac{1}{2 + \frac{8}{3}} = \frac{1}{14/3} = \frac{3}{14} < .25.$$

You know the price had to go down: there is more supply of the security.

The implied allocations are thus:

	t=0	t=1	
		$\theta = 1$	$\theta = 2$
Agent 1:	$2 - \frac{1}{2} = 1.5$	4	4
Agent 2:	$4 + \frac{1}{2} = 4.5$	2	2

If one unit of R is introduced:

The first order conditions become, with market clearing conditions imposed:

$$\text{Agent 1} \quad \begin{cases} P_Q = \frac{1}{2 + Q^1} \\ P_R = \frac{1}{4 + R^1} \end{cases}$$

$$\text{Agent 2} \quad \begin{cases} P_Q = \frac{1}{2} \left(\frac{1}{4 + Q^2} \right) = \frac{1}{2(4 - Q^1)} \\ P_R = \frac{1}{2} \left(\frac{1}{2 + R^2} \right) = \frac{1}{2} \left(\frac{1}{3 - R^1} \right) \end{cases}$$

So, P_Q is unchanged, and $P_Q = 1/4$ $Q^1 = 2, Q^2 = -2$.

Solving for P_R :

$$\frac{1}{4 + R^1} = \frac{1}{2(3 - R^1)}, \quad R^1 = \frac{2}{3}; R^2 = 1 - R^1 = \frac{1}{3}$$

$$P_{R^1} = \frac{1}{4 + R^1} = \frac{1}{4 + \frac{2}{3}} = \frac{1}{\frac{14}{3}} = \frac{3}{14} < .25$$

	t=0	t=1	
		$\theta = 1$	$\theta = 2$
Agent 1:	$2 - \frac{1}{2} - \frac{2}{3} \frac{3}{14} = 2 - \frac{9}{14}$	4	$4 + 2/3$
Agent 2:	$4 + \frac{1}{2} - \frac{1}{3} \frac{3}{14} = 4 + \frac{6}{14}$	2	$2 + 1/3$

The firm is indifferent as to which security it sells – either way it receives the same thing. Either way a Pareto optimum is achieved since, with no short sales constraints, the market is complete. Thus a C.E. is Pareto Optimal.

Agent 1 wishes to transfer income to period $t=1$. The introduction of more securities of either type will reduce the cost to him of doing that. Agent 2, however, will receive lower prices – either way – for the securities he issues. He will be hurt.

8.8 a. At a P.O. allocation there is no waste and there are no possibilities to redistribute goods and make everyone better off. From the viewpoint of social welfare there seems to be no argument not to search for the realization of a Pareto Optimum. Beyond considerations of efficiency, however, considerations of social justice might suggest some non-optimal allocations are in fact socially preferable to some Pareto optimal ones. These issues are at the heart of many political discussions in a world where redistribution across agents is not costless. From a purely financial perspective, we associate the failure to reach a Pareto optimal allocation with the failure to smooth various agents' consumptions across time or states as much as would in fact be feasible. Again there is a loss in welfare and this is socially relevant: we should care.

b. The answer to a indicates we should care since complete markets are required to guarantee that a Pareto optimal allocation is reached. Are markets complete? certainly not! Are we far from complete markets? Would the world be much better with significantly more complete markets? This is a subject of passionate debates that cannot be resolved here. You may want to re-read the concluding comments of Chapter 1 at this stage.

Chapter 9

- 9.1. a. The CCAPM is an intertemporal model whereas the CAPM is a one-period model. The CCAPM makes a full investors homogeneity assumption but does not require specific utility functions.
- b. The key contribution of the CCAPM resides in that the portfolio problem is indeed inherently intertemporal. The link with the real side of the economy is also more apparent in the CCAPM which does provide a better platform to think about many important questions in asset management.
- c. The two models are equivalent in a one-period exchange economy since then aggregate consumption and wealth is the same. More generally, the prescriptions of the two models would be very similar in situations where consumption would be expected to be closely correlated with variations in the value of the market portfolio.

- 9.2. a. $\max(0, S_{t+1}(\theta) - p^*)$
- b. $S_{t+1}(\theta) = \sum_{\tau=1}^{\infty} \delta^{\tau} \frac{1/c_{t+1+\tau}}{1/c_{t+1}(\theta)} c_{t+1+\tau} = \frac{\delta}{1-\delta} c_{t+1}(\theta)$
- c. $q_{t+1}(\theta) = p(\theta) \delta \frac{c_t}{c_{t+1}(\theta)}$
- d. The price of the option is,

$$C_t = \sum_{\theta' \in A} q_{t+1}(\theta') (S(\theta') - p^*)$$

where A is a set of states θ' for which $(S_{t+1}(\theta') - p^*) \geq 0$.

- 9.3. a. $(S_{t+1}(\theta) - p^*)$
- b. $S_{t+1}(\theta) = \sum_{\tau=1}^{\infty} \delta^{\tau} \frac{1/c_{t+1+\tau}}{1/c_{t+1}(\theta)} c_{t+1+\tau} = \frac{\delta}{1-\delta} c_{t+1}(\theta)$
- c. $q_{t+1}(\theta) = p(\theta) \delta \frac{c_t}{c_{t+1}(\theta)}$
- d. The price of the forward contract is,

$$F_t(\theta) = \sum_{\theta} q_{t+1}(\theta) (S_{t+1}(\theta) - p^*) .$$

- 9.4. a. After maximization, the pricing kernel from date 0 to date t takes the form $\frac{\delta^t c_0}{\delta^0 c_t} = m_t$. Now the value of the wealth portfolio is $P_0 = E_0 \sum_{t=0}^T m_t e_t$. At equilibrium we have $e_t = c_t$.

Proportionality follows immediately from $E \sum_{t=0}^T m_t e_t = E \sum_{t=0}^T m_t c_t$. With log utility we even have

$$P_0 = \frac{1}{1-\delta} c_0.$$

b. Let us first define the return on the wealth portfolio as $\tilde{r}_1 = \frac{P_1 + e_1 - P_0}{P_0}$. Inserting prices and

rearranging gives $\tilde{r}_1 = \frac{\frac{\delta}{1-\delta} c_1 + c_1 - \frac{1}{1-\delta} c_0}{\frac{1}{1-\delta} c_0}$. Defining consumption growth as $\bar{c}_{t+1} = c_{t+1} - c_t$ we

$$\text{get } \tilde{r}_1 = \frac{1}{c_0} \bar{c}_1.$$

c.

$$\begin{aligned} P_0 &= 100 E_0 [m_1 + m_2] \\ &= 100 \delta E_0 \left[\frac{c_0}{c_1} + \delta \frac{c_0}{c_2} \right] \\ &= 100 \delta E_0 \left[\frac{(c_2 - c_1) c_0 + (1 + \delta) c_1 c_0}{c_1 c_2} \right] \\ &= 100 \delta E_0 \left[\left\{ \tilde{r}_2 + (1 + \delta) \right\} \frac{c_0}{c_2} \right] \end{aligned}$$

9.5.

$$\begin{aligned} P_i &= \sum q_s d_{is} \\ 1 &= \sum q_s \frac{d_{is}}{P_i} \\ &= \sum q_s R_{is} \\ &= \sum m_s \pi_s R_{is} \\ &= E(m R_i) \\ &= E(m) E(R_i) + \text{Cov}(m, R_i) \\ &= \frac{E(R_i)}{1 + r_f} + \text{Cov}(m, R_i) \\ R_f &= E(R_i) + \text{Cov}(m, R_i) R_f \\ &\text{and} \\ R_f &= E(R_M) + \text{Cov}(m, R_M) R_f \\ \frac{E(R_i) - R_f}{E(R_M) - R_f} &= \frac{\text{Cov}(m, R_i)}{\text{Cov}(m, R_M)} \\ E(R_i) - R_f &= \frac{\text{Cov}(m, R_i)}{\text{Cov}(m, R_M)} [E(R_M) - R_f] \end{aligned}$$

Note that if $\text{Cov}(m, R_M) = \text{Var}(R_M)$ we have exactly the CAPM equation. This relation holds for example with quadratic utility.

Chapter 10

10.1. a. Markets are complete. Find the state prices from

$$\begin{aligned} 5q_1 + 10q_2 + 15q_3 &= 8 \\ q_1 + q_2 + q_3 &= \frac{1}{1.1} \rightarrow \\ 3q_3 &= 1 \end{aligned} \quad \begin{aligned} q_1 &= 0.55151 \\ q_2 &= 0.02424 \\ q_3 &= 1/3 \end{aligned}$$

b. The put option has a price of $3q_1$. Risk neutral probabilities are derived from

$$\begin{aligned} \pi_1 &= 0.60667 \\ \pi_s &= 1.1q_s \text{ where } s = 1, 2, 3 \quad \pi_2 = 0.02667 \\ \pi_3 &= 0.36667 \end{aligned}$$

c. Consumption at date 0 is 1. The pricing kernel is given by

$$\begin{aligned} m_1 &= 1.83838 \\ m_s &= \frac{q_s}{p_s} \text{ where } s = 1, 2, 3 \quad m_2 = 0.06060 \\ m_3 &= 0.11111 \end{aligned}$$

10.2. $m_s = \frac{q_s}{p_s}$ where $s = 1, 2, 3$ Program for agent 1

$$\max_{c_1^1, c_1^2} \left\{ (10 + q_1 + 5q_2 - c_1^1 q_1 - c_1^2 q_2) + \left(\frac{1}{3} \ln c_1^1 + \frac{2}{3} \ln c_1^2 \right) \right\}$$

The FOC is

$$-q_1 + \frac{1}{3} \frac{1}{c_1^1} = 0$$

$$-q_2 + \frac{2}{3} \frac{1}{c_1^2} = 0$$

and similarly for agent 2. This yields

$$\frac{1}{c_1^1} = \frac{1}{c_2^1}, \frac{1}{c_1^2} = \frac{1}{c_2^2}$$

$$\Leftrightarrow c_1^1 = c_2^1, c_1^2 = c_2^2$$

Using the market clearing conditions we get

$$c_1^1 = c_2^1 = \frac{5}{2}$$

$$c_1^2 = c_2^2 = \frac{11}{2}$$

so that $q_1 = 2/15$ and $q_2 = 4/33$.

Now construct the risk neutral probabilities as follows:

$$\pi_1 = \frac{q_1}{q_1 + q_2}$$

$$\pi_2 = \frac{q_2}{q_1 + q_2}$$

which satisfy the required conditions to be probabilities.

Computation of the risk-free rate is as usual: $q_1 + q_2 = 1/(1+r_f)$.

The market value of endowments can be computed as follows

$$MV_i = \frac{1}{1+r_f} (\pi_1 e_i^1 + \pi_2 e_i^2) \equiv q_1 e_i^1 + q_2 e_i^2 .$$

10.3. Options and market completeness.

The put option has payoffs [1,1,1,0]. The payoff matrix is then

$$m = \begin{pmatrix} 0 & 1 & 0 & 1 \\ 0 & 1 & 1 & 1 \\ 0 & 0 & 1 & 1 \\ 1 & 1 & 1 & 0 \end{pmatrix}$$

Of course, the fourth row gives the payoffs of the put option. We have to solve the system

$$mw = \begin{pmatrix} 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \end{pmatrix}$$

The matrix on the RHS is the A-D securities payoff matrix. The solution is

$$w = \begin{pmatrix} 1 & -2 & 1 & 1 \\ 0 & 1 & -1 & 0 \\ -1 & 1 & 0 & 0 \\ 1 & -1 & 1 & 0 \end{pmatrix}$$

We could also have checked the determinant condition on matrix m, which states that for a square matrix (number of states = number of assets), if the determinant is not null, then the system has a unique solution. Here $\text{Det}(m) = -1$.

- 10.4. a. An A-D security is an asset that pays out 1 unit of consumption in a particular state of the world. The concept is very useful since if are able to extract A-D prices from traded assets they enable us to price every complex security. This statement is valid even if no A-D security is traded. To price a complex security from A-D prices, make up the portfolio of A-D securities providing the same state-by-state payoff as the security to be priced and check what is the cost of this portfolio.

b. Markets are not complete : Determinant of the payoff matrix = 0.

c. No: # of assets < # of states.

Completeness can be reached by adding a put on asset one with strike 12 ($\text{Det} = 126$).

A-D security from calls:

long call on B (strike 5), two short calls on B (strike 6), long call on B (strike 7)

$$\begin{pmatrix} 0 \\ 1 \\ 2 \\ 3 \end{pmatrix} - 2 \begin{pmatrix} 0 \\ 0 \\ 1 \\ 2 \end{pmatrix} + \begin{pmatrix} 0 \\ 0 \\ 0 \\ 1 \end{pmatrix} = \begin{pmatrix} 0 \\ 1 \\ 0 \\ 0 \end{pmatrix}$$

An A-D security with puts:

long put on B (strike 8), two short puts on B (strike 7), long put on B (strike 6)

$$\begin{pmatrix} 3 \\ 2 \\ 1 \\ 0 \end{pmatrix} - 2 \begin{pmatrix} 2 \\ 1 \\ 0 \\ 0 \end{pmatrix} + \begin{pmatrix} 1 \\ 0 \\ 0 \\ 0 \end{pmatrix} = \begin{pmatrix} 0 \\ 0 \\ 1 \\ 0 \end{pmatrix}$$

Chapter 12

12.1. a.

$$EU = 1 + .96(.5 \times \ln 1.2 + .5 \times \ln .833) + (.96)^2 (.25 \times \ln 1.44 + .5 \times \ln 1 + .25 \times \ln .6944) = 0$$

b.

The maximization problem of the representative agent is

$$\max [\ln(c_0) + \delta (\pi_{11} \ln(c_{11}) + \pi_{12} \ln(c_{12})) + \delta^2 (\pi_{21} \ln(c_{21}) + \pi_{22} \ln(c_{22}) + \pi_{23} \ln(c_{23}))]$$

$$\text{s.t. } e_0 + q_{11}e_{11} + q_{12}e_{12} + q_{21}e_{21} + q_{22}e_{22} + q_{23}e_{23} = c_0 + q_{11}c_{11} + q_{12}c_{12} + q_{21}c_{21} + q_{22}c_{22} + q_{23}c_{23}$$

(take the consumption at date 0 as a numeraire, its price is at 1;

q_{ij} is time 0 price of AD security that pays 1 unit of consumption at date i in state j)

The Lagrangian is given by

$$L = EU + \lambda \begin{pmatrix} e_0 + q_{11}e_{11} + q_{12}e_{12} + q_{21}e_{21} + q_{22}e_{22} + q_{23}e_{23} \\ -c_0 + q_{11}c_{11} + q_{12}c_{12} + q_{21}c_{21} + q_{22}c_{22} + q_{23}c_{23} \end{pmatrix}$$

FOC's are:

$$\frac{\partial L}{\partial c_0} = \frac{1}{c_0} - \lambda = 0$$

$$\frac{\partial L}{\partial c_{11}} = \pi_{11} \delta \frac{1}{c_{11}} - \lambda q_{11} = 0$$

.

.

.

$$\frac{\partial L}{\partial c_{23}} = \pi_{23} \delta^2 \frac{1}{c_{23}} - \lambda q_{23} = 0$$

A-D prices, risk neutral probabilities, and the pricing kernel can be derived easily from the FOC's. For example, at date $t = 0$ we have

$$q_{11} = \pi_{11} \delta \frac{1}{c_{11} \lambda} = \pi_{11} \delta \frac{c_0}{c_{11}} = \pi_{11} \frac{MU_{11}}{MU_0} = \pi_{11} m_{11}$$

...

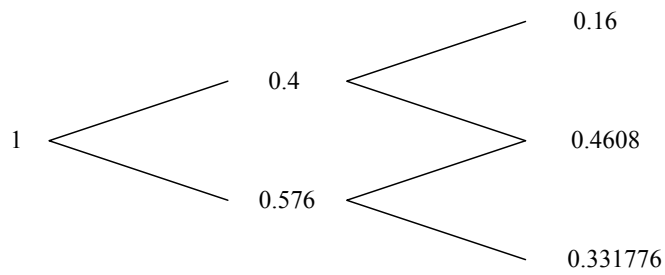
$$q_{23} = \pi_{23} \delta^2 \frac{1}{c_{23} \lambda} = \pi_{23} \delta^2 \frac{c_0}{c_{23}} = \pi_{23} \frac{MU_{23}}{MU_0} = \pi_{23} m_{23}$$

where m_{ij} is the pricing kernel. Risk neutral probabilities at date one are given by:

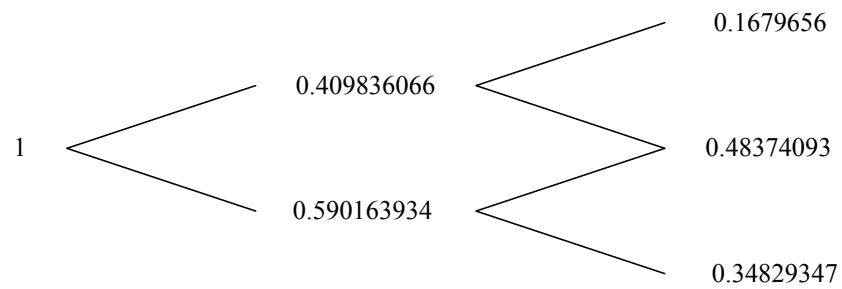
$$\pi_{11}^{RN} = \frac{q_{11}}{q_{11} + q_{12}} \text{ and } \pi_{12}^{RN} = \frac{q_{12}}{q_{11} + q_{12}}$$

$$\pi_{21}^{RN} = \frac{q_{21}}{q_{21} + q_{22} + q_{23}}, \pi_{22}^{RN} = \frac{q_{22}}{q_{21} + q_{22} + q_{23}} \text{ and } \pi_{23}^{RN} = \frac{q_{23}}{q_{21} + q_{22} + q_{23}}$$

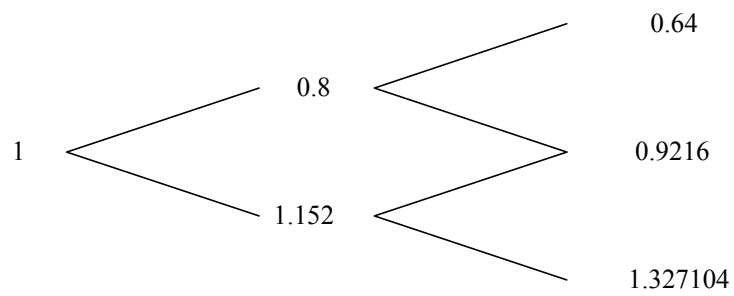
state prices



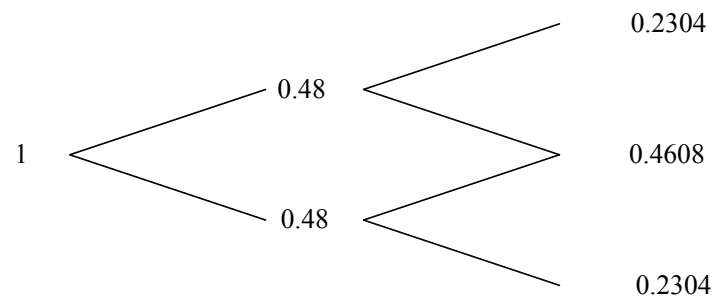
risk neutral probabilities



pricing kernel



c.
valuation



value 2.8816

d.
The one period interest rate at date zero is:

$$r_{0,1} = \frac{1}{(q_{11} + q_{12})} - 1 = 2.459\% .$$

The two period interest rate at date zero is:

$$r_{0,2} = \sqrt{\frac{1}{(q_{21} + q_{22} + q_{23})}} - 1 = 2.459\% .$$

Even though the economy is stochastic with log utility there is no term premia.

The price of a one period bond is $q_b(1) = \frac{1}{1.02459} = .976$ and the price of a two period

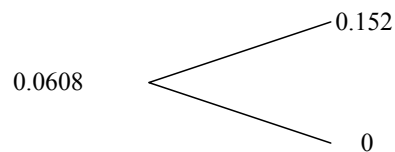
bond is $q_b(2) = \frac{1}{(1.02459)^2} = .953$.

e. The valuation of the endowment stream is

	2.8816	2.352	1.44
price	1.8816	1.152	
space		1.6333333	1
		0.8	
			0.69444444

At date one and two we have one value with payoffs (upper cell) and the value after the cash flow arrived (lower cell).

The value of the option, using either state prices, pricing kernel, or risk neutral valuation, is
option value



f.

The price process is as in e. Now we need to solve for u, d, R, and risk neutral probabilities.

$$u = \frac{2.352}{1.8816} = \frac{1.44}{1.152} = 1.25$$

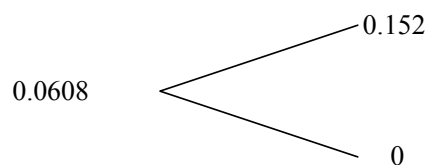
$$d = \frac{1.6333}{1.8816} = .8681$$

$$R = 1 + r = 1.02459$$

$$q_{11} = \frac{R - d}{u - d} = \frac{1.02459 - .8681}{1.25 - .8681} = .4098$$

(Compare this value with q_{11} in b))

The value of the option is
option value

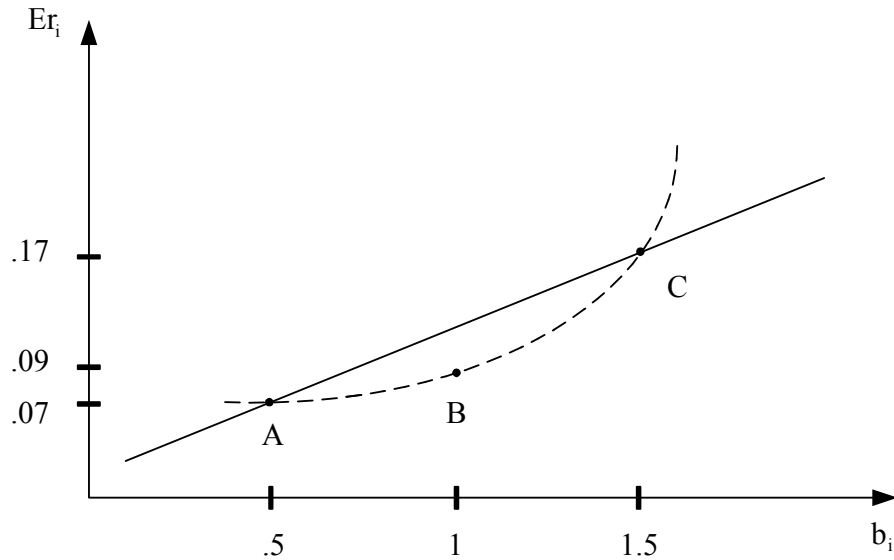


g.

In part b we saw that pricing via A-D prices, risk-neutral probabilities, and pricing kernel are essentially the same. These methods rely on the payoffs of the endowment stream. In contrast to b, risk neutral probabilities are elicited in part f from the price process. Of course, the risk-neutral probabilities are the same as in b. This is not surprising since prices are derived from utility maximization of the relevant cash flows. Thus risk-neutral probabilities of the cash flow stream coincide with the risk neutral probabilities of the price of the asset.

Chapter 13

13.1 a.



b. Using A, B ; we want

$$b_p = 0 = w_A b_A + (1 - w_A) b_B$$

$$0 = w_A (.5) + (1 - w_A)(1)$$

$$.5w_A = 1; w_A = 2, w_B = -1$$

Using B, C :

$$b_p = 0 = w_B b_B + w_C b_C$$

$$0 = w_B (1) + w_C (1.5)$$

$$0 = w_B + 1.5 - 1.5w_B$$

$$.5w_B = 1.5; w_B = 3, w_C = -2$$

c. We need to find the proportions of A and C that give the same b as asset B.

$$\text{Thus, } b_B = 1 = w_A b_A + (1 - w_A) b_C$$

$$1 = w_A (.5) + (1 - w_A)(1.5)$$

$$\Rightarrow w_A = \frac{1}{2}; w_C = \frac{1}{2}$$

With these proportions :

$$Er_p = \frac{1}{2} Er_A + \frac{1}{2} Er_C$$

$$= \frac{1}{2} (.07) + \frac{1}{2} (.17) = .12 > .09 = Er_B$$

$$r_p = .12 + 1F_1 + e_p$$

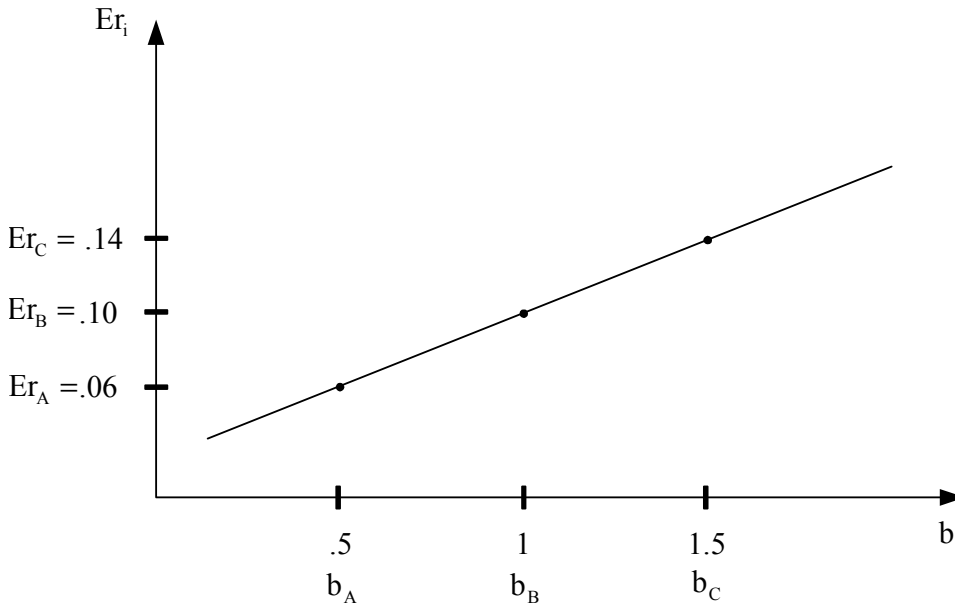
$$R_{p_{w_B=1}} = .09 + 1F_1 + e_B, \quad \text{cov}(e_p, e_B) \equiv 0.$$

Now we assume these assets are each well diversified portfolios so that $e_p = e_B \equiv 0$.

An arbitrage portfolio consist of shorting B and buying the portfolio composed of $w_A = w_C = \frac{1}{2}$ in equal amounts, you will earn 3% riskless.

d. As a result the prices of A, C will rise and their expected returns fall. The opposite will happen to B.

e.



There is no longer an arbitrage opportunity : expected returns are consistent with relative systematic risk.

13.2. a. Since $\text{cov}(\tilde{r}_M, \epsilon_j) = 0$,

$$\begin{aligned}\sigma_j^2 &= \text{var}(\alpha_j) + \text{var}(\beta_{jM} \tilde{r}_M) + \text{var}(\epsilon_j) \\ &= 0 + \beta_{jM}^2 \sigma_M^2 + \sigma_{\epsilon_j}^2 \\ &= \beta_{jM}^2 \sigma_M^2 + \sigma_{\epsilon_j}^2\end{aligned}$$

$$\begin{aligned}\text{b. } \sigma_{ij} &= \text{cov}(\alpha_i + \beta_{iM} \tilde{r}_M + \tilde{\epsilon}_i, \alpha_j + \beta_{jM} \tilde{r}_M + \tilde{\epsilon}_j) \\ &= \text{cov}(\beta_{iM} \tilde{r}_M + \tilde{\epsilon}_i, \beta_{jM} \tilde{r}_M + \tilde{\epsilon}_j) \quad (\text{constants do not affect covariances}) \\ &= \text{cov}(\beta_{iM} \tilde{r}_M + \tilde{\epsilon}_i, \beta_{jM} \tilde{r}_M) + \text{cov}(\beta_{iM} \tilde{r}_M + \tilde{\epsilon}_i, \tilde{\epsilon}_j) \quad (\text{since } \tilde{r}_M, \tilde{\epsilon}_j \text{ are independent}) \\ &= \text{cov}(\beta_{iM} \tilde{r}_M, \beta_{jM} \tilde{r}_M) + \text{cov}(\tilde{\epsilon}_i, \beta_{jM} \tilde{r}_M) + \text{cov}(\beta_{iM} \tilde{r}_M, \tilde{\epsilon}_j) + \text{cov}(\tilde{\epsilon}_i, \tilde{\epsilon}_j) \quad (\text{since } \tilde{r}_M, \tilde{\epsilon}_i \text{ are independent}) \\ &= \text{cov}(\beta_{iM} \tilde{r}_M, \beta_{jM} \tilde{r}_M) ; \text{ since by construction of the regression relationship all other covariances are zero.} \\ &= \beta_{iM} \beta_{jM} \text{cov}(\tilde{r}_M, \tilde{r}_M) = \beta_i \beta_j \sigma_M^2\end{aligned}$$

- 13.3. The CAPM model is an equilibrium model built on structural hypotheses about investors' preferences and expectations and on the condition that asset markets are at equilibrium. The APT observes market prices on a large asset base and derives, under the hypothesis of no arbitrage, the implied relationship between expected returns on individual assets and the expected returns on a small list of fundamental factors.
- Both models lead to a linear relationship explaining expected returns on individual assets and portfolios. In the case of the CAPM, the SML depends on a single factor, the expected excess return on the market portfolio. The APT opens up the possibility that more than one factor are priced in the market and are thus necessary to explain returns. The return on the market portfolio could be one of them, however. Both models would be compatible if the market portfolio were simply another way to synthesize the several factors identified by the APT: under the conditions spelled out in section 12.4, the two models are essentially alternative ways to reach the same 'truth'. Empirical results tend to suggest, however, that this is not likely to be the case. Going from expected returns to current price is straightforward but requires formulating, alongside expectations on future returns, expectations on the future price level and on dividend payments.
- 13.4. The main distinction is that the A-D theory is a full structural general equilibrium theory while the APT is a no-arbitrage approach to pricing. The former prices all assets from assumed primitives. The latter must start from the observations of quoted prices whose levels are not explained.
- The two theories are closer to one another, however, if one realizes that one can as well play the 'no arbitrage' game with A-D pricing. This is what we did in Chapter VIII. There the similarities are great: start from given unexplained market prices for 'complex' securities and extract from them the prices of the fundamental securities. Use the latter for pricing other assets or arbitrary cash flows.
- The essential differences are the following: A-D pricing focuses on the concept of states of nature and the pricing of future payoffs conditional on the occurrence of specific future states. The APT replaces the notion of state of nature with the 'transversal' concept of factor. While in the former the key information is the price of one unit of consumption good in a specific future date-state, in the latter the key ingredient extracted from observed prices is the expected excess return obtained for bearing one unit of a specified risk factor.
- 13.5. True. The APT is agnostic about beliefs. It simply requires that the observed prices and returns, presumably the product of a large number of agents trading on the basis of heterogeneous beliefs, are consistent in the sense that no arbitrage opportunities are left unexploited.

Chapter 15

- 15.1. a. These utility functions are well known. Agent 1 is risk-neutral, agent 2 is risk-averse.
b. A PO allocation is one such that agent 2 gets smooth consumption.
c. Given that agent 2 is risk-averse, he buys A-D₁ and sells AD₂, and gets a smooth consumption; Agent 1 is risk-neutral and is willing to buy or sell any quantity of A-D securities. We can say agent 2 determines the quantities, and agent 1 determines the prices of the AD securities.

Solving the program for agent 1 gives the following FOC:

$$q_1 = \delta\pi$$

$$q_2 = \delta(1 - \pi)$$

The price of AD securities depends only on the probability of each state.

Agent 2's optimal consumption levels are $c^2 = c^2(\theta_1) = c^2(\theta_2) = (2\delta(1 - \pi) + 1)/(1 + \delta)$ which is 1 if $\pi = 0.5$.

d. Note: it is not possible to transfer units of consumption across states. Price of the bond is δ . Allocation will not be PO.

Available security

t=0	t=1	
	θ_1	θ_2
$-p^b$	1	1

Let the desired holdings of this security by agents 1 and 2 be denoted by Q_1 and Q_2 respectively.

Agent maximization problems :

$$\text{Agent 1 : } \max_{Q_1} (1 - Q_1 p^b) + \delta \{ \pi(1 + Q_1) + (1 - \pi)(1 + Q_1) \}$$

$$\text{Agent 2 : } \max_{Q_2} \ln(1 - Q_2 p^b) + \delta \{ \pi \ln Q_2 + (1 - \pi) \ln(2 + Q_2) \}$$

The F.O.C.s are :

$$1. \text{ Agent 1 : } -p^b + \delta = 0 \text{ or } p^b = \delta.$$

$$\text{Agent 2 : } \frac{1}{1 - Q_2 p^b} (-p^b) + \delta \left\{ \frac{\pi}{Q_2} + \frac{(1 - \pi)}{2 + Q_2} \right\} = 0$$

2. We know also that $Q_1 + Q_2 = 1$ in competitive equilibrium in addition to these equations being satisfied. Substituting $p^b = \delta$ into the 2nd equation yields, after simplification,

$$\begin{aligned} (1 + \delta)(Q_2)^2 + (1 + 2\pi\delta)Q_2 - 2\pi &= 0 \\ Q_2 &= \frac{-(1 + 2\pi\delta) \pm \sqrt{(1 + 2\pi\delta)^2 - 4(1 + \delta)(-2\pi)}}{2(1 + \delta)} \\ &= \frac{-(1 + 2\pi\delta) \pm \sqrt{1 + 4\pi^2\delta^2 + 4\pi\delta + 8\pi + 8\pi\delta}}{2(1 + \delta)} \end{aligned}$$

$$Q_2 = \frac{-(1 + 2\pi\delta) \pm \sqrt{(1 + 2\pi\delta)^2 + 8\pi + 8\pi\delta}}{2(1 + \delta)}$$

$$\text{with } Q_1 = 1 - Q_2$$

3. Suppose $\pi = .5$ and $\delta = 1/3$

$$Q_2 = \frac{-(4/3) \pm \sqrt{(4/3)^2 + 8/6 + 4}}{2(4/3)}$$

$$= -\frac{1}{2} \pm \frac{3}{8} \left(\frac{8}{3}\right)$$

$$Q_2 = \frac{1}{2} \quad (\text{We want the positive root. Otherwise agent 2 will have no consumption in the } \theta = 1 \text{ state})$$

$$Q_1 = \frac{1}{2}.$$

15.2 When markets are incomplete :

- i) MM does not hold: the value of the firm may be affected by the financial structure of the firm.
- ii) It may not be optimal for the firm's manager to issue the socially preferable set of financial instruments

15.3 a. Write the problem of a risk neutral agent :

$$\text{Max } 30 - p^Q Q^2 + 1/3 \cdot 1/2 (15 + Q^2) + 2/3 \cdot 1/2 \cdot 15$$

$$\text{FOC: } -p^Q + 1/6 = 0 \text{ thus } p^Q = 1/6 \text{ necessarily!}$$

This is generic: risk neutrality implies no curvature in the utility function. If the equilibrium price differs from 1/6, the agent will want to take infinite positive or negative positions!
At that price, check that the demand for asset Q by agent 1 is zero:

$$\text{Max } 20 - 1/6 Q^1 + 1/2 \cdot 1/3 \ln(1 + Q^1) + 1/2 \cdot 2/3 \ln(5)$$

$$\text{FOC: } -1/6 + 1/6 \cdot 1/(1 + Q^1) = 0$$

$$Q^1 = 0$$

Thus there is no risk sharing. The initial allocation is not Pareto – optimal. The risk averse agent is exposed to significant risk at date 1. If the state probabilities were 1/2, nothing would change, except that the equilibrium price becomes:

$$p^Q = 1/4$$

b. $p^Q = 1/6$, $Q^1 = 0$ (the former FOC is not affected), $p^R = 1/3$

FOC of agent 1 wrt R:

$$-1/3 + (1/2)2/3 \cdot 1/(5 + R^1) = 0$$

$$1 = 5 + R^1$$

$$R^1 = -4$$

So agent 1 sells 4 units of asset R. He reduces his $t = 1$ risk. At date 1, he consumes 1 unit in either state. He is compensated by increasing his date 0 consumption by $p^R R^1 = 1/3(4) = 4/3$. The allocation is Pareto optimal, as expected from the fact that markets are now complete.

Post trade allocation:

	t = 0	t = 1	t = 2
Agent 1	20 4/3	1	1
Agent 2	28 2/3	15	19

15.4

	t=0	t=1	
		$\theta = 1$	$\theta = 2$
Agent 1	4	6	1
Agent 2	6	3	4

$$U^1(c_0, \tilde{c}_1(\theta)) = \frac{1}{2} \ln c_0^1 + E \ln c_1^1(\theta)$$

$$U^2(c_0, \tilde{c}_1(\theta)) = \frac{1}{2} \ln c_0^2 + E \ln c_1^2(\theta)$$

$$\text{Prob}(\theta_1) = .4 \quad \text{Prob}(\theta_2) = .6$$

a. Initial utilities –

$$\begin{aligned} \text{Agent 1:} \quad U^1(c_0, \tilde{c}_1(\theta)) &= \frac{1}{2} \ln(4) + .4 \ln(6) + .6 \ln(1) \\ &= \frac{1}{2} (1.386) + .4(1.79) \\ &= .693 + .716 = 1.409 \end{aligned}$$

$$\begin{aligned} \text{Agent 2:} \quad U^2(c_0, \tilde{c}_1(\theta)) &= \frac{1}{2} \ln(6) + .4 \ln(3) + .6 \ln(4) \\ &= 3 + .439 + .832 \\ &= 4.271 \end{aligned}$$

c. Firm's output

$$\begin{array}{ccc}
 & t=1 & \\
 \theta=1 & & \theta=2 \\
 -p & \frac{1}{2} & \frac{1}{3}
 \end{array}$$

Agent 1's Problem (presuming only this security is issued).

$$\max_{Q_1} \frac{1}{2} \ln(4 - pQ_1) + .4 \ln(6 + 2Q_1) + .6 \ln(1 + 3Q_1)$$

Agent 2's Problem

$$\max_{Q_2} \frac{1}{2} (6 - pQ_2) + .4 \ln(3 + 2Q_2) + .6 \ln(4 + 3Q_2)$$

The F.O.C.'s are:

$$\text{Agent 1: } \frac{1}{2} \left(\frac{p}{4 - pQ_1} \right) = .4 \left(\frac{1}{6 + 2Q_1} \right) (2) + .6 \left(\frac{1}{1 + 3Q_1} \right) (3)$$

$$\text{Agent 2: } \frac{1}{2} p = .4 \left(\frac{1}{3 + 2Q_2} \right) (2) + .6 \left(\frac{1}{4 + 3Q_2} \right) (3)$$

$$Q_2 = 1 - Q_1$$

These can be simplified to

$$(i) \quad \frac{p}{4 - pQ_1} = \frac{1.6}{6 + 2Q_1} + \frac{3.6}{1 + 3Q_1}$$

$$(ii) \quad p = \frac{1.6}{5 - 2Q_1} + \frac{3.6}{7 - 3Q_1}$$

The solution to this set equations via matlab is

$$\begin{aligned}
 p &= 1.74 \\
 Q_1 &= 1.245
 \end{aligned}$$

Thus $Q_2 = 1 - 1.245 = -.245$ (short sale).

Thus $V_F = 1.74$.

The post-trade allocations are

	t=0	$\theta = 1$	t=1 $\theta = 2$
Agent 1:	$4 - (1.74)(1.245) = 1.834$	$6 + 2(1.245) = 8.49$	$1 + 3(1.245) = 4.735$
Agent 2:	$6 - (1.74)(-.245) = 6.426$	$3 - 2(.245) = 2.51$	$4 - 3(.245) = 3.265$

Post-trade (ex ante) utilities:

$$\begin{aligned}
 \text{Agent 1: } & \frac{1}{2} \ln(1.834) + .4 \ln(8.49) + .6 \ln(4.735) \\
 & = .3032 + .4(2.14) + .6(1.555) \\
 & = .3032 + .856 + .933 \\
 & = 2.0922
 \end{aligned}$$

$$\begin{aligned}
 \text{Agent 2: } & \frac{1}{2} \ln(6.426) + .4 \ln(2.51) + .6 \ln(3.265) \\
 & = 3.213 + .368 + .70996 \\
 & = 4.291
 \end{aligned}$$

Nearly all the benefit goes to agent 1. This is not entirely surprising as the security payoffs are more useful to him for consumption smoothing.

For agent 2, the marginal utility of a unit of consumption in period 1 is less than the marginal utility of a unit in period 0. His consumption pattern across states in $t=1$ is also relatively smooth, and the security available for sale is not particularly useful in correcting the existing imbalance. Taken together, he is willing to “sell short” the security or, equivalently, to borrow against the future.

The reverse is true for agent 1 especially on the issue of consumption smoothing across $t=1$ states: he has very little endowment in the more likely state. Furthermore the security pays relatively more in this particular state. Agent 1 thus wishes to save and acquires “most” of the security.

If the two states were of equal probability agent 1 would have a bit less need to smooth, and thus his demand would be relatively smaller. We would expect p to be smaller in this case.

c. The Arrow-Debreu securities would offer greater opportunity for risk sharing among the agents without the presence of the firm. (We would expect V_F to be less than in b).

However, each agent would most likely have a higher utility ex ante (post-trade).

d. Let the foreign government issue 1 unit of the bond paying (2.2); let its price be p .

Agent Problems:

$$\text{Agent 1: } \max_{Q_1} \frac{1}{2} \ln(4 - pQ_1) + .4 \ln(6 + 2Q_1) + .6 \ln(1 + 2Q_1)$$

$$\text{Agent 2: } \max_{Q_2} \frac{1}{2}(4 - pQ_2) + .4 \ln(3 + 2Q_2) + .6 \ln(4 + 2Q_2)$$

Where, in equilibrium $Q_1 + Q_2 = 1$

F.O.C's:

$$\text{Agent 1: } \frac{1}{2} \left(\frac{p}{4 - pQ_1} \right) = \frac{.4(2)}{6 + 2Q_1} + \frac{.6(2)}{1 + 2Q_1}$$

$$\text{Agent 2: } \frac{1}{2} p = \frac{.4(2)}{3 + 2Q_2} + \frac{.6(2)}{4 + 2Q_2}$$

Substituting $Q_2 = 1 - Q_1$, these equations become:

$$\frac{p}{4 - pQ_1} = \frac{1.6}{6 + 2Q_1} + \frac{2.4}{1 + 2Q_1}$$

$$p = \frac{1.6}{5 - 2Q_1} + \frac{2.4}{6 - 2Q_1}$$

Solving these equations using matlab yields

$$p = 1.502$$

$$Q_1 = 1.4215$$

$$\text{and thus } Q_2 = -.4215$$

The bond issue will generate $p = 1.502$.

Post Trade allocations:

	t=0	t=1	t=1
		$\theta = 1$	$\theta = 2$
Agent 1:	$4 - (1.502)(1.4215) = 1.865$	$6 + 2(1.4215) = 8.843$	$1 + 2(1.4215) = 3.843$
Agent 2:	$6 - (1.502)(-.4215) = 6.633$	$3 + 2(-.4215) = 2.157$	$4 + 2(-.4215) = 3.157$

The utilities are:

$$\begin{aligned}\text{Agent 1 : } & \frac{1}{2} \ln(1.865) + .4 \ln(8.843) + .6 \ln(3.843) \\ & = .3116 + .8719 + .8078 \\ & = 1.9913\end{aligned}$$

$$\begin{aligned}\text{Agent 2 : } & \frac{1}{2} (6.633) + .4 \ln(2.157) + .6 \ln(3.157) \\ & = 3.3165 + .3075 + .690 \\ & = 4.314 .\end{aligned}$$

Once again; both agents are better off after trade. Most of the benefits still go to agent 1; however, the incremental benefit to him is less than in the prior situation because the security is less well situated to his consumption smoothing needs.

e. When the bond is issued by the local government, one should specify i) where the proceeds from the bond issue go, and ii) how the $t=1$ payments in the bond contracts will be financed. In a simple closed economy, the most natural assumption is that the proceeds from the issue are redistributed to the agents in the economy and similarly that the payments are financed from taxes levied on the same agents. If these redistributive payments and taxes are lump-sum transfers, they will not affect the decisions of individuals, nor the pricing of the security. But the final allocation will be modified and closer (equal ?) to the initial endowments. In more general contexts, these payments may have distortionary effects.

15.5 a.

$$\begin{aligned}\text{Agent 1 : } & \max \frac{1}{2} x_1^1 + \frac{1}{2} x_2^1 \\ \text{s.t.} & \\ & q_1 x_1^1 + q_2 x_2^1 \leq q_1 e_1^1 + q_2 e_2^1\end{aligned}$$

$$\begin{aligned}\text{Agent 2 : } & \max \frac{1}{2} \ln x_1^2 + \frac{1}{2} \ln x_2^2 \\ \text{s.t.} & \\ & q_1 x_1^2 + q_2 x_2^2 \leq q_1 e_1^2 + q_2 e_2^2\end{aligned}$$

Substituting the budget constraint into the objective function :

$$\text{Agent 1 : } \max_{x_2^1} \frac{1}{2} \left(\frac{q_1 e_1^1 + q_2 e_2^1 - q_2 x_2^1}{q_1} \right) + \frac{1}{2} \ln x_2^1$$

$$\text{Agent 2 : } \max_{x_2^2} \frac{1}{2} \ln \left(\frac{q_1 e_1^2 + q_2 e_2^2 - q_2 x_2^2}{q_1} \right) + \frac{1}{2} \ln x_2^2$$

FOC's

$$\text{Agent 1 : } \frac{1}{2} \left(\frac{-q_2}{q_1} \right) + \frac{1}{2} = 0 \Rightarrow q_1 = q_2$$

Agent 2 : $\frac{1}{2} \left(\frac{q_1}{q_1 e_1^2 + q_2 e_2^2 - q_2 x_2^2} \right) \left(\frac{q_2}{q_1} \right) = \frac{1}{2} \left(\frac{1}{x_2^2} \right)$ which, taking into account that the two prices are equal, solves for

$$x_2^2 = x_1^2 = \frac{e_1^2 + e_2^2}{2}, \text{ i.e.,}$$

agent 2's consumption is fully stabilized.

b. Each agent owns one half of the firm, which can employ simultaneously two technologies :

	t=0	t=1	
		$\theta = 1$	$\theta = 2$
<u>Technology 1</u>	-y	y	y
<u>Technology 2</u>	-y	3y	0

Let x be the portion of input invested in technology 2. Since there are 2 units invested in total, 2-x is invested in technology 1. In total we have :

	t=0	t=1	
		$\theta = 1$	$\theta = 2$
<u>Invested in tech. 1</u>	2-x	2-x	2-x
<u>Invested in tech. 2</u>	x	3x	0
<u>Each firm owner receives</u>		1+x	1 - $\frac{1}{2}$ x

Considering that the two Arrow-Debreu prices necessarily remain equal, agent 2 solves

$$\max_{x, x_2^2} \frac{1}{2} \ln(2 + \frac{1}{2} x + e_1^2 + e_2^2 - x_2^2) + \frac{1}{2} \ln x_2^2$$

It is clear that he wants x to be as high as possible, that is, x = 2.

The FOC wrt x_2^2 solves for

$$x_2^2 = x_1^2 = \frac{2 + \frac{1}{2} x + e_1^2 + e_2^2}{2}.$$

Again there is perfect consumption insurance for the risk averse agent (subject to feasibility, that is, an interior solution).

Agent 1 solves

$$\max_x \frac{1}{2} (2 + \frac{1}{2} x + e_1^1 + e_2^1 - x_2^1) + \frac{1}{2} x_2^1$$

Clearly he also wants x to be as high as possible. So there is agreement between the two firm owners to invest everything ($x = 2$) in the second more productive but riskier technology. For agent 1, this is because he is risk neutral. Agent 2, on the other hand, is fully insured, thanks to complete markets. Given that fact, he also prefers the more productive technology even though it is risky.

c. There cannot be any trade in the second period ; agents will consume their endowments at that time.

Agent 1 solves

$$\max_x \frac{1}{2}(1+x+e_1^1) + \frac{1}{2}(1-\frac{1}{2}x+e_2^1) =$$

$$\max_x 1 + \frac{1}{4}x + \frac{e_1^1 + e_2^1}{2} ; \text{ clearly he still wants to invest as much as possible in technology 2.}$$

Agent 2 solves

$$\max_x \frac{1}{2} \ln(1+x+e_1^2) + \frac{1}{2} \ln(1-\frac{1}{2}x+e_2^2)$$

which, after derivation, yields

$$x = \frac{1 + 2e_2^2 - e_1^2}{2}.$$

That is, the (risk averse) agent 2 in general wants to invest in the risk-free technology. There is thus disagreement among firm owners as to the investment policy of the firm. This is a consequence of the incomplete market situation.

d. The two securities are now

	t=1	
	$\theta = 1$	$\theta = 2$
<u>A bond</u>	1	1
<u>Technology 2</u>	$1+x$	$1-\frac{1}{2}x$

These two securities can replicate (1,0) and (0,1). To replicate (1,0), for instance, invest a in the bond and b in the firm where a and b are such that

$$a + (1+x)b = 1$$

$$a + (1-\frac{1}{2}x)b = 0$$

This system implies $b = \frac{2}{3}x$ and $a = 1 - \frac{2}{3}x(1+x)$.

Given that the markets are complete, both agents will agree to invest $x=2$. Thus $b = \frac{4}{3}$ and $a=0$ replicate (1,0).

Chapter 16

- 16.1. The maximization problem for the speculator's is:

$$\max_f EU[c^* + (p^f - p)f]$$

Let us rewrite the program in the spirit of Chapter IV:

$W(f) = E\{U(c^* + (p^f - p)f)\}$. The FOC can then be written

$W'(f) = E\{U'(c^* + (p^f - p)f)(p^f - p)\} = 0$. From $(U'' < 0)$ we find

$W''(f) = E\{U''(c^* + (p^f - p)f)(p^f - p)^2\} < 0$. This means that $f > 0$ iff

$W'(0) = \{U'(c^*)\}E(p^f - p) > 0$. From $U' > 0$ we have $f > 0$ iff $E(p^f - p) > 0$. The two other cases follow immediately.

- 16.2. Let us reason with the help of an example. Suppose the current futures price p^f is \$100. The marginal cost of producing the corresponding commodity is \$110. The producer's expectation as to the $t=1$ spot price for her output is \$120.

The producer could speculate in the sense of deciding to produce an extra unit of output since she expects to be able to sell it tomorrow at a price that covers her marginal cost with a margin (\$10) that we can even assume sufficient to compensate for the risk being taken.

The rule we have derived in this chapter would, however, suggest that this is the wrong decision. Under the spelled out hypotheses, the futures price is the definite signal of production (Equation 16.3).

It is easy to see, indeed, that if the producer wants to take a position on the basis of her expectations on the $t=1$ spot price, she will make a larger unit profit by speculating on the futures market rather than on the spot market. This is because the cost of establishing a 'spot' speculative position is the cost of producing the commodity, that is, \$110 while the cost of establishing a speculative position on the futures market is \$100 only. If the producers' expectations turn out to be right, she will make \$20 on a unit speculative futures position while she will make only \$10 by speculating on the spot market, that is, by producing the commodity. Note that this reasoning is generic as long as the futures price is below the cost of production. If this was not the case, producing would not involve any speculation in the sense that the output could be sold profitably on the futures market at the known futures price.

- 16.3 When the object exchanged is a financial asset and investors have heterogeneous information, a price increase may reveal some privately held information leading some buyers and sellers to reevaluate their positions. As a consequence, the increase in price may well lead to a fully rational increase in demand. In that sense the law of demand does not apply in such a context.
- 16.4 When investors share the same information, the only source of trading are liquidity needs of one or another group of investors. In the CCAPM with homogeneous investors, there simply is no trading. Prices support the endowment allocations. In a world of heterogeneous information, the liquidity-based trades are supplemented by exchanges between investors who disagree (and in

some sense agree to disagree) about the fundamental valuation of the object being exchanged: I sell because I believe at the current price the stock of company A is overvalued knowing that the investor who buys from me is motivated by the opposite assessment. Obviously trading volume is necessarily higher in the second world, which appears to be closer to the one we live in than the former.

16.5

$$E\tilde{\pi} = p y \bar{e} - \frac{1}{2} y^2$$

$$\text{var } \tilde{\pi} = p^2 y^2 \text{var } \tilde{e} + z^2 \text{var } \tilde{q} - 2 z p y \text{cov}(\tilde{e}, \tilde{q})$$

FOC:

$$y: \quad p \bar{e} - y - \gamma [p^2 y \text{var } \tilde{e} - z p \text{cov}(\tilde{e}, \tilde{q})] = 0$$

This is a maginal product = marginal cost condition where the latter includes a risk premium.

$$z: \quad z \text{var } \tilde{q} - p y \text{cov}(\tilde{e}, \tilde{q}) = 0$$

$$z = \frac{p y \text{cov}(\tilde{e}, \tilde{q})}{\text{var } \tilde{q}}$$

$p y$ = value placed at risk by fluctuations of exchange rates.

Optimal position in hedging instrument 2 is the product of $p y$ by a coefficient that we may call β , which is the sensitivity of \tilde{e} to changes in the price of the hedging instrument as in

$$\tilde{e} = \alpha + \beta \tilde{q} + \tilde{\varepsilon}.$$

If the coefficient β is 1, the optimal position in the hedging instrument is exactly equal to the value placed at risk $p y$. Given the value of the optimal z , the FOC wrt y can be written as:

$$p \bar{e} - y - \gamma p^2 y \text{var } \tilde{e} \left[1 - \frac{[\text{cov}(\tilde{e}, \tilde{q})]^2}{\text{var } \tilde{q} \text{var } \tilde{e}} \right] = 0$$

$$p \bar{e} - y - \gamma p^2 y \text{var } \tilde{e} [1 - R^2] = 0$$

where R^2 is the correlation coefficient in the above regression of \tilde{q} on \tilde{e} .

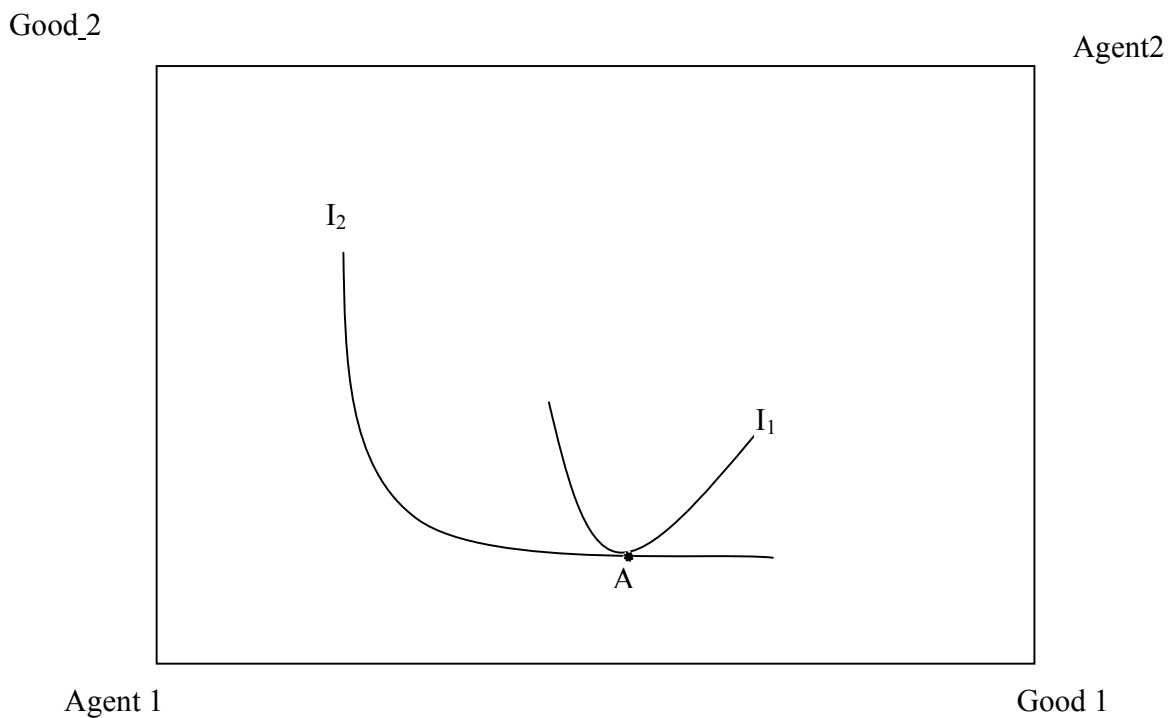
Intermediate Financial Theory
Danthine and Donaldson

Additional Exercises

Chapter 1

- 1.6. Assume a 2 goods-2 agents economy and well-behaved utility functions. Explain why the competitive equilibrium should be on the contract curve.
- 1.7. What is unusual in Figure 1 below? Is there a PO allocation ? Can it be obtained as a competitive equilibrium ? What is the corresponding assumption of the 2nd theorem of welfare, and why is it important?

Figure 1.2 : The Edgeworth-Bowley Box: An Unusual Configuration



Chapter 4

- 4.9. If you are exposed to a 50/50 probability of gaining or losing CHF 1'000 and an insurance that removes the risk costs CHF 500, at what level of wealth will you be indifferent between taking the gamble or paying the insurance? That is, what is your certainty equivalent wealth for this gamble? Assume that your utility function is $U(Y) = -1/Y$.
What would the solution be if the utility function were logarithmic?
- 4.10. Assume that you have a logarithmic utility function on wealth $U(Y) = \ln Y$ and that you are faced with a 50/50 probability of winning or losing CHF 1'000. How much will you pay to avoid this risk if your current level of wealth is CHF 10'000? How much would you pay if your level of wealth is CHF 1'000'000? Did you expect that the premium you were willing to pay would increase/decrease? Why?
- 4.11. Assume that security returns are normally distributed. Compare portfolios A and B, using both first and second-order stochastic dominance :

Case 1	Case 2	Case 3
$\sigma_a > \sigma_b$	$\sigma_a = \sigma_b$	$\sigma_a < \sigma_b$
$E_a = E_b$	$E_a > E_b$	$E_a < E_b$

- 4.12. An agent faces a risky situation in which he can lose some amount of money with probabilities given in the following table:

Loss	Probability
1000	10%
2000	20%
3000	35%
5000	20%
6000	15%

This agent has a utility function of wealth of the form

$$U(Y) = \frac{Y^{1-\gamma}}{1-\gamma} + 2$$

His initial wealth level is 10000 and his γ is equal to 1.2.

a. Calculate the certainty equivalent of this prospect for this agent. Calculate the risk premium. What would be the certainty equivalent of this agent if he would be risk neutral?

b. Describe the risk premium of an agent whose utility function of wealth has the form implied by the following properties: $U'(Y) > 0$ and $U''(Y) < 0$

- 4.13. An agent with a logarithmic utility function of wealth tries to maximize his expected utility. He faces a situation in which he will incur a loss of L with probability π . He has the possibility to insure against this loss. The insurance premium depends on the extent of the coverage. The amount covered is denoted by h and the price of the insurance per unit of coverage is p (hence the amount he has to spend on the insurance will be hp).

- a. Calculate the amount of coverage h demanded by agent as a function of his wealth level Y , the loss L , the probability π and the price of the insurance p .
- b. What is the expected gain of an insurance company offering such a contract ?
- c. If there is perfect competition in the insurance market (zero profit), what price p will the insurance company set?
- d. What amount of insurance will the agent buy at the price calculated under c. What is the influence of the form of the utility function ?

4.14. Given the following probability distributions for risky payoffs \tilde{x} and \tilde{z} :

x	Probability (x)	z	Probability (z)
-10	.1	2	.2
5	.4	3	.5
10	.3	4	.2
12	.2	30	.1

- a. If the only available choice is 100 % of your wealth in \tilde{x} or 100 % in \tilde{z} and you choose on the basis of mean and variance, which asset is preferred ?
 - b. According to the second-order stochastic dominance criterion, how would you compare them ?
- 4.15. There is an individual with a well-behaved utility function, and initial wealth Y . Let a lottery offer a payoff of G with probability π and a payoff of B with probability $1-\pi$.
- a. If the individual already owns this lottery denote the minimum price he would sell it for by P_s . Write down the expression P_s has to satisfy.
 - b. If he does not own it, write down the expression P_b (the maximum price he would be willing to pay for it) has to satisfy.
 - c. Assume now that $\pi=1/2$, $Y=10$, $G=6$, $B=26$, and the utility function is $U(Y)=Y^{1/2}$. Find buying and selling prices. Are they equal? Explain why not. Generally, can they ever be equal?

4.16. Consider the following investments:

Investment 1		Investment 2	
Payoff	Prob.	Payoff	Prob.
1	0.25	3	0.33
7	0.50	5	0.33
12	0.25	8	0.34

Check that neither investment dominates the other on the basis of

- The Mean-Variance criterion
- First Order Stochastic Dominance
- Second Order Stochastic Dominance

How could you rank these investments ?

Chapter 5

5.6. An individual with a utility function $U(c) = -\exp(-Ac)$ (where $A = (1/30)\ln 4$) and an initial wealth of \$50 must choose a portfolio of two assets. Each asset has a price of \$50. The first asset is riskless and pays off \$50 next period in each of the two possible states. The risky asset pays off z_s in state $s=1,2$. Suppose also that the individual cares only about next period consumption (denoted by c_1 or c_2 depending on the state). The probability of state 1 is denoted by π .

a. If the individual splits his wealth equally between the two assets, fill in the following table assuming that each of three scenarios is considered.

Scenario	(z_1, z_2)	π	(c_1, c_2)	$E(c)$	$\text{Var}(c)$
1	(20, 80)	1/5			
2	(38, 98)	1/2			
3	(30, 90)	1/3			

b. How would this individual rank these three scenarios? Explain and give reasons to support your argument.

c. Show that under each scenario the individual's optimal decision is to invest an equal amount on each of the two assets.

5.7. Consider an agent with a well-behaved utility function who must balance his portfolio between a riskless asset and a risky asset. The first asset, with price p_1 has the certain payoff z_1 while the second asset, with price p_2 pays off \tilde{z}_2 which is a random variable. The agent has an initial wealth Y_0 and he only cares about next period. Assuming that he holds x_1 units of the riskless asset and x_2 units of the risky asset;

a. Write down his expected utility function in terms of Y_0, x_1, x_2, z_1 , and \tilde{z}_2 . Write down his budget constraint as well. Find the equation (the FOC) that the optimal demand for the risky asset has to satisfy.

b. Assume now that his utility function is of the form: $U(Y) = a - be^{-AY}$ with $a, b > 0$. Show that the demand for the risky asset is independent of the initial wealth. Explain intuitively why this is so.

5.8. Consider the savings problem of Section 4.4 :

$$\max_{s \geq 0} EU\{(y_0 - s) + \delta U(s\tilde{x})\}$$

$$\text{Assume } U(c) = Ec - \frac{1}{2}\chi\sigma_c^2$$

Show that if \tilde{x}_A SSD \tilde{x}_B ($w/E\tilde{x}_A = E\tilde{x}_B$), then $s_A > s_B$.

Chapter 7

- 7.7. Show that maximizing the Sharpe ratio, $\frac{E(r_p) - r_f}{\sigma_p}$, yields the same tangency portfolio that was obtained in the text.
- Hint: Formulate the Lagrangian and solve the problem.
- 7.8. Think of a typical investor selecting his preferred portfolio along the Capital Market Line. Imagine:
1. A 1% increase in both the risk free rate and the expected rate of return on the market, so that the CML shifts in a parallel fashion
 2. An increase in the expected rate of return on the market without any change in the risk free rate, so that the CML tilts upward.
- In these two situations, describe how the optimal portfolio of the typical investor is modified.
- 7.9. Questions about the Markowitz model and the CAPM.
- a. Explain why the efficient frontier must be concave.
 - b. Suppose that there are N risky assets in an economy, each being the single claim to a different firm (hence, there are N firms). Then suppose that some firms go bankrupt, i.e. their single stock disappears; how is the efficient frontier altered?
 - c. How is the efficient frontier altered if the borrowing (risk-free) rate is higher than the lending rate? Draw a picture.
 - d. Suppose you believe that the CAPM holds and you notice that an asset (call it asset A) is above the Security Market Line. How can you take advantage of this situation? What will happen to stock A in the long run?
- 7.10. Consider the case without a riskless asset. Take any portfolio p . Show that the covariance vector of individual asset returns with portfolio p is linear in the vector of mean returns if and only if p is a frontier portfolio.
- Hint: To show the "if" part is straightforward. To show the converse begin by assuming that $Vw = ae + b\mathbf{1}$ where V is the variance-covariance matrix of returns, e is the vector of mean returns, and $\mathbf{1}$ is the vector of ones.
- 7.11. Show that the covariance of the return on the minimum variance portfolio and that on any portfolio (not only those on the frontier) is always equal to the variance of the rate of return on the MVP.
- Hint: consider a 2-assets portfolio made of an arbitrary portfolio p and the MVP, with weights a and $1-a$. Show that $a=0$ satisfies the variance minimizing program; the conclusion follows.
- 7.12. Find the frontier portfolio that has an identical variance as that of its zero-covariance portfolio. (That is, determine its weights.)
- 7.13. Let there be two risky securities, a and b . Security a has expected return of 13% and volatility of 30%. Security b has expected return of 26% and volatility of 60%. The two securities are uncorrelated.
- a. Compute the portfolio on the efficient frontier that is tangent to a line from zero, the zero beta portfolio associated with that portfolio, and the minimum-variance portfolio.

- b. Assume a risk-free rate of 5%. Compute the portfolio of risky assets that investors hold. Does this portfolio differ from the tangency portfolio computed under a) ? If yes, why?
- 7.14 a. Given risk-free borrowing and lending, efficient portfolios have no unsystematic risk. True or false?
- b. If the agents in the economy have different utility functions the market portfolio is not efficient. True or false?
- c. The CAPM makes no provision for investor preference for skewness. True or false?

Chapter 8

8.9. Consider an exchange economy with two states. There are two agents with the same utility function $U(c) = \ln(c)$. State 1 has a probability of π . The agents are endowed with the units of the consumption good at each state. Their endowments across themselves and across states are not necessarily equal. Total endowment of this consumption good is e_1 in state 1 and e_2 in state 2. Arrow-Debreu state prices are denoted by q_1 and q_2 .

a. Write down agents' optimization problems and show that

$$\frac{q_1}{q_2} = \frac{\pi}{1-\pi} \left(\frac{y_2}{y_1} \right)$$

Assuming that $q_1 + q_2 = 1$ solve for the state prices. Hint: Recall the simple algebraic fact that $\frac{a}{b} = \frac{c}{d} = \frac{a+c}{b+d}$.

b. Suppose there are two types of asset in the economy. A riskless asset (asset 1) pays off 1 (unit of the consumption good) in each state and has market price of $P_1=1$. The risky asset (asset 2) pays off 0.5 in state 1 and 2 in state 2. Aggregate supplies of the two assets are Q_1 and Q_2 . If the two states are equally likely, show that the price of the risky asset is

$$P_2 = \frac{5Q_1 + 4Q_2}{4Q_1 + 5Q_2}$$

Hint: Note that in this case state-contingent consumption of the agents are assured, in equilibrium, through their holdings of the two assets. To solve the problem you will need to use the results of section a). There is no need to set up another optimization problem.

Chapter 10

10.5. A-D pricing

Consider two 5-year coupon bonds with different coupon rates which are simultaneously traded.

	<u>Price</u>	<u>Coupon</u>	<u>Maturity Value</u>
Bond 1	1300	8 %	1000
Bond 2	1200	6.5 %	1000

For simplicity, assume that interest payments are made once per year. What is the price of a 5-year A-D security when we assume the only state is the date.

10.6. You anticipate receiving the following cash flow which you would like to invest risk free.

t = 0	1	2	3
	\$1m	\$1,25m	

The period denotes one year. Risk free discount bonds of various maturities are actively traded, and the following price data is reported:

t = 0	1	2	3
-950	1000		
-880		1000	
-780			1000

- Compute the term structure implied by these bond prices.
 - How much money can you create, risk free, at $t = 3$ from the above cash flow using the three mentioned instruments?
 - Show the transactions whereby you guarantee (lock in) its creation at $t = 3$.
- 10.7. Consider a world with two states of nature. You have the following term structure of interest rates over two periods:

$$r_1^1 = 11.1111, r_2^1 = 25.0000, r_1^2 = 13.2277, r_2^2 = 21.2678$$

where the subscript denotes the state at the beginning of period 1, and the superscript denotes the period.

For instance $\frac{1}{(1+r_j^2)^2}$ is the price at state j at the beginning of period 1 of a riskless asset paying 1 two periods later. Construct the stationary (same every period) Arrow-Debreu state price matrix.

Chapter 13

13.6. Assume that the following two-factor model describes returns

$$r_i = a_i + b_{i1}F_1 + b_{i2}F_2 + e_i$$

Assume that the following three portfolios are observed.

Portfolio	Expected returns	b_{i1}	b_{i2}
A	12.0	1	0.5
B	13.4	3	0.2
D	12.0	3	-0.5

a. Find the equation of the plane that must describe equilibrium returns.

b. If $\tilde{r}_M - r_f = 4$, find the values for the following variables that would make the expected returns consistent with equilibrium determined by the CAPM.

i) r_f

ii) β_{pi} , the market beta of the pure portfolio associated with factor i

13.7. Based on a single factor APT model, the risk premium on a portfolio with unit sensitivity is 8% ($\lambda_1 = 8\%$). The risk free rate is 4%. You have uncovered three well-diversified portfolios with the following characteristics:

<u>Portfolio</u>	<u>Factor Sensitivity</u>	<u>Expected Return</u>
A	.80	10.4%
B	1.00	10.0%
C	1.20	13.6%

Which of these three portfolios is not in line with the APT?

13.8 A main lesson of the CAPM is that “diversifiable risk is not priced”. Is this important result supported by the various asset pricing theories reviewed in this book? Discuss.

As a provision of supplementary material we describe below how the APT could be used to construct an arbitrage portfolio to profit of security mispricing. The context is that of equity portfolios. The usefulness of such an approach will, of course, depend upon “getting the right set of factors” so that the attendant regressions have high R^2 .

13.9 An APT Exercise in Practice

- Step 1: select the factors; suppose there are J of them.
- Step 2: For a large number of firms N (big enough so that when combined in an approximately equally weighted portfolio of them the unique risks diversify away to approx. zero) undertake the following time series regressions on historical data:

$$\begin{array}{ll}
 \text{Firm 1: IBM} & \tilde{r}_{\text{IBM}} = \hat{\alpha}_{\text{IBM}} + \hat{b}_{\text{IBM},1}\tilde{F}_1 + \dots + \hat{b}_{\text{IBM},J}\tilde{F}_J + \tilde{e}_{\text{IBM}} \\
 \text{Firm 2: BP} & \tilde{r}_{\text{BP}} = \hat{\alpha}_{\text{BP}} + \hat{b}_{\text{BP},1}\tilde{F}_1 + \dots + \hat{b}_{\text{BP},J}\tilde{F}_J + \tilde{e}_{\text{BP}} \\
 \vdots & \vdots \quad \vdots \quad \vdots \quad \vdots \\
 \text{Firm N: GE} & \tilde{r}_{\text{GE}} = \hat{\alpha}_{\text{GE}} + \hat{b}_{\text{GE},1}\tilde{F}_1 + \dots + \hat{b}_{\text{GE},J}\tilde{F}_J + \tilde{e}_{\text{GE}}
 \end{array}$$

The return to each stock are regressed on the same J factors; what differs is the factor sensitivities $\{\hat{b}_{\text{IBM},1}, \dots, \hat{b}_{\text{GE},J}\}$. Remember that:

$$\hat{b}_{\text{BP},J} = \frac{\text{cov}(\tilde{r}_{\text{BP}}, \tilde{F}_J^{\text{HIST}})}{\sigma_{F_J}^2}$$

(want high R^2)

- Step 3: first assemble the following data set:

$$\begin{array}{ll}
 \text{Firm 1: IBM} & \text{AR}_{\text{IBM}}, \hat{b}_{\text{IBM},1}, \hat{b}_{\text{IBM},2}, \dots, \hat{b}_{\text{IBM},J} \\
 \text{Firm 2: BP} & \text{AR}_{\text{BP}}, \hat{b}_{\text{BP},1}, \hat{b}_{\text{BP},2}, \dots, \hat{b}_{\text{BP},J} \\
 \vdots & \vdots \quad \vdots \quad \vdots \quad \vdots \\
 \text{Firm N: GE} & \text{AR}_{\text{GE}}, \hat{b}_{\text{GE},1}, \hat{b}_{\text{GE},2}, \dots, \hat{b}_{\text{GE},J}
 \end{array}$$

The $\text{AR}_{\text{IBM}}, \text{AR}_{\text{BP}}, \dots$ etc. represent the average returns on the N stocks over the historical period chosen for the regression.

Then, regress the average returns on the factor sensitivities (we have N data points corresponding to the N firms)

$$\begin{array}{c}
 \text{(these vary across the N firms)} \\
 \swarrow \quad \searrow \quad \swarrow \\
 A\tilde{r}_i = r_f + \hat{\lambda}_1 \tilde{b}_{i1} + \hat{\lambda}_2 \tilde{b}_{i2} + \dots + \hat{\lambda}_J \tilde{b}_{iJ} \\
 \text{we obtain estimates } \{\hat{\lambda}_1, \dots, \hat{\lambda}_J\}
 \end{array}$$

In the regression sense this determines the “best” linear relationship among the factor sensitivities and the past average returns for this sample of N stocks.

This is a “cross sectional” regression. (Want a high R^2)

- Step 4: Compare, for the N assets, their actually observed returns with what should have been observed given their factor sensitivities; compute α_j 's:

$$\alpha_{IBM} = \underbrace{AR_{IBM}}_{\text{IBM's actually observed historical return}} - \underbrace{\left[r_f + \hat{\lambda}_1 \hat{b}_{IBM,1} + \dots + \hat{\lambda}_J \hat{b}_{IBM,J} \right]}_{\text{predicted return given its factors intensities } \hat{b}_{IBM,1}, \dots, \hat{b}_{IBM,J} \text{ according to the regression in step 3}}$$

$$\vdots$$

$$\alpha_{GE} = AR_{GE} - \left[r_f + \hat{\lambda}_1 \hat{b}_{GE,1} + \dots + \hat{\lambda}_J \hat{b}_{GE,J} \right]$$

Note that

$\alpha_j > 0$ implies the average returns exceeded what would be justified by the factor intensities => undervalued;

$\alpha_j < 0$ implies the average returns fell short of what would be justified by the factor intensities => overvalued.

e. Step 5: Form an arbitrage portfolio of the N stocks:

if $\alpha_j > 0$ - assume a long position

if $\alpha_j < 0$ - assume a short position

since N is large $e_p \equiv 0$, so ignore “unique” risks.

Remarks:

1.: In step 4 we could substitute independent (otherwise obtained) estimates of AR_i 's, and not use the historical averages.

2.: Notice that nowhere do we have to forecast future values of the factors.

3.: In forming the arbitrage portfolio we are implicitly assuming that the over and under pricing we believe exists will be eliminated in the future – to our advantage!

Chapter 15

- 15.6 Consider two agents in the context of a pure exchange economy in which there are two dates ($t = 0, 1$) and two states at $t = 1$. The endowments of the two agents are different ($e_1 \neq e_2$). Both agents have the same utility function :

$$U(c_0, c_1(\theta)) = \ln c_0 + E \ln c_1(\theta),$$

but they differ in their beliefs. In particular, agent 1 assigns probability $\frac{3}{4}$ to state 1, while agent 2 assigns state 1 a probability $\frac{1}{4}$. The agents trade Arrow-Debreu claims and the supply of each claim is 1. Neither agent receives any endowment at $t=1$.

- Derive the equilibrium state claim prices. How are they related to the relative endowments of the agents ? How are the relative demands of each security related to the agents' subjective beliefs ?
- Suppose rather than trading state claims, each agent is given a_i units of a riskless security paying one unit in each future state. Their $t=0$ endowments are otherwise unaffected. Will there be trade ? Can you think of circumstances where no trade will occur ?
- Now suppose that a risky asset is also introduced into this economy. What will be the effects ?
- Rather than introducing a risky asset, suppose an entrepreneur invents a technology that is able to convert x units of the riskless asset into x units each of $(1,0)$ and $(0,1)$. How is x and the value of these newly created securities related ? Could the entrepreneur extract a payment for the technology ? What considerations would influence the magnitude of this payment ?

Intermediate Financial Theory
Danthine and Donaldson

Solutions to Additional Exercises

Chapter 1

- 1.6. Consider a two agent –two good economy. Assume well-behaved utility functions (in particular, indifference curves don't exhibit flat spots). At a competitive equilibrium, both agents maximize their utility given their budget constraints. This leads each of them to select a bundle of goods corresponding to a point of tangency between one of his or her indifference curves and the price line. Tangency signifies that the slope of the IC and the slope of the budget line (the price ratio) are the same. But both agents face the same market prices. The slope of their indifference curves are thus identical at their respective optimal point.

Now consider the second requirement of a competitive equilibrium: that market clear. This means that the respective optimal choices of each of the two agents correspond to the same point of the Edgeworth-Bowley box.

Putting the two elements of this discussion together, we have that a competitive equilibrium is a point in the box corresponding to a feasible allocation where both agents' indifference curves are tangent to the same price line, have the same slope, and, consequently, are tangent to one another. Since the contract curve is the locus of all such points in the box at which the two agents' indifference curves are tangent, the competitive equilibrium is on the contract curve.

Of course, we could have obtained this result simply by invoking the First Welfare Theorem.

- 1.7. Indifference curves of agent 2 are non-convex.
Point A is a PO : the indifference curves of the two agents are tangent. This PO cannot be obtained as a Competitive Equilibrium, however. Let a price line tangent to I_1 at point A. It is also tangent to I_2 , but “in the wrong direction”: it corresponds to a local minimum for agent 2 who, at those prices, can reach higher utility levels. The difficulty is generic when indifference curves have such a shape. The geometry is inescapable, underlining the importance of the assumption that preferences should be convex.

Chapter 4

4.9 Certainty equivalent.

The problem to be solved is: find Y such that

$$\frac{1}{2}U(Y+1000) + \frac{1}{2}U(Y-1000) \equiv U(Y-500)$$

$$\frac{1}{(Y+1000)} + \frac{1}{(Y-1000)} = \frac{2}{(Y-500)}$$

$$\frac{(Y+1000) + (Y-1000)}{(Y+1000)(Y-1000)} = \frac{2}{(Y-500)}$$

$$\frac{2Y}{(Y^2 - 1000^2)} = \frac{2}{(Y-500)}$$

$$Y^2 - 1000^2 = Y^2 - 500Y$$

$$Y = 2000$$

The logarithmic utility function is solved in the same way; the answer is $Y = 1250$.

4.10. Risk premium.

The problem to be solved is: find P such that

$$\frac{1}{2}(\ln(Y+1000) + \ln(Y-1000)) \equiv \ln(Y-P)$$

$$Y-P = \exp\left(\frac{1}{2}(\ln(Y+1000) + \ln(Y-1000))\right)$$

$$P = Y - \exp(.)$$

where P is the insurance premium.

$$P(Y=10000) = 50.13$$

$$P(Y=100000) = 0.50$$

The utility function is DARA, so the outcome (smaller premium associated with higher wealth) was expected.

4.11. Case 1 Case 2 Case 3

$$\sigma_a > \sigma_b \quad \sigma_a = \sigma_b \quad \sigma_a < \sigma_b$$

$$E_a = E_b \quad E_a > E_b \quad E_a < E_b$$

Case 1: cannot conclude with FSD, but B SSD A

Case 2: A FSD B, A SSD B

Case 3: cannot conclude (general case)

4.12. a. $U(Y - CE) = EU(Y - L(\theta))$

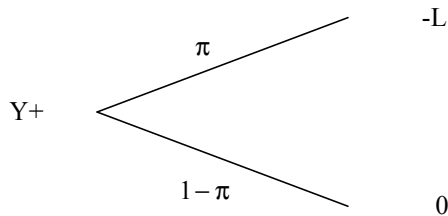
$$\frac{(10,000 + CE)^{-.2}}{-.2} = .10 \frac{(10,000 - 1,000)^{-.2}}{-.2} + .20 \frac{(10,000 - 2,000)^{-.2}}{-.2}$$

$$\begin{aligned}
& + .35 \frac{(10,000 - 3,000)^{-2}}{-.2} + .20 \frac{(10,000 - 5,000)^{-2}}{-.2} + .15 \frac{(10,000 - 6,000)^{-2}}{-.2} \\
(10,000 + CE)^{-2} &= .173846 \\
(10,000 + CE)^2 &= \frac{1}{.173846} = 5.752 \\
CE &= -3702.2 \\
E\tilde{L} &= -\{1(1000) + .2(2000) + .35(3000) + .2(5000) + .15(6000)\} \\
&= -3450 \\
CE(\tilde{z}, y) &= E(\tilde{z}) - \Pi(y, \tilde{z}) \\
\Pi(y, \tilde{z}) &= 252.2
\end{aligned}$$

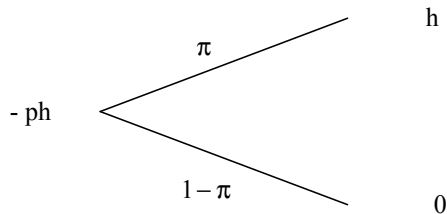
If the agent were risk neutral the CE = - 3450

b. If $U'(y) > 0, U''(y) > 0$, the agent loves risk. The premium would be negative here.

4.13. Current Wealth :



Insurance Policy :



Certainly $p \leq 1$

a. Agent solves

$$\max_h \pi \ln(y - ph - L + h) + (1 - \pi) \ln(y - ph)$$

The F.O.C. is

$$\frac{\pi(1 - p)}{y - L + h(1 - p)} = \frac{p(1 - \pi)}{y - ph}, \text{ which solves for}$$

$$h = Y \left(\frac{\pi}{p} \right) - \left(\frac{1 - \pi}{1 - p} \right) (Y - L)$$

Note : if $p = 0$, $h = \infty$; if $\pi = 1$, $ph = Y$.

b. expected gain is $ph - \pi L$

$$\begin{aligned}
c. \quad ph &= p \left(Y \left(\frac{\pi}{p} \right) - \left(\frac{1 - \pi}{1 - p} \right) (Y - L) \right) = \pi L \\
&\Rightarrow p = \pi
\end{aligned}$$

$$d. \quad h = Y \left(\frac{\pi}{p} \right) - \left(\frac{1 - \pi}{1 - p} \right) (Y - L)$$

$$h = Y \left(\frac{\pi}{\pi} \right) - \left(\frac{1-\pi}{1-\pi} \right) (Y - L) = L.$$

The agent will perfectly insure. None ; this is true for all risk averse individuals.

4.14.

	$\tilde{x}, \pi(\tilde{x})$	$\tilde{z}, \pi(\tilde{z})$
a.	$E\tilde{x} = -10(.1) + 5(.4) + 10(.3) + 12(.2) = -1 + 2 + 3 + 2.4 = 6.4$	
	$E\tilde{z} = .2(2) + 3(.5) + 4(.2) + 30(.1) = .4 + 1.5 + .8 + 3 = 5.7$	
	$\sigma_{\tilde{x}}^2 = .1(-10 - 6.4)^2 + .4(5 - 6.4)^2 + .3(10 - 6.4)^2 + .2(12 - 6.4)^2$	
	$= 26.9 + .78 + 3.9 + 6.27$	
	$= 37.85$	
	$\sigma_{\tilde{x}} = 6.15$	
	$\sigma_{\tilde{z}}^2 = .2(2 - 5.7)^2 + .5(3 - 5.7)^2 + .2(4 - 5.7)^2 + .1(30 - 5.7)^2$	
	$= 2.74 + 3.65 + .58 + 59.04$	
	$= 66.01$	
	$\sigma_{\tilde{z}} = 8.12$	

There is mean variance dominance in favor of \tilde{x} :

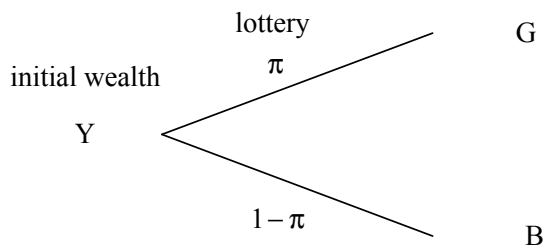
$E\tilde{x} > E\tilde{z}$ and $\sigma_{\tilde{x}} < \sigma_{\tilde{z}}$. The latter is due to the large outlying payment of 30.

b. 2nd order stochastic dominance :

r	$F_{\tilde{x}}(r)$	$\int_0^r F_{\tilde{x}}(t)dt$	$F_{\tilde{z}}(r)$	$\int_0^r F_{\tilde{z}}(t)dt$	$\int_0^r [F_{\tilde{x}}(t) - F_{\tilde{z}}(t)]dt$
-10	.1	.1	0	0	.1
-9	.1	.2	0	0	.2
-8	.1	.3	0	0	.3
-7	.1	.4	0	0	.4
-6	.1	.5	0	0	.5
-5	.1	.6	0	0	.6
-4	.1	.7	0	0	.7
-3	.1	.8	0	0	.8
-2	.1	.9	0	0	.9
-1	.1	1.0	0	0	1.0
0	.1	1.1	0	0	1.1
1	.1	1.2	0	0	1.2
2	.1	1.3	.2	.2	1.1
3	.1	1.4	.7	.9	.5
4	.1	1.5	.9	1.8	-.3

Since the final column is not of uniform sign, we cannot make any claim about relative 2nd order SD.

4.15.



a. If he already owns the lottery, P_s must satisfy

$$U(Y + P_s) = \pi U(Y + G) + (1 - \pi)U(Y + B)$$

$$\text{or } P_s = U^{-1}(\pi U(Y + G) + (1 - \pi)U(Y + B)) - Y.$$

b. If he does not own the lottery, the maximum he would be willing to pay, P_b , must satisfy :

$$U(Y) = \pi U(Y - P_b + G) + (1 - \pi)U(Y - P_b + B)$$

c. Assume now that $\pi = \frac{1}{2}$, $G = 26$, $B = 6$, $Y = 10$. Find P_s , P_b .

P_b satisfies :

$$U(10) = \frac{1}{2}U(10 - P_b + 26) + \frac{1}{2}U(10 - P_b + 6)$$

$$10^{\frac{1}{2}} = \frac{1}{2}(36 - P_b)^{\frac{1}{2}} + \frac{1}{2}(16 - P_b)^{\frac{1}{2}}$$

$$6.32 = (36 - P_b)^{\frac{1}{2}} + (16 - P_b)^{\frac{1}{2}}$$

$$P_b \approx 13.5$$

P_s satisfies :

$$(10 + P_s)^{\frac{1}{2}} = \frac{1}{2}(10 + 26)^{\frac{1}{2}} + \frac{1}{2}(10 + 6)^{\frac{1}{2}}$$

$$= \frac{1}{2}(6) + \frac{1}{2}(4) = 5$$

$$10 + P_s = 25$$

$$P_s = 15$$

Clearly, $P_b < P_s$. If the agent already owned the asset his minimum wealth is $10 + 6 = 16$. If he is considering buying, its wealth is 10. In the former case, he is less risk averse and the lottery is worth more.

If the agent is risk neutral, $P_s = P_b = \pi G + (1 - \pi)B$. To check it out, assume $U(x) = x$:

$$P_s : U(Y + P_s) = \pi U(Y + G) + (1 - \pi)U(Y + B)$$

$$Y + P_s = \pi(Y + G) + (1 - \pi)(Y + B) = Y + \pi G + (1 - \pi)B$$

$$P_s = \pi G + (1 - \pi)B$$

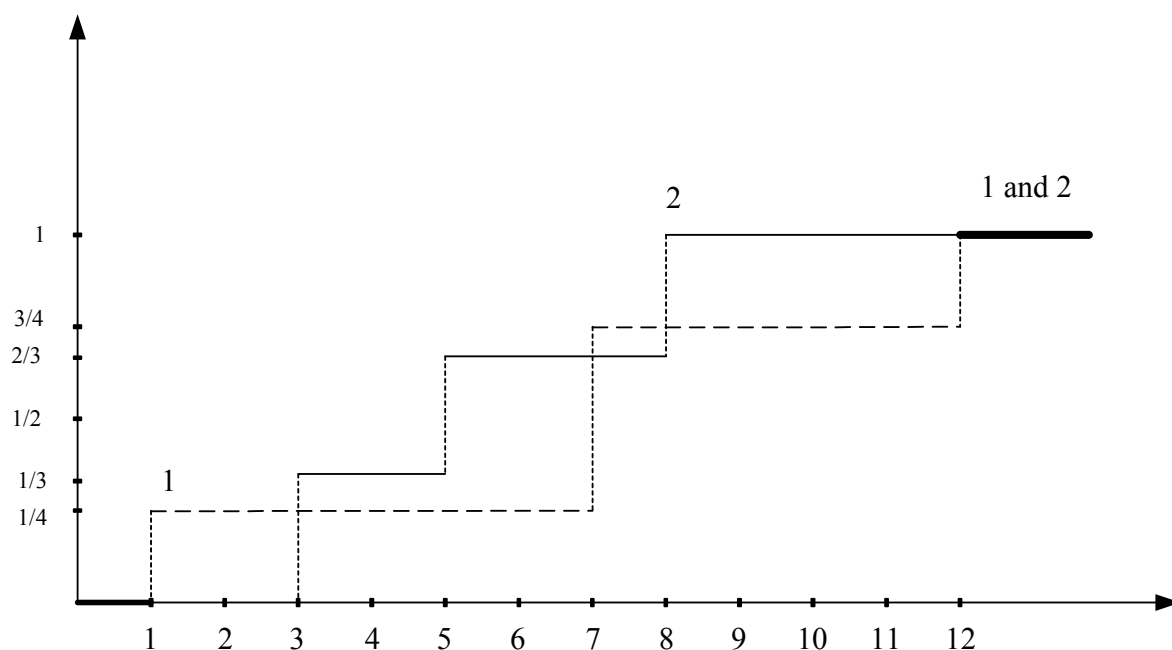
$$P_b : U(Y) = \pi U(Y - P_b + G) + (1 - \pi)U(Y - P_b + B)$$

$$Y = \pi(Y - P_b + G) + (1 - \pi)(Y - P_b + B)$$

$$P_b = \pi G + (1 - \pi)B$$

4.16. Mean-variance: $Ex_1 = 6.75$, $(\sigma_1)^2 = 15.22$; $Ex_2 = 5.37$, $(\sigma_2)^2 = 4.25$; no dominance.

FSD: No dominance as the following graph shows:



SSD:

x	$\int_0^x f_1(t)dt$	$\int_0^x F_1(t)dt$	$\int_0^x f_2(t)dt$	$\int_0^x F_2(t)dt$	$\int_0^x [F_1(t) - F_2(t)]dt$
0	0	0	0	0	0
1	.25	.25	0	0	.25
2	.25	.50	0	0	.50
3	.25	.75	0	0	.75
4	.25	1	.33	.33	.67
5	.25	1.25	.33	.66	.67
6	.25	1.50	.66	1.32	.18
7	.50	2	.66	1.98	.02
8	.50	2.50	1	2.98	-.48

There is no SSD as the sign of $\int_0^x [F_1(t) - F_2(t)]dt$ is not monotone.

Using Expected utility. Generally speaking, one would expect the more risk averse individuals to prefer investment 2 while less risk averse agents would tend to favor investment 1.

Chapter 5

5.6. a.

Scenario	(z_1, z_2)	π	(c_1, c_2)	$E(c)$	$Var(c)$
1	(20, 80)	1/5	(35, 65)	59	144
2	(38, 98)	1/2	(44, 74)	59	225
3	(30, 90)	1/3	(40, 70)	60	200

b.

• Mean-Variance analysis:

1 is preferred to 2 (same mean, but lower variance)

3 is preferred to 2 (higher mean and lower variance)

1 and 3 can not be ranked with standard mean-variance analysis

• Stochastic Dominance:

No investment opportunity FSD one of the other investments.

Investment 1 SSD Scenario 2 (mean preserving spread).

• Expected Utility (with assumed U)

$3 > 1 > 2$

c.

Scenario 1

$$EU(a) = \frac{1}{5}[-\exp(-A[50 - .6a50])] + \frac{4}{5}[-\exp(-A[50 + .6a50])]$$

$$\frac{\partial EU(a)}{\partial a} = \frac{1}{5}(.6A50)[-\exp(-A[50 - .6a50])] - \frac{4}{5}(.6A50)[-\exp(-A[50 + .6a50])] = 0$$

$$a = -\frac{\ln\left(\frac{1}{4}\right)}{(1.2A50)} = .5$$

The scenarios 2 and 3 can be solved along the same lines.

5.7. a. $\max_{x_1, x_2} EU(x_1 z_1 + x_2 \tilde{z}_2)$

$$\text{s.t. } p_1 x_1 + p_2 x_2 \leq Y_0$$

Since we assume (maintained assumption) $U'(\cdot) > 0$, $p_1 x_1 + p_2 x_2 = Y_0$, and $x_1 = \frac{Y_0 - p_2 x_2}{p_1}$

The problem may thus be written :

$$\max_{x_2} EU\left[\left(\frac{Y_0 - p_2 x_2}{p_1}\right)z_1 + x_2 \tilde{z}_2\right]$$

The necessary and sufficient F.O.C. (under the customary assumption) is :

$$EU'\left[\left(\frac{Y_0 - p_2 x_2}{p_1}\right)z_1 + x_2 \tilde{z}_2\right]\left(\tilde{z}_2 - \frac{p_2}{p_1} z_1\right) = 0$$

b. Suppose $U(y) = a - be^{-Ay}$; the above equation becomes :

$$AbE\left\{\exp\left[\left(\frac{Y_0 - p_2 x_2}{p_1}\right)z_1 + x_2 \tilde{z}_2\right]\left(\tilde{z}_2 - \frac{p_2}{p_1} z_1\right)\right\} = 0$$

equivalently,

$$\text{AbE} \left\{ \exp \left[\frac{Y_0 z_1}{p_1} \right] \exp \left[\frac{-p_2 x_2}{p_1} z_1 + x_2 \tilde{z}_2 \right] \left(\tilde{z}_2 - \frac{p_2}{p_1} z_1 \right) \right\} = 0$$

The first term which contains Y_0 can be eliminated from the equation. The intuition for this result is in the fact that the stated utility is CARA, that is, the rate of absolute risk aversion is constant and independent of the initial wealth.

5.8. The problem with linear mean variance utility is

$$\max(y_0 - s) + \delta \left[s E\tilde{x} - \frac{1}{2} \chi s^2 \sigma_x^2 \right]$$

$$\text{FOC} \quad -1 + \delta E\tilde{x} - s \chi \sigma_x^2 = 0$$

$$\text{or} \quad s = \frac{\delta E(\tilde{x}) - 1}{\chi \sigma_x^2}$$

Clearly s is inversely related with σ_x^2 . For a given $E\tilde{x}$ (x_B is a mean-preserving spread of x_A),

$$s_A > s_B$$

Chapter 7

- 7.7. A ray in R^2 is defined by $y - y_1 = n(x - x_1)$. Rewrite this in the following way $y = n(x - x_1) + y_1$ and apply it to the problem:

$$E(r_p) = \frac{E(r_p) - r_f}{\sigma_p} (\sigma_p - 0) + r_f.$$

This can be maximized with respect to the Sharpe ratio. Of course, we get $\sigma_p = 0$; i.e. the slope is infinite.

Now we constrain x to be $x = \sigma_p = \sqrt{\frac{C}{D} \left(E(r_p) - \frac{A}{C} \right)^2 + \frac{1}{C}}$. Inserting this back leads to

$E(r_p) = \theta \sqrt{\frac{C}{D} \left(E(r_p) - \frac{A}{C} \right)^2 + \frac{1}{C}} + r_f$ where θ is the Sharpe ratio. From this it is easy to solve for σ_p^2 and the Sharpe ratio. (A, B, C, and D are the notorious letters defined in Chapter 6.)

- 7.8. 1) Not possible to say without further knowledge of preferences. The reason is that with both risk-free and risky returns higher, there is what is called a 'wealth effect': with given amount of initial wealth to invest, the end-of-period is unambiguously expected to be higher: the investor is 'richer'. But we have not made any assumption as to whether at higher wealth level he/she will be more or less risk averse. The CAPM does not require to specify IARA or CARA or DARA utility functions (although we know that we could build the model on a quadratic (IARA) utility function, this is not the only route.)
2) Here it is even more complicated; the efficient frontier is higher: there is a wealth effect, but it is also steeper: there is also a substitution effect. Everything else equal, the risky portfolio is more attractive. It is more likely that an investor will select a riskier optimal portfolio in this situation, but one cannot rule out that the wealth effect dominates and at the higher expected end-of-period wealth the investor decides to invest more conservatively.
- 7.9. Questions about the Markowitz model and the CAPM.
- If it were not, one could build a portfolio composed of two efficient portfolios that would not be itself efficient. Yet, the new portfolio's expected return would be higher than the frontier portfolio with the same standard deviation, in violation of the efficiency property of frontier portfolios.
 - With a lower number of risky assets, one expects that the new frontier will be contained inside the previous one as diversification opportunities are reduced.
 - The efficient frontier is made of three parts, including a portion of the frontier. Note that borrowers and lenders do not take positions in the same "market portfolio".
 - Asset A is a good buy: it pays on average a return that exceeds the average return justified by its beta. If the past is a good indication of the pattern of future returns, buying asset A offers the promise of an extra return compared to what would be fair according to the CAPM. What could expect that in the longer run many investors will try to exploit such an opportunity and that, as a consequence, the price of asset A will increase with the expected return decreasing back to the SML level.
- 7.10. "If" part has been shown in Chapter 6.

"Only if" : start with $Vw = ae + b\mathbf{1}$; premultiply by V^{-1}

$w_p = aV^{-1}e + bV^{-1}1 = aA\left(\frac{V^{-1}e}{A}\right) + bC\left(\frac{V^{-1}1}{C}\right)$ where $\frac{V^{-1}e}{A}$, and $\frac{V^{-1}1}{C}$ are frontier portfolios with means $\frac{B}{A}$, and $\frac{A}{C}$ respectively. Since $aA+bC=1$ (Why?) the result follows.

- 7.11. We build a portfolio with P and the MVP, with minimum variance. Then, the weights a and $(1-a)$ must satisfy the condition

$$\min_a \left\{ a^2 \times \sigma_p^2 + 2 \times a \times (1-a) \times \text{cov}(r_p, r_{MVP}) + (1-a)^2 \times \sigma_{MVP}^2 \right\}.$$

The FOC is

$$2 \times a \times \sigma_p^2 + 2 \times (1-a) \times \text{cov}(r_p, r_{MVP}) - 2 \times (1-a) \times \sigma_{MVP}^2 = 0.$$

Since MVP is the minimum variance portfolio, $a=0$ must satisfy the condition, which simplifies to $\text{cov}(r_p, r_{MVP}) = \sigma_{MVP}^2$.

- 7.12. For any portfolio on the frontier we have

$$\sigma^2(\tilde{r}_p) = \frac{C}{D} \left(E(\tilde{r}_p) - \frac{A}{C} \right)^2 + \frac{1}{C}.$$

where A, B, C , and D are our notorious numbers. Additionally, we know that

$E(\tilde{r}_{zcp}) = \frac{A}{C} - \frac{D/C^2}{E(\tilde{r}_p) - A/C}$. Since the zero covariance portfolio is also a frontier portfolio we have

$$\sigma^2(\tilde{r}_{zcp}) = \frac{C}{D} \left(\frac{D/C^2}{E(\tilde{r}_p) - A/C} \right)^2 + \frac{1}{C}.$$

Now, we need to have $\sigma^2(\tilde{r}_p) = \sigma^2(\tilde{r}_{zcp})$. This leads to

$$\frac{C}{D} \left(E(\tilde{r}_p) - \frac{A}{C} \right)^2 + \frac{1}{C} = \frac{C}{D} \left(\frac{D/C^2}{E(\tilde{r}_p) - A/C} \right)^2 + \frac{1}{C}$$

$$E(\tilde{r}_p) - \frac{A}{C} = \frac{D/C^2}{E(\tilde{r}_p) - A/C}.$$

$$E(\tilde{r}_p) - \frac{A}{C} = \frac{\sqrt{D}}{C}$$

$$E(\tilde{r}_p) = \frac{A}{C} + \frac{\sqrt{D}}{C}$$

Given $E(\tilde{r}_p)$, we can use (6.15) from chapter 6 to find the portfolio weights.

- 7.13. a. As shown in Chapter 5, we find the slope of the mean-variance frontier and utilize it in the equation of the line passing through the origin. If we call the portfolio we are seeking "p", then it follows that

$E(r_p) = \frac{D + A^2 C}{C^2 A}$ where A, B, C , and D are our notorious numbers. The zero-covariance portfolio of p is such that $E(r_{zcp}) = 0$.

$$E(r_{zcp}) = \frac{A}{C} - \frac{D/C^2}{E(r_p) - A/C} = 0$$

$$\Leftrightarrow E(r_p) = \frac{D}{CA} + \frac{A}{C} = 0.1733$$

b. We need to compare the weights of the portfolio p with $w_T = \frac{V^{-1}(e - r_f \mathbf{1})}{A - r_f C}$. The two portfolios should differ because we are comparing the tangency points of two different lines on the same mean-variance frontier. Note also that the intercepts are different:

$$E(r_{zcp}) = 0.05$$

$$\Leftrightarrow E(r_p) = \frac{D/C^2}{[A/C - 0.05]} + \frac{A}{C} = 0.1815$$

- 7.14. a. True
 b. False: the CAPM holds even with investors have different rates of risk aversion. It however requires that they are all mean-variance maximizers.
 c. True. Only the mean and variance of a portfolio matters. They have no preference for the third moment of the return distribution. Portfolio including derivative instruments may exhibit highly skewed return distribution. For non-quadratic utility investors the prescriptions of the CAPM should, in that context, be severely questioned.

Chapter 8

8.9. a.

The optimization problem of the first agent is

$$\text{Max} EU(c) \quad \text{s.t.} \quad q_1 c_{11} + q_2 c_{12} = q_1 e_{11} + q_2 e_{12}.$$

The FOC's are,

$$\pi \frac{1}{c_{11}} = \lambda q_1$$

$$(1-\pi) \frac{1}{c_{12}} = \lambda q_2$$

$$q_1 c_{11} + q_2 c_{12} = q_1 e_{11} + q_2 e_{12}$$

where λ is the Lagrange multiplier of the problem. Clearly, if we define

$$c_{11} = y_1, c_{12} = y_2 \text{ we have}$$

$$\frac{q_1}{q_2} = \frac{\pi}{(1-\pi)} \frac{y_2}{y_1}.$$

A-D can be derived as follows

$$\frac{q_1}{q_2} = \frac{\pi}{(1-\pi)} \frac{c_{12}}{c_{11}} = \frac{\pi}{(1-\pi)} \frac{c_{22}}{c_{21}} = \frac{\pi}{(1-\pi)} \frac{c_{12}}{c_{11}} + \frac{\pi}{(1-\pi)} \frac{c_{22}}{c_{21}} = \frac{\pi}{(1-\pi)} \frac{e_2}{e_1}.$$

Using $1 = q_1 + q_2$ and after some manipulation we get

$$q_1 = \frac{\pi e_2}{(1-\pi)e_1 + \pi e_2}$$

$$q_2 = \frac{(1-\pi)e_1}{(1-\pi)e_1 + \pi e_2}.$$

b.

If $\pi = (1-\pi)$ A-D prices are

$$q_1 = \frac{e_2}{e_1 + e_2}$$

$$q_2 = \frac{e_1}{e_1 + e_2}$$

The price of the risky asset is

$$P_2 = \frac{1}{2} q_1 + 2 q_2.$$

Now we insert A-D prices and since endowments are

$$e_1 = Q_1 + \frac{1}{2} Q_2$$

$$e_2 = Q_1 + 2 Q_2$$

the pricing formula

$$P_2 = \frac{5Q_1 + 4Q_2}{4Q_1 + 5Q_2}$$

follows.

Chapter 10

- 10.5. The dividends (computed on the face value, 1000) are $d_1 = 80$, $d_2 = 65$. The ratio is d_1/d_2 ; buying 1 unit of bond 1 and selling d_1/d_2 units of bond 2, we can build a 5-yr zero-coupon bond with following payoffs:

	Price	Maturity Value
Bond 1	-1300	+1080
Bond 2	$+1200d_1/d_2$	$1065d_1/d_2$
<hr/>		
	176.92	-230.77

The price of the 5-yr A-D security is then $176.92/230.77 = 23/30$.

10.6. a. r_1 : $950 = \frac{1000}{(1+r_1)}; (1+r_1) = \frac{1000}{950}; r_1 = .05263$

r_2 : $880 = \frac{1000}{(1+r_2)^2}; (1+r_2) = \left(\frac{1000}{880}\right)^{1/2}; r_2 = .0660$

r_3 : $780 = \frac{1000}{(1+r_3)^3}; (1+r_3) = \left(\frac{1000}{780}\right)^{1/3}; r_3 = .0863$

b. We need the forward rates ${}_1f_2$ and ${}_2f_1$.

(i) $(1+r_1)(1+{}_1f_2)^2 = (1+r_3)^3$

$$(1+{}_1f_2) = \left[\frac{(1.0863)^3}{(1.05263)} \right]^{1/2} = 1.1035$$

$$(1+r_2)^2(1+{}_2f_1) = (1+r_3)^3$$

$$(1+{}_2f_1) = \frac{(1.0863)^3}{(1.0660)^2} = 1.1281$$

So ${}_1f_2 = .1035$, and ${}_2f_1 = .1281$.

The $CF_{T=3} = (1.25M)(1+{}_2f_1) + 1M(1+{}_1f_2)^2$

$$= (1.25M)(1.1281) + 1M(1.1035)^2$$

$$= 1.410 M + 1.2177 M$$

$$= 2.6277 M.$$

c. To lock in the ${}_2f_1$ applied to the 1.25M, consider the following transactions :

We want to replicate

	t=0	1	2	3
			-1.25M	1.410M
Short : 1250 2 yr bond				
Long : 1410 3 yr bond				

Consider the corresponding cash flows :

	t=0	1	2	3
Short : (1250)(880)= +1,100,000			-1,250,000	
Long : -(1410)(780)= - 1,100,000				1,410,000
Total		0	-1,250,000	1,410,000

To lock in the ${}_1f_2$ (compounded for two periods) applied to the 1M ; consider the following transactions :

t=0	1	2	3
	-1M		1.2177M

Short : 1000 1 yr bonds.

Long : 1217.7 3 yr bonds.

Consider the corresponding cash flows :

t=0	1	2	3
Short : (1000)(950)=	950,000	-1,000,000	
Long : -(1217.7)(780)=	- 950,000		+1,217,700
Total	0	-1,000,000	+1,217,700

The portfolio that will allow us to invest the indicated cash flows at the implied forward rates is :

Short : 1000, 1 yr bond

Short : 1000, 2 yr bond

Long : 1410 + 1217.7 = 2627.7, 3 yr bond

- 10.7. If today's state is state 1, to get \$1.- for sure tomorrow using Arrow-Debreu prices, I need to pay $q_{11} + q_{12}$; thus

$$q_{11} + q_{12} = \frac{1}{1+r_1^1} = .9 \quad (i)$$

Similarly, if today's state is state 2:

$$q_{21} + q_{22} = \frac{1}{1+r_2^1} = .8 \quad (ii)$$

Given the matrix of Arrow-Debreu prices

$$q = \begin{pmatrix} q_{11} & q_{12} \\ q_{21} & q_{22} \end{pmatrix}; \quad q^2 = \begin{pmatrix} q_{11} & q_{12} \\ q_{21} & q_{22} \end{pmatrix} \begin{pmatrix} q_{11} & q_{12} \\ q_{21} & q_{22} \end{pmatrix}.$$

To get \$1.- for sure two periods from today, I need to pay

$$q_{11}q_{11} + q_{12}q_{21} + q_{11}q_{12} + q_{12}q_{22} = \frac{1}{(1+r_1^2)^2} = .78 \quad (iii)$$

If state 2 today

$$q_{21}q_{11} + q_{22}q_{21} + q_{21}q_{12} + q_{22}q_{22} = \frac{1}{(1+r_2^2)^2} = .68 \quad (iv)$$

The 4 equations (i) to (iv) can be solved for the 4 unknown Arrow-Debreu prices as per

$$q_{11} = 0.6$$

$$q_{22} = 0.3$$

$$q_{21} = 0.4$$

$$q_{12} = 0.4$$

Chapter 13

13.6.

a. From the main APT equation and the problem data, one obtains the following system :

$$12.0 = Er_A = \lambda_0 + \lambda_1 1 + \lambda_2 0.5 \quad (i)$$

$$13.4 = Er_B = \lambda_0 + \lambda_1 3 + \lambda_2 0.2 \quad (ii)$$

$$12.0 = Er_C = \lambda_0 + \lambda_1 3 - \lambda_2 0.5 \quad (iii)$$

This system can easily be solved for

$$\lambda_0 = 10 ; \lambda_1 = 1 ; \lambda_2 = 2$$

Thus, the APT tells us that

$$Er_i = 10 + 1 b_{i1} + 2 b_{i2}$$

b. (i) If there is a risk free asset one must have

$$\lambda_0 = r_f = 10$$

(ii) Let P_i be the pure factor portfolio associated with factor i . One has $\lambda_i = \bar{r}_{P_i} - r_f$. Furthermore if the CAPM holds one should have

$$\lambda_i = \bar{r}_{P_i} - r_f = \beta_{P_i} (\bar{r}_M - r_f).$$

Thus

$$\lambda_1 = 1 = \beta_{P1} 4 \rightarrow \beta_{P1} = \frac{1}{4}, \text{ and}$$

$$\lambda_2 = 2 = \beta_{P2} 4 \rightarrow \beta_{P2} = \frac{1}{2}.$$

13.7.

<u>Expected APT Return</u>	<u>Expected Returns</u>
$r_A = 4\% + .8(8\%) = 10.4\%$	10.4%
$r_B = 4\% + 1(8\%) = 12.0\%$	12.0%
$r_C = 4\% + 1.2(8\%) = 13.6\%$	13.6%

The Expected return of B is less than what is consistent with the APT.

This provides an arbitrage opportunity. Consider a combination of A and C that gives the same factor sensitivity as B.

$$w_A (.8) + \underbrace{w_C}_{1-w_A} (1.2) = 1.00$$

$$\Rightarrow w_A = w_C = 1/2$$

What is the expected return on this portfolio:

$$\begin{aligned} E\left(r_{1/2A, 1/2C}\right) &= \frac{1}{2} E(\tilde{r}_A) + \frac{1}{2} E(\tilde{r}_C) \\ &= \frac{1}{2} (10.4\%) + \frac{1}{2} (13.6\%) \\ &= 12\% \end{aligned}$$

This clearly provides an arbitrage opportunity: short portfolio B, buy a portfolio of $\frac{1}{2}$ A and $\frac{1}{2}$ C.

13.8.

That diversifiable risk is not priced has long been considered as the main lesson of the CAPM. While defining systematic risk differently (in terms of the consumption portfolio rather than the market portfolio), the CCAPM leads to the same conclusion. So does the APT with possibly yet another definition for systematic risk, at least in the case where the market portfolio risk does not encompass the various risk factors identified by the APT. The Value additivity theorem seen in Chapter 7 proves that Arrow-Debreu pricing also leads to the same implication. The equivalence between risk neutral prices and Arrow-Debreu prices in complete markets guarantees that the same conclusion follows from the martingale pricing theory. When markets are incomplete, some risks that we would understand as being diversifiable may no longer be so. In those situations a reward from holding these risks may be forthcoming.

Chapter 15

15.6. a. Agent problems :

$$\text{Agent 1 : } \max_{Q_1^1, Q_2^1} \ln(e_1 - p_1 Q_1^1 - p_2 Q_2^1) + \frac{3}{4} \ln Q_1^1 + \frac{1}{4} \ln Q_2^1$$

$$\text{Agent 2 : } \max_{Q_1^2, Q_2^2} \ln(e_2 - p_1 Q_1^2 - p_2 Q_2^2) + \frac{3}{4} \ln Q_1^2 + \frac{1}{4} \ln Q_2^2$$

FOCs :

Agent 1 :

$$(i) \quad Q_1^1 : \frac{p_1}{e_1 - p_1 Q_1^1 - p_2 Q_2^1} = \frac{3}{4Q_1^1}$$

$$(ii) \quad Q_2^1 : \frac{p_2}{e_1 - p_1 Q_1^1 - p_2 Q_2^1} = \frac{1}{4Q_2^1}$$

(i) and (ii) together imply $p_2 Q_2^1 = \frac{1}{3} p_1 Q_1^1$. This is not surprising ; agent 1 places a higher subjective probability on the first state. Taking this into account, (i) implies $p_1 Q_1^1 = \frac{3}{8} e_1$ and thus $p_2 Q_2^1 = \frac{1}{8} e_1$.

The calculations thus far are symmetric and (iii) and (iv) solve for $p_1 Q_1^2 = \frac{1}{8} e_1$ while $p_2 Q_2^2 = \frac{3}{8} e_2$.

Now suppose there is 1 unit of each security available for purchase.

$$a) \quad Q_1^1 + Q_1^2 = 1$$

$$b) \quad Q_2^1 + Q_2^2 = 1$$

Substituting the above demand function into these equations gives :

$$a) \quad \left(\frac{3}{8p_1} \right) e_1 + \left(\frac{1}{8p_1} \right) e_2 = 1$$

$$p_1 = \frac{3}{8} e_1 + \frac{1}{8} e_2$$

$$b) \quad \left(\frac{3}{8p_2} \right) e_1 + \left(\frac{1}{8p_2} \right) e_2 = 1$$

$$p_2 = \frac{1}{8} e_1 + \frac{3}{8} e_2$$

If $e_2 > e_1$, then $p_2 > p_1$;

If $e_2 < e_1$, then $p_1 > p_2$

b. There is now only one security

	t=1	t=2
$-p^b$	θ_1	θ_2
	1	1

Agent Endowments

	t=1	t=2
Agent 1	e_1	a_1
Agent 2	e_2	a_2

Now, there can be trade here even with this one asset, if, say $e_1 = 0$, $a_1 > 0$, $e_2 > 0$, $a_2 = 0$ to take an extreme case.

If $e_1 = e_2 = 0$, then there will be no trade as the security payoff do not allow the agents to tailor their consumption plans to their subjective probabilities.

c. Suppose we introduce a risky asset

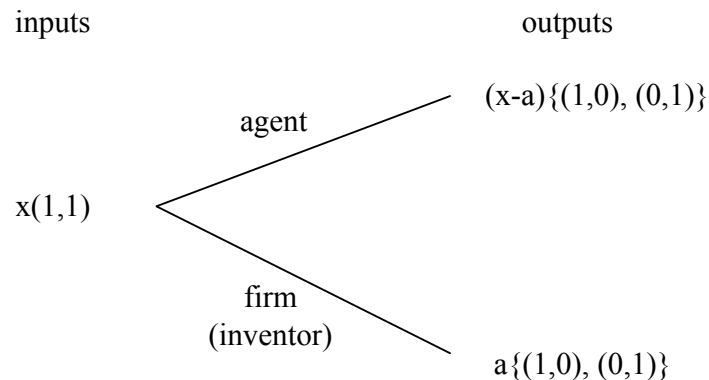
t=0	t=1	
	θ_1	θ_2
-p	z_1	z_2

where $z_1 \neq z_2, z_1, z_2 > 0$.

Combinations of this security and the riskless one can be used to construct the state claims. This will be welfare improving relative to the case where only the riskless asset is traded. The final equilibrium outcome and the extent to which each agents welfare is improved will depend upon the relative endowments of the risky security assigned to each agent ; and the absolute total quantity bestowed on the economy.

d. The firm can convert x units of $(1,1)$ into x units of $\{(1,0), (0,1)\}$. These agents (relative to having only the riskless asset) would avail themselves of this technology, and then trade the resultant claims to attain a more preferred consumption state.

Furthermore, the agents would be willing to pay for such a service in the following sense :



Clearly, if $a = x$, the agents would ignore the inventor. However, each agent would be willing to pay something. Assuming the inventor charges the same a to each agent, the most he could charge would be that a at which one of the agents were no better off ex ante than if he did not trade.

Suppose the inventor could choose to convert $x, 2x, 3x, \dots, nx$ securities (x understood to be small). The additional increment he could charge would decline as n increased.