Lecture Notes in Financial Economics

© by Antonio Mele
The London School of Economics and Political Science

January 2007

Contents

“Many of the models in the literature are not general equilibrium models in my sense. Of those that are, most are intermediate in scope: broader than examples, but much narrower than the full general equilibrium model. They are narrower, not for carefully-spelled-out economic reasons, but for reasons of convenience. I don’t know what to do with models like that, especially when the designer says he imposed restrictions to simplify the model or to make it more likely that conventional data will lead to reject it. The full general equilibrium model is about as simple as a model can be: we need only a few equations to describe it, and each is easy to understand. The restrictions usually strike me as extreme. When we reject a restricted version of the general equilibrium model, we are not rejecting the general equilibrium model itself. So why bother testing the restricted version?”

Fischer Black, 1995, p. 4, *Exploring General Equilibrium*, The MIT Press.

Preface

The present *Lecture Notes in Financial Economics* are based on my teaching notes for advanced undergraduate and graduate courses on financial economics, macroeconomic dynamics and financial econometrics. These *Lecture Notes* are still too underground. Many derivations are inelegant, proofs and exercises are not always separated from the main text, economic motivation and intuition are not developed as enough as they deserve, and the English is informal. Moreover, I didn't include (yet) material on asset pricing with asymmetric information, monetary models of asset prices, and asset prices determination within overlapping generation models; or on more applied topics such as credit risk - and their related derivatives. Finally, I need to include more extensive surveys for each topic I cover. I plan to revise my *Lecture Notes* in the near future. Naturally, any comments on this version are more than welcome.

Antonio Mele
January 2007

Part I

Foundations

1

The classical capital asset pricing model

1.1 Static portfolio selection problems

1.1.1 The wealth constraint

We consider an economy with m risky assets, and some safe asset. Let $q_1, \cdots, q_m$ be the prices associated to the risky assets and q_0 the price of the riskless asset. We wish to build up a portfolio of all these assets. Let θ_i, $i = 0, 1 \cdots, m$, be the number of asset no. i in this portfolio. Initial wealth is simply $w = q_0\theta_0 + q \cdot \theta = q_0\theta_0 + \sum_{i=1}^m q_i\theta_i$. Final wealth is $w^+ = x_0\theta_0 + \sum_{i=1}^m x_i\theta_i$, where x_i is the payoff promised by the i-th asset, viz

$$w^+ = x_0\theta_0 + \sum_{i=1}^m x_i\theta_i = \frac{x_0}{q_0}\theta_0 q_0 + \sum_{i=1}^m \frac{x_i}{q_i}\theta_i q_i \equiv R\pi_0 + \sum_{i=1}^m \tilde{R}_i\pi_i,$$

where $\pi_i \equiv \theta_i q_i$ is the wealth invested in the i-th asset, $R \equiv \frac{x_0}{q_0}$, and $\tilde{R}_i \equiv \frac{x_i}{q_i}$. Let $\tilde{b}_i \equiv \tilde{R}_i - 1$, $r \equiv R - 1$ and $b \equiv E(\tilde{b})$. We have $w = \pi_0 + \sum_{i=1}^m \pi_i$, and,

$$\begin{aligned} w^+ &= Rw + \pi^\top(\tilde{R} - \mathbf{1}_m R) = Rw + \pi^\top(\tilde{R} - \mathbf{1}_m - \mathbf{1}_m(R-1)) \\ &= Rw + \pi^\top(\tilde{b} - \mathbf{1}_m r) = Rw + \pi^\top(b - \mathbf{1}_m r) + \pi^\top(\tilde{b} - b), \end{aligned}$$

where π is a m-dimensional vector. For $m \leq d$,

$$\underset{m\times 1}{\tilde{b}} = \underset{m\times d}{a} \cdot \underset{d\times 1}{\tilde{\epsilon}} \quad \Rightarrow \quad \tilde{b} - b = a \cdot \epsilon$$

where a is a matrix of constants and $\tilde{\epsilon}$ is distributed according to a given distribution with expectation $\bar{\epsilon}$ and variance-covariance matrix equal to the identity matrix.

We have,

$$w^+ = Rw + \pi^\top(b - \mathbf{1}_m r) + \pi^\top a\epsilon, \tag{1.1}$$

where ϵ has the same distribution as $\tilde{\epsilon}$, but expectation zero.[1]

[1] Eq. (2.?) is a very useful formulation of the problem and will be used to derive multiperiod models. It already reveals that something very intuitive must happen: there must be a $\lambda : b - \mathbf{1}_m r = a\lambda$, which implies that $E(w^+) = Rw + \pi' a\lambda$. See chapter 5 for further details.

The objective now is to maximize the expectation of a function of w^+ w.r.t. π. We have,

$$\begin{aligned} E\left[w^+(\pi)\right] &= Rw + \pi^\top (b - \mathbf{1}_m r) \\ var\left[w^+(\pi)\right] &= \pi^\top \sigma \pi \end{aligned}$$

where $\sigma \equiv aa^\top$. Let $\sigma_i^2 \equiv \sigma_{ii}$. We assume that σ has full-rank, and that,

$$\sigma_i^2 > \sigma_j^2 \Rightarrow b_i > b_j \text{ all } i, j, \tag{1.2}$$

which implies that $r < \min_j(b_j)$.

1.1.2 The program

The program is:

$$\left| \begin{array}{l} \hat{\pi} = \arg\max_{\pi \in \mathbb{R}^m} E\left\{w^+(\pi)\right\} \\ \text{s.t. } var\left[w^+(\pi)\right] = w^2 \cdot v_p^2 \quad (\nu) \end{array} \right.$$

where ν is the Lagrange multiplier. Let $L = Rw + \pi^\top(b - \mathbf{1}_m r) - \nu(\pi^\top \sigma \pi - w^2 \cdot v_p^2)$. The first order conditions are,

$$\begin{cases} \hat{\pi} &= \dfrac{1}{2\nu}\sigma^{-1}(b - \mathbf{1}_m r) \\ \hat{\pi}^\top \sigma \hat{\pi} &= w^2 \cdot v_p^2 \end{cases}$$

By plugging the first condition into the second condition,

$$\frac{w^2 \cdot v_p^2}{Sh} = \left(\frac{1}{2\nu}\right)^2$$

where

$$Sh \equiv (b - \mathbf{1}_m r)^\top \sigma^{-1} (b - \mathbf{1}_m r),$$

is the *Sharpe market performance*.[2] We have $\dfrac{1}{2\nu} = \mp \frac{w \cdot v_p}{\sqrt{Sh}}$, and we take the positive solution to ensure efficiency.

Substituting $\dfrac{1}{2\nu} = \frac{w \cdot v_p}{\sqrt{Sh}}$ back into the first condition,

$$\frac{\hat{\pi}}{w} = \frac{\sigma^{-1}(b - \mathbf{1}_m r)}{\sqrt{Sh}} \cdot v_p.$$

Now the value of the problem is $E\left[w^+(\hat{\pi}(w \cdot v_p))\right]$, and following a standard convention we define the *expected portfolio return* as:

$$\mu_p(v_p) \equiv \frac{E\left[w^+(\hat{\pi}(v_p))\right] - w}{w}.$$

Some simple computations leave,

$$\mu_p(v_p) = r + \sqrt{Sh} \cdot v_p, \tag{1.3}$$

where now v_p^2 is easily seen as being the *portfolio return variance*:

$$var\left[\frac{w^+(\hat{\pi}) - w}{w}\right] = var\left[\frac{w^+(\hat{\pi})}{w}\right] = \frac{1}{w^2}\hat{\pi}^\top \sigma \hat{\pi} = v_p^2.$$

The relation in eq. (1.??) is known as the *Capital Market Line* (CML).

[2] Sharpe ratios on individual assets are defined as $\frac{b_i - r}{\sigma_i}$.

1.1.3 The program without a safe asset

In this case we have simply that $w = \sum_{i=1}^{m} q_i \theta_i$ and $w^+ = \sum_{i=1}^{m} x_i \theta_i$, where x_i is the payoff going to asset i: $w^+ = \sum_{i=1}^{m} x_i \theta_i = \sum_{i=1}^{m} \frac{x_i}{q_i} \theta_i q_i \equiv \sum_{i=1}^{m} \tilde{R}_i \pi_i = \sum_{i=1}^{m} \tilde{R}_i \pi_i = \sum_{i=1}^{m} (\tilde{R}_i - 1 + 1)\pi_i = \sum_{i=1}^{m} \tilde{b}_i \pi_i + \sum_{i=1}^{m} \pi_i = \sum_{i=1}^{m} (b + \tilde{b}_i - b)\pi_i + \sum_{i=1}^{m} \pi_i = \pi^\top b + \pi^\top a\epsilon + w$. Finally, $w = \pi^\top \mathbf{1}_m$. We have:

$$\begin{aligned} E\left[w^+(\pi)\right] &= \pi^\top b + w,\ w = \pi^\top \mathbf{1}_m \\ var\left[w^+(\pi)\right] &= \pi^\top \sigma \pi \end{aligned}$$

The program is then:

$$\left| \begin{array}{l} \hat{\pi} = \arg\max_{\pi \in \mathbb{R}} E\left\{w^+(\pi)\right\} \\ \text{s.t.} \left\{ \begin{array}{l} var\left[w^+(\pi)\right] = w^2 \cdot v_p^2 \quad (\nu_1) \\ \pi^\top \mathbf{1}_m = w \quad (\nu_2) \end{array} \right. \end{array} \right.$$

where ν_i are Lagrange multipliers.

The appendix shows that the solution for the portfolio is:

$$\begin{aligned} \frac{\hat{\pi}}{w} &= \frac{\sigma^{-1}\mathbf{1}_m}{\gamma} + \frac{1}{\alpha\gamma - \beta^2}\left(\gamma\mu_p(v_p) - \beta\right)\left(\sigma^{-1}b - \frac{\sigma^{-1}\beta}{\gamma}\mathbf{1}_m\right) \\ &= \frac{\gamma\mu_p(v_p) - \beta}{\alpha\gamma - \beta^2}\sigma^{-1}b + \frac{\alpha - \beta\mu_p(v_p)}{\alpha\gamma - \beta^2}\sigma^{-1}\mathbf{1}_m, \end{aligned}$$

where $\alpha \equiv b^\top \sigma^{-1} b$, $\beta \equiv \mathbf{1}_m^\top \sigma^{-1} b$ and $\gamma \equiv \mathbf{1}_m^\top \sigma^{-1} \mathbf{1}_m$.

Furthermore, let the normalized value of the program be:

$$\mu_p(v_p) \equiv \frac{E\left[w^+(\hat{\pi})\right] - w}{w}.$$

The appendix also shows that:

$$v_p^2 = \frac{1}{\gamma}\left[1 + \frac{1}{\alpha\gamma - \beta^2}\left(\gamma\mu_p(v_p) - \beta\right)^2\right],$$

and given that $\alpha\gamma - \beta^2 > 0$, the *global minimum variance portfolio* achieves variance $v_p^2 = \gamma^{-1}$ and expected return $\mu_p = \beta/\gamma$. For each v_p, there are two values of $\mu_p(v_p)$ that solve equation (2.??). Clearly the optimal choice is the one with higher μ_p, which implies that the *efficient portfolios frontier* (2.??) in the (v_p, μ_p)-space is positively sloped.

To summarize, we had to solve the following problem:

$$\left| \begin{array}{l} \hat{\pi} = \arg\max_{\pi \in \mathbb{R}} E\left[w^+(\pi)\right] \\ \text{s.t.} \left\{ \begin{array}{l} var\left[w^+(\pi)\right] = v_p^2 \\ \pi^\top \mathbf{1}_m = w \end{array} \right. \end{array} \right.$$

and the solution was of the form: $\hat{\pi}(v_p)$, from which one obtains the map:

$$v_p \mapsto E\left[w^+\left(\hat{\pi}(v_p)\right)\right].$$

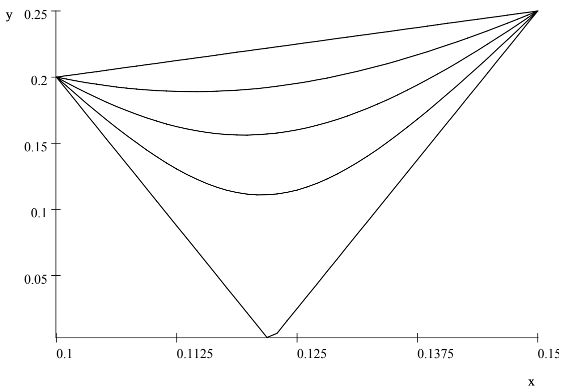


FIGURE 1.1. Shown on the X-asis is μ_p and shown on the Y-axis is v_p. From top to bottom: portfolio frontiers corresponding to $\rho = 1, 0.5, 0, -0.5, -1$. Parameters are set to $b_1 = 0.10$, $b_2 = 0.15$, $\sigma_1 = 0.20$, $\sigma_2 = 0.25$. For each frontier, efficient portfolios are those yielding the lowest volatility for a given return.

It's a concave function, and it can be interpreted as a sort of "production function": it produces expected returns using levels of risk as inputs (see, e.g., figure 2.3 below). The choice of *which* portfolio has effectively to be selected then depends on agents' preferences.

EXAMPLE 1.1. Suppose $m = 2$. Here there is not need to optimize. We have $\frac{E\left\{w^+(\pi)\right\}-w}{w} = \frac{\pi_1}{w}b_1 + \frac{\pi_2}{w}b_2$, with $\frac{\pi_1}{w} + \frac{\pi_2}{w} = 1$, and then

$$\begin{cases} \dfrac{E\left[w^+(\pi)\right]-w}{w} & \equiv \mu_p = & b_1 + (b_2 - b_1)\dfrac{\pi_2}{w} \\ var\left[\dfrac{w^+(\pi)}{w}\right] & \equiv v_p^2 = & \left(1-\dfrac{\pi_2}{w}\right)^2\sigma_1^2 + 2\left(1-\dfrac{\pi_2}{w}\right)\dfrac{\pi_2}{w}\sigma_{12} + \left(\dfrac{\pi_2}{w}\right)^2\sigma_2^2 \end{cases}$$

whence:

$$v_p = \frac{1}{b_2 - b_1}\sqrt{\left(b_2-\mu_p\right)^2\sigma_1^2 + 2\left(b_2-\mu_p\right)\left(\mu_p - b_1\right)\rho\sigma_1\sigma_1 + \left(\mu_p - b_1\right)^2\sigma_2^2}$$

When $\rho = 1$,

$$\mu_p = b_1 + \frac{(b_1 - b_2)(\sigma_1 - v_p)}{\sigma_2 - \sigma_1}.$$

In the general case, diversification pays when asset returns are not perfectly positively correlated (see figure 2.3). It is even possible to obtain a portfolio that is less risky than than the less risky asset. And risk can be zeroed with $\rho = -1$. In some cases,

Next, we turn to portfolio issues: it is easily ckecked that

$$\frac{\hat{\pi}}{w} = \ell_1\frac{\pi_d}{w} + \ell_2\frac{\pi_g}{w}, \quad \ell_1 + \ell_2 = 1,$$

where

$$\begin{cases} \ell_1 & \equiv & \dfrac{\nu_1 w}{2}\beta & = & \dfrac{\mu_p\gamma - \beta}{\alpha\gamma - \beta^2}\beta \\ \ell_2 & \equiv & \dfrac{\nu_2 w}{2}\gamma & = & \dfrac{\alpha - \beta\mu_p}{\alpha\gamma - \beta^2}\gamma \end{cases}$$

and

$$\begin{cases} \dfrac{\pi_d}{w} & \equiv & \dfrac{\sigma^{-1} b}{\beta}, \quad \beta \equiv \mathbf{1}_m^\top \sigma^{-1} b \\ \dfrac{\pi_g}{w} & \equiv & \dfrac{\sigma^{-1} \mathbf{1}_m}{\gamma}, \quad \gamma \equiv \mathbf{1}_m^\top \sigma^{-1} \mathbf{1}_m \end{cases}$$

$\frac{\pi_g}{w}$ is the *global minimum variance portfolio* because minimum variance occurs at $(v, \mu) = \left(\sqrt{\frac{1}{\gamma}}, \frac{\beta}{\gamma}\right)$, in which case $\ell_1 = 0$ and $\ell_2 = 1$. In general, any portfolio on the frontier can be obtained by letting ℓ_1 and ℓ_2 vary and using $\frac{\pi_d}{w}$ and $\frac{\pi_g}{w}$ as instruments. It's a *Mutual-Funds theorem.*

1.1.4 The market, or "tangency", portfolio

Definition 1.2. *The market portfolio is the portfolio at which the CML* (2.??) *and the efficient portfolios frontier* (2.??) *intersect.*

In fact, the market portfolio is the point at which the CML is *tangent* at the efficient portfolio frontier. This is so because agents have access to wider possibilities of choice on the CML (all risky assets *plus* the riskless asset). The existence of the market portfolio requires a restriction on R. Let (v_M, μ_M) be the market portfolio, and suppose that it exists. As figure 2.4 shows, the CML dominates the efficient portfolio frontier AMC. The important point here is that any point on the CML is a combination of safe assets with the market portfolio M. An investor with high risk-aversion would like to choose a point such as Q, say; and an investor with low risk-aversion would like to choose a point such as P, say. But no matter how risk-adverse an individual is, she will always have the interest to choose a combination of safe assets with the "pivotal", market portfolio M. In other terms, the market portfolio doesn't depend on the risk-attitudes of any investor. It's a *two-funds separation theorem.*

Finally, the dotted line MZ represents the continuation of the rM line when the interest rate for borrowing is higher than the interest rate for lending. Until M, the CML is still rM. From M onwards, the resulting CML is then the one that dominates between MZ and MA. As an example, the CML compatible with the scheme shown in the figure is the rMA curve.

We assume that

$$r < \frac{\beta}{\gamma}.$$

To characterize the market portfolio analytically, we have two analytical strategies:

- The first one is perhaps the best known in the literature: the tangency portfolio π_M belongs to AMC if $\pi_M^\top \mathbf{1}_m = w$, where π_M also belongs to CML and is therefore such that:

$$\frac{\pi_M}{w} = \frac{\sigma^{-1}(b - \mathbf{1}_m r)}{\sqrt{Sh}} \cdot v_M.$$

  Therefore, we must be looking for the value v_M which solves

$$w = \mathbf{1}_m^\top \pi_M = w \cdot \mathbf{1}_m^\top \frac{\sigma^{-1}(b - \mathbf{1}_m r)}{\sqrt{Sh}} \cdot v_M,$$

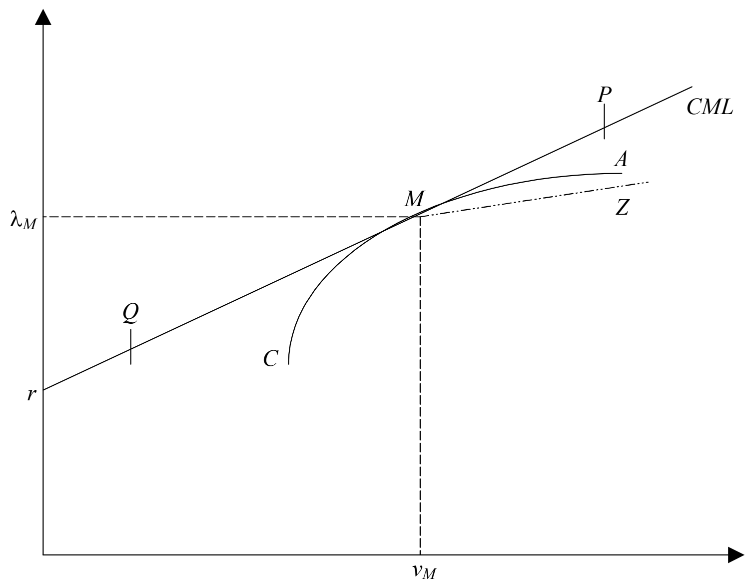


FIGURE 1.2.

i.e.

$$v_M = \frac{\sqrt{Sh}}{\mathbf{1}_m^\top \sigma^{-1}(b - \mathbf{1}_m r)} = \frac{\sqrt{Sh}}{\beta - \gamma r},$$

and plug it back into the expression of π_M to obtain:

$$\frac{\pi_M}{w} = \frac{\sigma^{-1}(b - \mathbf{1}_m r)}{\mathbf{1}_m^\top \sigma^{-1}(b - \mathbf{1}_m r)} = \frac{1}{\beta - \gamma r}\sigma^{-1}(b - \mathbf{1}_m r). \qquad (1.4)$$

Of course nothing is invested in the riskless asset with π_M. Furthermore, it belongs to the efficient portfolio frontier for two reasons: 1) It is not above it because this would contradict the efficiency of AMC (which is obtained by only investing in risky assets); 2) It is not below because by construction it belongs to the CML which, as shown before, dominates the efficient portfolio frontier.

- The second analytical strategy consists in directly exploiting the tangency condition of CML with AMC at point M:

$$\text{slope\_of\_}CML = \sqrt{Sh} = \frac{\alpha\gamma - \beta^2}{\gamma\mu_M - \beta} v_M = \text{slope\_of\_}AMC$$

where we used the fact that $\left.\frac{\partial v}{\partial \mu}\right|_M = \frac{1}{v_M}\frac{\gamma\mu_M - \beta}{\alpha\gamma - \beta^2}$. After using $\mu_M = r + \sqrt{Sh} \cdot v_M$ and rearranging terms:

$$\begin{cases} \mu_M = r + \sqrt{Sh} \cdot v_M \\ v_M = \dfrac{(\gamma r - \beta)\sqrt{Sh}}{\alpha\gamma - \beta^2 - \gamma \cdot Sh} = \dfrac{\sqrt{Sh}}{\beta - \gamma r} \end{cases}$$

which is exactly what found in the previous point.

The previous considerations now allow us to justify why the tangency portfolio is called "market portfolio". As it is clear, any portfolio can be attained by investing in zero-net supply lending/borrowing funds and in portfolio M. Therefore, in this mean-variance economy, everyone is holding some proportions of M and since in aggregate there is no net borrowing or lending, one has that in aggregate, all agents have portfolio holdings that sum up to the market portfolio, which is therefore the value-weighted portfolio of all assets in the economy. There are important connections between results on the market portfolio and results for dynamic models to be presented in later chapters.

1.2 The CAPM

The CAPM (Capital Asset Pricing Model) provides an asset evaluation formula. Here we follow the construction of Sharpe (1964).[3] We work directly with portfolio *returns*. Create an α-parametrized portfolio which has α units of wealth invested in asset i and $1-\alpha$ units of wealth invested in the market portfolio:

$$\begin{cases} \tilde{\mu}_p & \equiv & \alpha b_i + (1-\alpha)\mu_M \\ \tilde{v}_p & \equiv & \sqrt{(1-\alpha)^2\sigma_M^2 + 2(1-\alpha)\alpha\sigma_{iM} + \alpha^2\sigma_i^2} \end{cases} \tag{1.5}$$

where $\sigma_M \equiv v_M$. Clearly point M in the (v_P, μ_P)-space belongs to such an α-parametrized curve. By example 2.6, its geometric structure is then as in curve $A'Mi$ in figure 2.5. The reason for which curve $A'Mi$ lies below curve AMC is due to diversification: AMC can be obtained through all existing assets and must clearly dominate the frontier which can be obtained with only the two assets i and M. In other terms, if curve $A'Mi$ was to intersect curve AMC, this would mean that a feasible combination of assets (composed by a proportion α of asset i and a proportion $1-\alpha$ of assets in the portfolio: the sum will be 1, again!) dominates AMC, which is impossible because AMC is, by construction, the most efficient, and feasible combination of assets. For the same reason, $A'Mi$ cannot intersect AMC (otherwise $A'Mi$ could dominate AMC in some region). Therefore, $A'Mi$ is tangent at AMC in M, which is itself tangent at the CML in M by the analysis of the previous section.

The idea now is to equate the two slopes of $A'Mi$ and AMC in M and derive a restriction on the expected returns b_i. Because (2.??) is an α-parametrized curve, it's enough to compute the two objects $d\tilde{\mu}_p/\,d\alpha$ and $d\tilde{v}_p/\,d\alpha$ at $\alpha = 0$. We have

$$\frac{d\tilde{\mu}_p}{d\alpha} = b_i - \mu_M, \text{ all } \alpha,$$

and $d\tilde{v}_p/\,d\alpha = (-(1-\alpha)\sigma_M^2 + (1-2\alpha)\sigma_{iM} + \alpha\sigma_i^2)/\,\tilde{v}_p$, from which we get:

$$\left.\frac{d\tilde{v}_p}{d\alpha}\right|_{\alpha=0} = \frac{1}{\tilde{v}_p|_{\alpha=0}}\left(\sigma_{iM} - \sigma_M^2\right) = \frac{1}{\sigma_M}\left(\sigma_{iM} - \sigma_M^2\right).$$

[3] Sharpe, W.F. (1964): "Capital Asset Prices: a Theory of Market Equilibrium under Conditions of Risk," *Journal of Finance*, Vol. XIX, 3, 425-442.

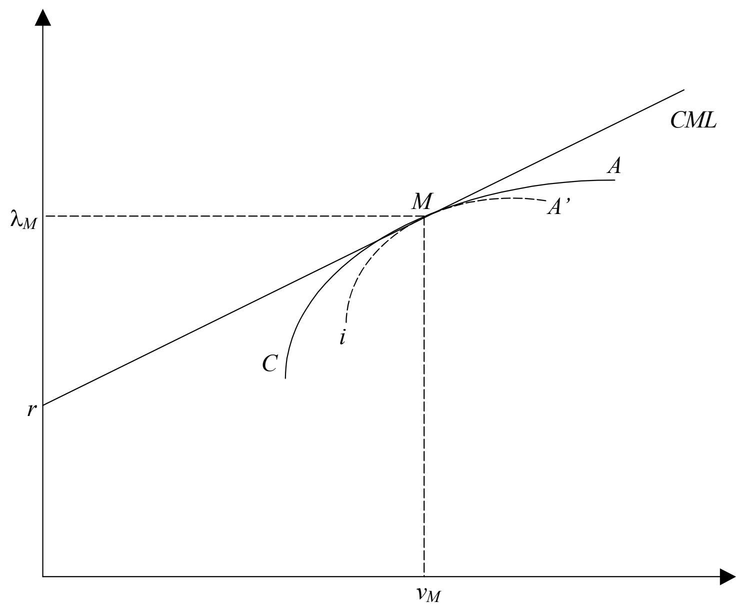


FIGURE 1.3.

Therefore,

$$\left.\frac{d\tilde{\mu}_p(\alpha)}{d\tilde{v}_p(\alpha)}\right|_{\alpha=0} = \frac{b_i - \mu_M}{\frac{1}{\sigma_M}\left(\sigma_{iM} - \sigma_M^2\right)}. \tag{1.6}$$

On the other hand, the slope of the CML is $(\mu_M - r)/\, v_M$ and by comparing such a slope with (2.??) we obtain by rearranging terms $b_i - \mu_M + r - r = (\mu_M - r)(\sigma_{iM} - v_M^2)/\, v_M^2$, or

$$b_i - r = \beta_i\left(\mu_M - r\right),\ \beta_i \equiv \frac{\sigma_{iM}}{v_M^2},\ i = 1, \cdots, m. \tag{1.7}$$

The previous relation is called the *Security Market Line* (SML).

An alternative derivation of the SML is the following one. Recall that $\frac{\pi_M}{w} = \frac{1}{\beta-\gamma r}\sigma^{-1}\left(b - \mathbf{1}_m r\right)$. Compute the vector of covariances of the m asset returns with the market portfolio:

$$cov\left(\tilde{x}, \tilde{x}_M\right) = cov\left(\tilde{x}, \tilde{x}\frac{\pi_M}{w}\right) = \sigma\frac{\pi_M}{w} = \frac{1}{\beta - \gamma r}\left(b - \mathbf{1}_m r\right). \tag{1.8}$$

Premultiply the previous equation by $\frac{\pi_M^\top}{w}$ to obtain:

$$v_M^2 = \frac{\pi_M^\top}{w}\sigma\frac{\pi_M}{w} = \frac{\pi_M^\top}{w}\frac{1}{\beta - \gamma r}\left(b - \mathbf{1}_m r\right) = \frac{1}{(\beta - \gamma r)^2}Sh,$$

or

$$v_M = \frac{\sqrt{Sh}}{\beta - \gamma r}, \quad \text{not new.} \tag{1.9}$$

By (2.26),

$$\sigma_{iM} \equiv cov\left(\tilde{x}_i, \tilde{x}_M\right) = \frac{1}{\beta - \gamma r}\left(b_i - r\right), \quad i = 1, \cdots, m.$$

By replacing (2.27) into the previous equation and rearranging:

$$b_i = r + \frac{\sqrt{Sh}\sigma_{iM}}{v_M}, \quad i = 1, \cdots, m.$$

But we also know that $\sqrt{Sh} = \frac{\mu_M - r}{v_M}$, and replacing it into the previous equation gives the result:

$$b_i = r + \frac{\sigma_{iM}}{v_M^2} (\mu_M - r), \quad i = 1, \cdots, m.$$

Note, the SML can also be interpreted as a projection of the excess returns on asset i (i.e. $\tilde{b}_i - r$) on the excess returns on the market portfolio (i.e. $\tilde{b}_M - r$):

$$\tilde{b}_i - r = \beta (\mu_M - r) + \varepsilon_i, \quad i = 1, \cdots, m,$$

from which we get

$$\sigma_i^2 = \beta^2 v_M + var(\varepsilon_i), \quad i = 1, \cdots, m.$$

The quantity $\beta^2 v_M$ is *systematic* risk, and $var(\varepsilon_i)$ is *non-systematic, idiosynchratic risk* which can be eliminated with diversification. (As the # assets goes to infinity. See APT and factor analysis below for a general analysis of this phenomenon.)

Assets with $\beta_i > 1$ may be called "aggressive" assets; assets with $\beta_i < 1$ may be called "conservative" assets.

Some notes: recall that every asset must lie below the frontier. After the construction of the frontier, the assets must still lie under the frontier, because the frontier itself was constructed with the assets. If, for some reasons, some of the assets were *on* the frontier under the construction of the frontier, the frontier itself should also change to reflect such asset changes.

The CAPM can also be used to evaluate risky projects. Let

$$V = \text{value of a project} = \frac{E(C^+)}{1 + r_C},$$

where C^+ is future cash flow and r_C is the risk-adjusted discount rate for this project. This is a standard MBA textbook formula.

We have:

$$\begin{aligned} \frac{E(C^+)}{V} &= 1 + r_C \\ &= 1 + r + \beta_C (\mu_M - r) \\ &= 1 + r + \frac{cov\left(\frac{C^+}{V} - 1, \tilde{x}_M\right)}{v_M^2} (\mu_M - r) \\ &= 1 + r + \frac{1}{V} \frac{cov(C^+, \tilde{x}_M)}{v_M^2} (\mu_M - r) \\ &= 1 + r + \frac{1}{V} cov\left(C^+, \tilde{x}_M\right) \frac{\lambda}{v_M}, \end{aligned}$$

where $\lambda \equiv \frac{\mu_M - r}{v_M}$, the unit market risk-premium.

Rearranging terms in the previous equation leaves:

$$V = \frac{E(C^+) - \frac{\lambda}{v_M} cov(C^+, \tilde{x}_M)}{1 + r}. \tag{1.10}$$

The certainty equivalent $\bar{C}$ is defined as:

$$\bar{C} : V = \frac{E\left(C^{+}\right)}{1+r_{C}} = \frac{\bar{C}}{1+r},$$

or,

$$\bar{C} = (1+r)\, V,$$

and using relation (2.??),

$$\bar{C} = E\left(C^{+}\right) - \frac{\lambda}{v_{M}} cov\left(C^{+}, \tilde{x}_{M}\right).$$

1.3 Appendix 1: Analytics details for the mean-variance portfolio choice

1.3.1 The primal program

Let $L = \pi^\top b + w - \nu_1(\pi^\top \sigma \pi - w^2 \cdot v_p^2) - \nu_2(\pi^\top \mathbf{1}_m - w)$. The first order conditions are,

$$\begin{aligned} \hat{\pi} &= \frac{1}{2\nu_1}\sigma^{-1}(b - \nu_2 \mathbf{1}_m) \\ \hat{\pi}^\top \sigma \hat{\pi} &= w^2 \cdot v_p^2 \\ \hat{\pi}^\top \mathbf{1}_m &= w \end{aligned}$$

Using the first and the third of the previous first order conditions,

$$w = \mathbf{1}_m^\top \hat{\pi} = \frac{1}{2\nu_1}(\underbrace{\mathbf{1}_m^\top \sigma^{-1} b}_{\equiv \beta} - \nu_2 \underbrace{\mathbf{1}_m^\top \sigma^{-1} \mathbf{1}_m}_{\equiv \gamma}) \equiv \frac{1}{2\nu_1}(\beta - \nu_2 \gamma),$$

and then:

$$\nu_2 = \frac{\beta - 2w\nu_1}{\gamma}.$$

By replacing back into the portfolio first order condition we get:

$$\hat{\pi} = \frac{w}{\gamma}\sigma^{-1}\mathbf{1}_m + \frac{1}{2\nu_1}\sigma^{-1}\left(b - \frac{\beta}{\gamma}\mathbf{1}_m\right).$$

Now

$$E\left\{w^+(\hat{\pi})\right\} - w = \hat{\pi}^\top b = \frac{w}{\gamma}\underbrace{\mathbf{1}_m^\top \sigma^{-1} b}_{\equiv \beta} + \frac{1}{2\nu_1}(\underbrace{b^\top \sigma^{-1} b}_{\equiv \alpha} - \frac{\beta}{\gamma}\underbrace{\mathbf{1}_m^\top \sigma^{-1} b}_{\equiv \beta}) = \frac{w}{\gamma}\beta + \frac{1}{2\nu_1}\left(\alpha - \frac{\beta^2}{\gamma}\right),$$

and

$$\begin{aligned} var\left\{w^+(\hat{\pi})\right\} &= \hat{\pi}^\top \sigma \hat{\pi} \\ &= \left[\frac{w}{\gamma}\mathbf{1}_m^\top \sigma^{-1} + \frac{1}{2\nu_1}\left(b^\top - \frac{\beta}{\gamma}\mathbf{1}_m^\top\right)\sigma^{-1}\right]\left[\frac{w}{\gamma}\mathbf{1}_m + \frac{1}{2\nu_1}\left(b - \frac{\beta}{\gamma}\mathbf{1}_m\right)\right] \\ &= \frac{w^2}{\gamma} + \left(\frac{1}{2\nu_1}\right)^2\left(\alpha - \frac{\beta^2}{\gamma}\right) \\ &= w^2 \cdot v_p^2. \end{aligned}$$

Therefore, by defining $\mu_p(v_p) \equiv \frac{E\{w^+(\hat{\pi})\} - w}{w}$ we get:

$$\begin{aligned} \mu_p(v_p) &= \frac{\beta}{\gamma} + \frac{1}{2\nu_1 w}\left(\alpha - \frac{\beta^2}{\gamma}\right) \\ v_p^2 &= \frac{1}{\gamma} + \left(\frac{1}{2\nu_1 w}\right)^2\left(\alpha - \frac{\beta^2}{\gamma}\right) \end{aligned} \tag{1.11}$$

with the usual interpretation of v_p^2.

The first condition in (2.??) can be solved for $2\nu_1 w$:

$$\frac{1}{2\nu_1 w} = \left(\alpha\gamma - \beta^2\right)^{-1}\left(\gamma\mu_p(v_p) - \beta\right),$$

from which we get the solution for the portfolio:

$$\frac{\hat{\pi}}{w} = \frac{\sigma^{-1}\mathbf{1}_m}{\gamma} + \left(\alpha\gamma - \beta^2\right)^{-1}\left(\gamma\mu_p(v_p) - \beta\right)\left(\sigma^{-1}b - \frac{\sigma^{-1}\beta}{\gamma}\mathbf{1}_m\right).$$

Also, by substituting $2\nu_1 w$ into the second condition in (2.??) leaves:

$$v_p^2 = \frac{1}{\gamma}\left[1 + \left(\alpha\gamma - \beta^2\right)^{-1}\left(\gamma\mu_p(v_p) - \beta\right)^2\right]. \tag{1.12}$$

The second condition in (2.??) also reveals that:

$$\left(\frac{1}{2\nu_1 w}\right)^2 = \frac{\gamma v_p^2 - 1}{\alpha\gamma - \beta^2},$$

and given that $\alpha\gamma - \beta^2 > 0$,[4] we may then confirm the properties of the *global minimum variance portfolio* stated in sect. 2.?.

1.3.2 The dual program

Naturally, the previous results can be obtained by solving the dual program:

$$\left|\begin{array}{l} \hat{\pi} = \arg\min_{\pi\in\mathbb{R}^m} var\left[\dfrac{w^+(\pi)}{w}\right] \\ \text{s.t.} \left\{\begin{array}{l} E\left\{w^+(\pi)\right\} = E_p \quad (\nu_1) \\ \pi^\top \mathbf{1}_m = w \quad (\nu_2) \end{array}\right. \end{array}\right.$$

Set $L = \frac{1}{w^2}\pi^\top\sigma\pi - \nu_1(\pi^\top b + w - E_p) - \nu_2(\pi^\top\mathbf{1}_m - w)$. The first order conditions are

$$\left\{\begin{array}{lcl} \dfrac{\hat{\pi}}{w^2} & = & \dfrac{\nu_1}{2}\sigma^{-1}b + \dfrac{\nu_2}{2}\sigma^{-1}\mathbf{1}_m \\ \hat{\pi}^\top b & = & E_p - w \\ \hat{\pi}^\top\mathbf{1}_m & = & w \end{array}\right.$$

By replacing the first relation into the second one,

$$E_p - w = \hat{\pi}^\top b = w^2\Big(\frac{\nu_1}{2}\underbrace{b^\top\sigma^{-1}b}_{\equiv\alpha} + \frac{\nu_2}{2}\underbrace{\mathbf{1}_m^\top\sigma^{-1}b}_{\equiv\beta}\Big) \equiv w^2\left(\frac{\nu_1}{2}\alpha + \frac{\nu_2}{2}\beta\right),$$

and by replacing the first relation into the third one,

$$w = \hat{\pi}^\top\mathbf{1}_m = w^2\Big(\frac{\nu_1}{2}\underbrace{b^\top\sigma^{-1}\mathbf{1}_m}_{\equiv\beta} + \frac{\nu_2}{2}\underbrace{\mathbf{1}_m^\top\sigma^{-1}\mathbf{1}_m}_{\equiv\gamma}\Big) \equiv w^2\left(\frac{\nu_1}{2}\beta + \frac{\nu_2}{2}\gamma\right).$$

Let $\mu_p \equiv \frac{E_p - w}{w}$. The solution for the Lagrange multipliers can be written as

$$\begin{array}{lcl} \dfrac{\nu_1 w}{2} & = & \dfrac{\mu_p\gamma - \beta}{\alpha\gamma - \beta^2} \\ \dfrac{\nu_2 w}{2} & = & \dfrac{\alpha - \beta\mu_p}{\alpha\gamma - \beta^2} \end{array}$$

Therefore, the solution for the portfolio is,

$$\frac{\hat{\pi}}{w} = \frac{\gamma\mu_p - \beta}{\alpha\gamma - \beta^2}\sigma^{-1}b + \frac{\alpha - \beta\mu_p}{\alpha\gamma - \beta^2}\sigma^{-1}\mathbf{1}_m.$$

[4] Explain why.

Finally, the value of the program is,

$$var\left[\frac{w^{+}(\hat{\pi})}{w}\right]=\frac{1}{w^{2}}\hat{\pi}^{\top}\sigma\hat{\pi}=\frac{1}{w}\hat{\pi}^{\top}\frac{\mu_{p}\gamma-\beta}{\alpha\gamma-\beta^{2}}b+\frac{1}{w}\hat{\pi}^{\top}\frac{\alpha-\mu_{p}\beta}{\alpha\gamma-\beta^{2}}\mathbf{1}_{m}=\frac{\gamma\mu_{p}^{2}-2\beta\mu_{p}+\alpha}{\alpha\gamma-\beta^{2}}.$$

The previous relation is also:

$$\frac{\gamma\mu_{p}^{2}-2\beta\mu_{p}+\alpha}{\alpha\gamma-\beta^{2}}=\frac{\gamma^{2}\mu_{p}^{2}-2\beta\gamma\mu_{p}+\alpha\gamma}{(\alpha\gamma-\beta^{2})\gamma}=\frac{\gamma^{2}\mu_{p}^{2}-2\beta\gamma\mu_{p}+\beta^{2}+(\alpha\gamma-\beta^{2})}{(\alpha\gamma-\beta^{2})\gamma}=\frac{(\gamma\mu_{p}-\beta)^{2}}{(\alpha\gamma-\beta^{2})\gamma}+\frac{1}{\gamma},$$

which is exactly (2.17).

2
The CAPM in general equilibrium

2.1 Introduction

We develop general equilibrium foundations to the CAPM. First, we review the static model of general equilibrium - without uncertainty. Second, we emphasize the role of financial assets in a world of uncertainty, and then we derive the CAPM.

2.2 Static general equilibrium in a nutshell

We consider an economy with n agents and m commodities. Let w_{ij} denote the endowment of commodity i at the disposal of the j-th agent. Let the price vector be $p = (p_1, \cdots, p_m)$, where p_i is the price of commodity i. Let $w_i = \sum_{j=1}^n w_{ij}$ be the total endowment of commodity i in the economy, $i = 1, \cdots, m$, and $W = (w_1, \cdots, w_m)$ the corresponding endowments bundle of the economy.

Agent j has utility function $u_j(c_{1j}, \cdots, c_{mj})$, where $(c_{ij})_{i=1}^m$ denotes his consumption bundle. Utility functions satisfy the following conditions:

Assumption 2.1 (Preferences).

- A1 *Monotonicity.*
- A2 *Continuity.*
- A3 *Quasi-concavity*: $u_j(x) \geq u_j(y)$, *and* $\forall \alpha \in (0,1)$, $u_j(\alpha x + (1-\alpha)y) > u_j(y)$ *or,* $\frac{\partial u_j}{\partial c_{ij}}(c_{1j}, \cdots, c_{mj}) \geq 0$ *and* $\frac{\partial^2 u_j}{\partial c_{ij}^2}(c_{1j}, \cdots, c_{mj}) \leq 0$.

Let $B_j(p_1, \cdots, p_m) = \{(c_{1j}, \cdots, c_{mj}) : \sum_{i=1}^m p_i c_{ij} \leq \sum_{i=1}^m p_i w_{ij} \equiv R_j\}$, a bounded, closed and convex (i.e., a compact) set. Every agent $j = 1, \cdots, n$ solves the following program:

$$\max_{\{c_{ij}\}} u_j(c_{1j}, \cdots, c_{mj}) \quad \text{s.t.} \quad (c_{1j}, \cdots, c_{mj}) \in B_j(p_1, \cdots, p_m) \tag{2.1}$$

Because B_j is a compact set, this problem has a solution, since by assumption (A2) u_j is continuous, and a continuous function attains its maximum on a compact set. Moreover, Appendix A, proves that this maximum is unique.

Next, we write down the $m-1$ first order conditions in (1.1) and a m-th equation, representing the constraint of the program. For $j = 1, \cdots, n$,

$$\begin{cases} \dfrac{\frac{\partial u_j}{\partial c_{1j}}}{p_1} = \dfrac{\frac{\partial u_j}{\partial c_{2j}}}{p_2} \\ \cdots \\ \dfrac{\frac{\partial u_j}{\partial c_{1j}}}{p_1} = \dfrac{\frac{\partial u_j}{\partial c_{mj}}}{p_m} \\ \sum_{i=1}^{m} p_i c_{ij} = \sum_{i=1}^{m} p_i w_{ij} \end{cases} \tag{2.2}$$

This is a system of m equations in m unknowns (c_{ij}). Solutions to this system are vectors in $\mathbb{R}_+^m$, and are denoted as $\hat{C}_{ij} = (\hat{c}_{1j}, \cdots, \hat{c}_{mj})$, where each component is a function of prices and endowments:

$$\hat{C}_{ij}(p, w_{1j}, \cdots, w_{mj}) = (\hat{c}_{1j}(p, w_{1j}, \cdots, w_{mj}), \cdots, \hat{c}_{mj}(p, w_{1j}, \cdots, w_{mj})) . \tag{2.3}$$

We call functions $\hat{c}$ demand functions.

Sometimes, it is possible to invert the previous system in a very simple way. For example, suppose that the utility function is separable in its arguments. Let $\left[\frac{\partial u_j}{\partial c_{ij}}\right]^{-1}(\cdot)$ be the inverse function of $\frac{\partial u_j}{\partial c_{ij}}$. System (1.2) can be rewritten as:

$$\begin{cases} c_{2j} = \left[\frac{\partial u_j}{\partial c_{2j}}\right]^{-1} \left(\frac{p_2}{p_1}\frac{\partial u_j}{\partial c_{1j}}\right) \\ \cdots \\ c_{mj} = \left[\frac{\partial u_j}{\partial c_{mj}}\right]^{-1} \left(\frac{p_m}{p_1}\frac{\partial u_j}{\partial c_{1j}}\right) \\ \sum_{i=1}^{m} p_i c_{ij} = \sum_{i=1}^{m} p_i w_{ij} \end{cases} \tag{2.4}$$

By replacing the first $m-1$ equations into the m-th equation, one gets

$$\sum_{i=2}^{m} p_i \left[\frac{\partial u}{\partial c_{ij}}\right]^{-1} \left(\frac{p_i}{p_1}\frac{\partial u}{\partial c_{1j}}\right) = \sum_{i=1}^{m} p_i w_{ij} - p_1 c_{1j}.$$

By replacing the solution of c_{1j} obtained via the preceding equation into the first $m-1$ equations in (1.4) one can finally find the (unique) solution of $c_{2j}, \cdots, c_{mj}$.

Consider the following definition:

$$\hat{c}_i(p) = \sum_{j=1}^{n} \hat{c}_{ij}(p), \ i = 1, \cdots, m.$$

This is the total demand of commodity i.

In the previous program, prices are exogeneously given, and agents formulate "rational" plans taking as given such prices. More precisely, an action plan is a compete description of quantities demanded in correspondence with each possible price vector: this is well described by the fact that the consumption bundles in (1.3) depend on p. In fact, the objective of these lectures is to show how to determine prices when the agents' action plans are made consistent. Here the term "consistency" essentially means that the total "rationally formed" demand for any commodity i $\hat{c}_i(p)$ can not exceed the total endowments of the economy w_i, and in fact, below we will define an equilibrium as a price vector $\bar{p} : \hat{c}_i(\bar{p}) = w_i$ all i.

In the present introductory chapter, we consider the case of an economy without production: endowments are a bonanza, and the central aspect that will be focussed on will be how the final allocation of resources is to be directed by prices. This is a perspective that is radically different form the one proposed by the classical school (Ricardo, Marx, Sraffa, ...), in which the price determination could absolutely not be dissociated from the production process of the economy. To see the difference at work, notice that here we are going to build up a theory of price determination without any need to include the production sphere of the economy although, to make the model realistic, we will consider production processes in more advanced chapters of these lectures.

2.2.1 Walras' Law and homogeneity of degree zero of the excess demand functions

Let us plug the demand functions into the (satiated) constraint of program (1.1) to obtain:

$$0 = \sum_{i=1}^{m} p_i \left(\hat{c}_{ij}(p) - w_{ij} \right), \ \forall p, \tag{2.5}$$

where the notation has been alleviated by writing $\hat{c}_{ij}(p)$ instead of $\hat{c}_{ij} \left(p, w_{1j}, \cdots, w_{mj} \right)$.

Define the total excess demand going to the i-th commodity

$$e_i(p) = \hat{c}_i(p) - w_i, \ i = 1, \cdots, m, \ \forall p,$$

and aggregate relation (1.5) across agents to obtain:

$$\text{For all } p, \quad 0 = \sum_{j=1}^{n} \sum_{i=1}^{m} p_i (\hat{c}_{ij}(p) - w_{ij}) = \sum_{i=1}^{m} p_i e_i(p),$$

or, in vector notation:

$$0 = p \cdot E(p), \ \forall p,$$

where $E(p) = (e_1(p), \cdots, e_m(p))^\top$. The previous equality is the celebrated *Walras' law.*

Next, multiply p by $\lambda \in \mathbb{R}_{++}$. Since the constraint of program (1.1) does not change, the excess demand functions will be the same as before. Therefore,

The excess demand functions are homogeneous of degree zero, or $e_i(\lambda p) = e_i(p)$, $i = 1, \cdots, m$.

Sometimes, such a property of the excess demand functions is in tight connection with the concept of *absence of monetary illusion.*

2.2.2 Competitive equilibrium

DEFINITION 2.2 (Competitive equilibrium). A *competitive equilibrium* is a vector $\bar{p}$ in $\mathbb{R}_+^m$ such that $e_i(\bar{p}) \leq 0$ for all $i = 1, \cdots, m$, with at least one component of $\bar{p}$ being strictly positive. Furthermore, if there exists a $j : e_j(\bar{p}) < 0$, then $\bar{p}_j = 0$.

2.2.3 Back to Walras' law

Walras' law holds essentially because it is derived by aggregation of the agents' constraints, which are nothing but budget identities. In particular, Walras' law holds for any price vector and a fortiori for the equilibrium price vectors:

$$0=\sum_{i=1}^{m}\bar{p}_i e_i(\bar{p})=\sum_{i=1}^{m-1}\bar{p}_i e_i(\bar{p})+\bar{p}_m e_m(\bar{p}). \tag{2.6}$$

Now suppose that the first $m-1$ markets are in equilibrium: $e_i(\bar{p})\leq 0, i=1,\cdots,m-1$. By the definition of an equilibrium, $\text{sign}(e_i(p))\,\bar{p}_i=0$, or $\bar{p}_i=[\text{sign}(e_i(\bar{p}))+1]\,\bar{p}_i$. Therefore, by eq. (1.7),

$$\bar{p}_m e_m(\bar{p})=-\sum_{i=1}^{m-1}\bar{p}_i e_i(\bar{p})=-\sum_{i=1}^{m-1}\left[\text{sign}\left(e_i(\bar{p})\right)+1\right]\bar{p}_i e_i(\bar{p})=0.$$

Hence, $\bar{p}_m e_m(\bar{p})=0$, which implies that the m-th market is in equilibrium. Since the choice of the m-th market is arbitrary, we have that:

If $m-1$ markets are in equilibrium, then the remaining market is also in equilibrium.

2.2.4 The notion of numéraire

The excess demand functions are homogeneous of degree zero, and Walras' law implies that if $m-1$ markets are at the equilibrium, then the mth market is also at the equilibrium. We wish to link such results. One implication of the Walras' law is that system (1.6) in definition 1.2 has not m independent relationships. There are only $m-1$ independent relationships: once that the first $m-1$ relations are determined, the mth is automatically determined. This means that there exist $m-1$ independent relations and m unknowns: system (1.6) is indetermined, and there exists an infinity of solutions. If the mth price is chosen as an exogeneous datum, the result is that we obtain a system of $m-1$ equations and $m-1$ unknowns. Provided a solution exists, this is a function f of the mth price, viz $\bar{p}_i=f_i(\bar{p}_m)$, $i=1,\cdots,m-1$. It is thus very natural to refer to the mth commodity as the *numéraire.* As is clear, the general equilibrium can only determine a structure of relative prices. The scale of such a structure depends on the price level of the numéraire. In addition, it is easily checked that if function f_is are homogeneous of degree one, multiplying p_m by a strictly positive number λ does not change the relative prices structure. By the equilibrium condition, for all $i=1,\cdots,m$,

$$\begin{aligned} 0 &\geq e_i\left(\bar{p}_1,\bar{p}_2,\cdots,\lambda\bar{p}_m\right)=e_i\left(f_1(\lambda\bar{p}_m),f_2(\lambda\bar{p}_m),\cdots,\lambda\bar{p}_m\right)\\ &= e_i\left(\lambda\bar{p}_1,\lambda\bar{p}_2\cdots,\lambda\bar{p}_m\right)=e_i\left(\bar{p}_1,\bar{p}_2\cdots,\bar{p}_m\right), \end{aligned}$$

where the second equality is due to the assumed homogeneity property of the f_is, and the last equality holds because the e_is are homogeneous of degree zero.

REMARK 2.3. By defining relative prices of the form $\hat{p}_j=p_j/\,p_m$, one has that $p_j=\hat{p}_j\cdot p_m$ is a function which is homogeneous of degree one. In other terms, if $\lambda\equiv\bar{p}_m^{-1}$,

$$0\geq e_i\left(\bar{p}_1,\cdots,\bar{p}_m\right)=e_i\left(\lambda\bar{p}_1,\cdots,\lambda\bar{p}_m\right)\equiv e_i\left(\frac{\bar{p}_1}{\bar{p}_m},\cdots,1\right).$$

2.2.5 Optimality

Let $c^j = (c_{1j}, \cdots, c_{mj})$ be the allocation to agent j, $j = 1, \cdots, n$.

DEFINITION 2.4 (Pareto optimum). *An allocation $\bar{c} = (\bar{c}^1, \cdots, \bar{c}^n)$ is a Pareto optimum if $\sum_{j=1}^n (\bar{c}^j - w^j) \leq 0$ and there is no $c = (c^1, \cdots, c^n)$ such that $u_j(c^j) \geq u_j(\bar{c}^j)$, $j = 1, \cdots, n$, with one strictly inequality for at least one agent.*

THEOREM 2.5 (First welfare theorem). *Every competitive equilibrium is a Pareto optimum.*

PROOF. Let us suppose on the contrary that $\bar{c}$ is an equilibrium but not a Pareto optimum. Then there exists a $c : u_j(c^j) \geq u_j(\bar{c}^j)$ with one strictly inequality for at least one j. Let j^* be such j. The preceding assert can then be restated as follows: "Then there exists a $c : u_{j^*}(c^{j^*}) > u_{j^*}(\bar{c}^{j^*})$". Because $\bar{c}^{j^*}$ is optimal for agent j^*, $c^{j^*} \notin B_j(\bar{p})$, or $\bar{p}c^{j^*} > \bar{p}w_j$ and, by aggregating: $\bar{p}\sum_{j=1}^n c^j > \bar{p}\sum_{j=1}^n w^j$, which is unfeasible. It follows that c can not be an equilibrium. ‖

Now we show that any Pareto optimum can be "decentralized". That is, corresponding to a given Pareto optimum $\bar{c}$, there exist ways of redistributing around endowments as well as a price vector $\bar{p} : \{\bar{p}\bar{c} = \bar{p}w\}$ which is an equilibrium for the initial set of resources.

THEOREM 2.6 (Second welfare theorem). *Every Pareto optimum can be decentralized.*

PROOF. Let $\bar{c}$ be a Pareto optimum and $\tilde{B}_j = \{c^j : u_j(c^j) > u_j(\bar{c}^j)\}$. Let us consider the two sets $\tilde{B} = \bigcup_{j=1}^n \tilde{B}_j$ and $A = \left\{(c^j)_{j=1}^n : c^j \geq 0 \; \forall j, \; \sum_{j=1}^n c^j = w\right\}$. A is the set of all possible combinations of feasible allocations. By the definition of a Pareto optimum, there are no elements in A that are simultaneously in $\tilde{B}$, or $A \bigcap \tilde{B} = \emptyset$. In particular, this is true for all compact subsets B of $\tilde{B}$, or $A \bigcap B = \emptyset$. Because A is closed, by the Minkowski's separating theorem (see appendix) there exists a $p \in \mathbb{R}^m$ and two distincts numbers d_1, d_2 such that

$$p'a \leq d_1 < d_2 \leq p'b, \quad \forall a \in A, \; \forall b \in B.$$

This means that for all allocations $(c^j)_{j=1}^n$ preferred to $\bar{c}$, we have:

$$p' \sum_{j=1}^n w^j < p' \sum_{j=1}^n c^j,$$

or, by replacing $\sum_{j=1}^n w^j$ with $\sum_{j=1}^n \bar{c}^j$,

$$p' \sum_{j=1}^n \bar{c}^j < p' \sum_{j=1}^n c^j. \tag{2.7}$$

Next we show that $p > 0$. Let $\bar{c}_i = \sum_{j=1}^n \bar{c}_{ij}$, $i = 1, \cdots, m$, and partition $\bar{c} = (\bar{c}_1, \cdots, \bar{c}_m)$. Let us apply inequality (1.8) to $\bar{c} \in A$ and, for $\mu > 0$, to $c = (\bar{c}_1 + \mu, \cdots, \bar{c}_m) \in B$. We have $p_1\mu > 0$, or $p_1 > 0$. by symmetry, $p_i > 0$ for all i. Finally, we choose $c^j = \bar{c}^j + \mathbf{1}_m \frac{\epsilon}{n}, j = 2, \cdots, n$, $\epsilon > 0$ in (1.8), $p'\bar{c}^1 < p'c^1 + p'\mathbf{1}_m\epsilon$ or,

$$p'\bar{c}^1 < p'c^1,$$

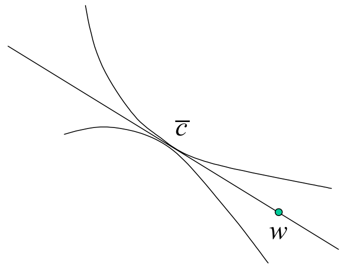


FIGURE 2.1. Decentralizing a Pareto optimum

for sufficiently small ϵ (ϵ has the form $\epsilon = \epsilon(c^1)$). This means that $u_1(c^1) > u_1(\bar{c}^1) \Rightarrow p'c^1 > p'\bar{c}^1$. This means that $\bar{c}^1 = \arg\max_{c^1} u_1(c^1)$ s.t. $p'c^1 = p'\bar{c}^1$. By symmetry, $\bar{c}^j = \arg\max_{c^j} u_j(c^j)$ s.t. $p'c^j = p'\bar{c}^j$ for all j. ‖

The previous theorem can be interpreted in terms of a *transfer payments equilibrium*. For any given Pareto optimum $\bar{c}^j$, a social planner can always give $\bar{p}w^j$ to each agent (with $\bar{p}\bar{c}^j = \bar{p}w^j$, where w^j is chosen by the planner), and agents choose $\bar{c}^j$. Figure 1.1 illustratres such a decentralization procedure within the celebrated Edgeworth's box. Suppose that the objective is to achieve $\bar{c}$. Given an initial allocation w chosen by the planner, each agent is given $\bar{p}w^j$. Under *laissez faire*, $\bar{c}$ will always be obtained. In other terms, agents are given a constraint of the form $pc^j = \bar{p}w^j$ and when w^j and $\bar{p}$ are chosen so that each agent is induced to choose $\bar{c}^j$, a supporting equilibrium price is $\bar{p}$. In this case, the marginal rates of substitutions are identical as established in the following celebrated result:

THEOREM 2.7 (Characterization of Pareto optima). *A feasible allocation $\bar{c} = (\bar{c}^1, \cdots, \bar{c}^n)$ is a Pareto optimum if and only if there exists*

$$\tilde{\phi} \in \mathbb{R}^{m-1}_{++} : \tilde{\bigtriangledown} u_j = \tilde{\phi},\ j = 1, \cdots, n, \quad \tilde{\bigtriangledown} u_j \equiv \left(\frac{\frac{\partial u_j}{\partial c_{2j}}}{\frac{\partial u_j}{\partial c_{1j}}}, \cdots, \frac{\frac{\partial u_j}{\partial c_{mj}}}{\frac{\partial u_j}{\partial c_{1j}}} \right) \tag{2.8}$$

PROOF. A Pareto optimum satisfies

$$\bar{c} \in \arg \max_{c \in \mathbb{R}^{m \cdot n}_{+}} u_1\left(c^1\right)$$
$$\text{s.t.} \begin{cases} u_j(c^j) \geq \bar{u}_j,\ j = 2, \cdots, n & (\lambda_j,\ j = 2, \cdots, n) \\ \sum_{j=1}^{n} (c^j - w^j) \leq 0 & (\phi_i,\ i = 1, \cdots, m) \end{cases}$$

The Lagrangian function associated with this program is

$$L = u_1(c^1) + \sum_{j=2}^{n} \lambda_j \left(u_j(c^j) - \bar{u}_j\right) - \sum_{i=1}^{m} \phi_i \sum_{j=1}^{n} \left(c_{ij} - w_{ij}\right),$$

and the first order conditions are

$$\begin{cases} \dfrac{\partial u_1}{\partial c_{11}} & = & \phi_1 \\ & \cdots & \\ \dfrac{\partial u_1}{\partial c_{m1}} & = & \phi_m \end{cases}$$

and, for $j = 2, \cdots, n$,

$$\begin{cases} \lambda_j \dfrac{\partial u_j}{\partial c_{1j}} & = & \phi_1 \\ & \cdots & \\ \lambda_j \dfrac{\partial u_j}{\partial c_{mj}} & = & \phi_m \end{cases}$$

Divide each system by its first equation. We obtain exactly (1.9) with $\tilde{\phi} = \left(\frac{\phi_2}{\phi_1}, \cdots, \frac{\phi_m}{\phi_1}\right)$.

The converse is straight forward. ‖

2.3 Time and uncertainty in general equilibrium

"A commodity is characterized by its physical properties, the date and the place at which it will be available."

Gerard Debreu (1959, chapter 2)

General equilibrium theory can be used to study a variety of fields by the use of the previous definition - from the theory of international commerce to finance. Unfortunately, this definition is of no use to deal with situations in which future events are uncertain. Debreu (1959, chapter 7) extended the previous definition to the uncertainty case. In the uncertainty case, the a commodity can be described through a list of physical properties, but the structure of dates and places is replaced by some event structure. An example stressing the difference between two kinds of contracts on the delivery of corn arising under conditions of certainty (case A) and uncertainty (case B) is given below:

A *The first agent will deliver 5000 tons of corn of a specified type to the second agent, who will accept the delivery at date t and in place ℓ.*

B *The first agent will deliver 5000 tons of corn of a specified type to the second agent, who will accept the delivery in place ℓ and in the event s_t at time t. If s_t does not occur at time t, no delivery will take place.*

In case B, the payment of the contract is made at the time of the contract, even if it is possible that in the future the buyer of the corn will not receive the corn in some states.

The static model of the previous chapter can now be used to model contracts of the kind of case B above. As an example, consider a two-period economy. Uncertainty affects the second period only, in which $s_n < \infty$ mutually exhaustive and exclusive states of nature may occur. We can now recover the model of the previous chapter once that we replace m with m^*, where $m^* = s_n \cdot m$, and m denotes as usually the number of commodities described by physical properties, dates and places. Therefore, an equilibrium of this economy is defined similarly as the equilibrium of the economy of the previous chapter. The only difference is that the dimension of the commodity space in the uncertainty case is higher. Since the conditions given in the previous chapter did not depend on the dimension of the commodity space (unless this is not infinite, an issue that is not treated in the present chapter), we can safely say that an equilibrium exists in an economy under uncertainty of this kind under the same conditions of the previous chapter. The extension to multiperiod economies is immediate. Next chapter makes extensions to infinite horizon economies.

The merit of such a construction is that it is very simple. Such a merit is at the same time the main inconvenient of the model. Indeed, the important assumption underlying such a construction is that there exists markets in which all commodities for all states of nature are exchanged. Such markets are usually referred to as "contingent". In addition, such contingent markets are also *complete*: a market is open in correspondence with every commodity in all states of nature. Therefore, agents can implement any feasible action plan, and the resources allocation is Pareto-optimal. In addition such a construction presumes the existence of $s_n \cdot m$ contingent markets. This is a strong assumption that we may wish to relax by introducing financial assets. Chapter 6 deals with issues concerning the structure of incomplete markets.

2.4 The role of financial assets

What is the role of financial assets in an uncertainty world? Arrow (1953) proposed the following interpretation. Instead of signing good-exchange contracts that are conditioned on the realization of certain events, agents might prefer to sign contracts giving rise to payoffs that are contingent on the realisation of the same events. In a second step, the various payoffs could then be used to satisfy the consumption needs related to the realization of the various events. Therefore, a financial asset is simply a contract whose payoff is a given amount of numéraire in the state of nature s if such a state will prevail in the future, and nil otherwise. In the remainder, we shall qualify such a kind of assets as elementary Arrow-Debreu assets. In more general terms, a financial asset is then a function $\mathcal{A} : \mathcal{S} \mapsto \mathbb{R}$, where $\mathcal{S}$ is a subset of all the future states of the world.

Let x be the number of existing financial assets. We can now link financial assets to commodities: it suffices to say that if state of nature s occurs tomorrow, the payment $\mathcal{A}_i(s)$ deriving from assets $\mathcal{A}_i$, could be used to finance a net transaction on the various commodity markets:

$$p(s) \cdot E(s) = \sum_{i=1}^{x} \theta^{(i)} \mathcal{A}_i(s), \ \forall s \in \mathcal{S},$$

where $p(s)$ and $E(s)$ denote the m-dimensional vectors of prices and excess demands referring to the m commodities contingent on the realization of state s for a given agent, and $\theta^{(i)}$ is the quantity of assets i hold by the same agent. More precise details will of course be given in the remainder of this chapter.

Clearly, the previous relation does not hold in general. A condition is that the number of financial assets be sufficiently high to let each agent face the heterogeneity of the states of nature. We see that completeness reduces to a simple size problem concerning exclusively the number of financial assets. Indeed, the proof that markets are complete insofar as $x = s_n$ is a simple linear algebra exercise developed in the following sections.

The important lesson that must be understood since the beginning of this chapter is that *in the models that we examine in these lectures*, the only role of financial assets is to transfer value from a state of nature to another in such a way that the resulting wealth be added to the state contingent endowments to finance state-contingent consumption. Notice finally that in so doing, we are reducing the dimension of the problem, since we will be considering equilibrium conditions in $s_n + m$ markets, instead of equilibrium conditions in $s_n \cdot m$ markets.

2.5 Arbitrage and optimality

The principle of absence of arbitrage opportunities is heuristically introduced in the next subsection, and embedded in a specific equilibrium model in subsection 2.3.2.

2.5.1 How to price a financial asset?

Consider an economy in which uncertainty is resolved by the realization of the event: "tomorrow it will be raining". M. X must implement an action plan conditioned on this event: if tomorrow will be sunny, M. X will need $c_s > 0$ units of money to buy sun-glasses; and if it will rain tomorrow, M. X will need $c_r > 0$ units of money to buy an umbrella. M. X has access to a financial market where m assets are exchanged, and he builds up a portfolio θ with which he tries to imitate the structure of payments that he needs tomorrow:

$$\begin{cases} \sum_{i=1}^{m} \theta_i q_i (1 + x_r^{(i)}) = c_r \\ \sum_{i=1}^{m} \theta_i q_i (1 + x_s^{(i)}) = c_s \end{cases} \tag{2.9}$$

where q_i is the price of financial asset i, and $x_r^{(i)}, x_s^{(i)}$ are the net returns of asset i in the two states of nature. M. X knows the values of $x_r^{(i)}$ and $x_s^{(i)}$. For the time being, no assumption is made as regards the resources needed to buy the assets. Section 2.3.2 presents a version of this problem that is articulated within a standard microeconomic framework; see, also, remark 2.1. Finally, note that no assumption is being made here as regards the preferences of M. X.

System (2.1) has 2 equations and m unknowns. If $m < 2$, there is no perfect hedging strategy (i.e. exact obtention of the desired pair $(c_i)_{i=r,s}$). In this case we say that markets are *incomplete*. The same phenomenon propagates to the case in which we have s_n states of nature. In this case, a necessary condition for market completeness is the existence of at least s_n assets. Indeed, the system to be solved is:

$$c = X \cdot \theta,$$

where

$$X = \begin{pmatrix} q_1(1 + x_{s_1}^{(1)}) & & q_m(1 + x_{s_1}^{(m)}) \\ & \ddots & \\ q_1(1 + x_{s_n}^{(1)}) & & q_m(1 + x_{s_n}^{(m)}) \end{pmatrix},$$

and $\theta \in \mathbb{R}^m$, $H \in \mathbb{R}^{s_n}$, and there is well defined solution if $\text{rank}(X) = s_n = m$:

$$\hat{\theta} = X^{-1} c.$$

In the previous case, $s_n = 2$; then assume that $m = 2$. We have:

$$\begin{cases} \hat{\theta}_1 = \dfrac{(1 + x_r^{(2)}) c_s - (1 + x_s^{(2)}) c_r}{q_1 \left[(1 + x_s^{(1)})(1 + x_r^{(2)}) - (1 + x_r^{(1)})(1 + x_s^{(2)}) \right]} \\ \hat{\theta}_2 = \dfrac{(1 + x_s^{(1)}) c_r - (1 + x_r^{(1)}) c_s}{q_2 \left[(1 + x_s^{(1)})(1 + x_r^{(2)}) - (1 + x_r^{(1)})(1 + x_s^{(2)}) \right]} \end{cases}$$

Finally, assume that the second asset is safe, or that it yields the same return in the two states of nature: $x_r^{(2)} = x_s^{(2)} \equiv r$. Let $x_s = x_s^{(1)}$ and $x_r = x_r^{(1)}$. The pair $(\hat{\theta}_1, \hat{\theta}_2)$ can then be rewritten as:

$$\begin{cases} \hat{\theta}_1 = \dfrac{c_s - c_r}{q_1 (x_s - x_r)} \\ \hat{\theta}_2 = \dfrac{(1 + x_s) c_r - (1 + x_r) c_s}{q_2 (1 + r) (x_s - x_r)} \end{cases}$$

As is clear, the problem solved here corresponds to an issue concerning the *replication* of a random variable. Here a random variable $(c_i)_{i=r,s}$ (where c_r and c_p are known!) has been replicated by a portfolio for hedging purposes. In this 2-states example, 2 independent assets were able to generate any 2-states variable. Now the natural further step is to understand what happens when we assume that there exists a third asset $\mathcal{A}$ promising exactly the same random variable $(c_i)_{i=r,s}$ that can also be generated with portfolio $\hat{\theta}$.

The answer is that if the current price of this asset is H, then it must be the case that:

$$H = V \equiv \hat{\theta}_1 q_1 + \hat{\theta}_2 q_2 \tag{2.10}$$

to avoid *arbitrage opportunities*. Indeed, if $V < H$, then you can buy $\hat{\theta}$ and sell at the same time $\mathcal{A}$. In so doing, you would get a profit $H - V$, with certainty. Indeed, $\hat{\theta}$ generates c_r if tomorrow it will rain and c_r if tomorrow it will not rain. In any case, $\hat{\theta}$ generates the payments that are exactly necessary to cover the contractual committments derived by the selling of $\mathcal{A}$. By a symmetric argument, $V > H$ is also impossible. Whence (2.2).

It remains to compute the r.h.s. of (2.2), thus obtaining the evaluation formula of $\mathcal{A}$:

$$H = \frac{1}{1+r} \left[P_1^* c_s + (1 - P_1^*) c_r \right], \; P_1^* = \frac{x_r^{(1)} - r}{x_s^{(1)} - x_r^{(1)}}.$$

H can be seen at the expectation of payments promised by $\mathcal{A}$, taken under a probability P^*, that are discounted at the certainty factor $1 + r$.

Remark 2.1. Here we were looking for an argument that could be used to evaluate $\mathcal{A}$ without making reference to agents' preferences. This does not mean that such agents will implement the strategy $\hat{\theta}$ to obtain the payoffs of $\mathcal{A}$: it may turn out, for instance, that the agents' budget constraints could not even allow them to implement such a portfolio. By contrast, such a strategy would be possible and in fact, instrumental to the obtention of arbitrage profits when relation (2.2) is not satisfied. In this case, any agent without resources can implement such an operation !

Now suppose that there is another contraction day that is submitted to the same structure of uncertainty resolution: at the following date, $\mathcal{A}$ gives c_{ss} if it will be sunny given that the previous day was sunny, c_{rs} if it will be sunny given that the previous day it was raining, ... By repeating the same previous arguments we obtain:

$$H = \frac{1}{(1+r)^2} \left[(P_1^*)^2 c_{ss} + P_1^*(1 - P_1^*) c_{sr} + (1 - P_1^*) P_1^* c_{rs} + (P_1^*)^2 c_{rr} \right],$$

and by considering T days,

$$H = \frac{1}{(1+r)^T} E_t (c_T). \tag{2.11}$$

As is clear, the previous approach can be used only when one adopts the assumption of complete markets. In addition, the notion of complete markets must be made more precise here. It is true that 2^T states of nature may occur from here to T days, but we simply do not need 2^T assets to duplicate $\mathcal{A}$, since we can trade T days. The dimension of the problem can be naturally reduced by intermediate trading, with dynamic strategies that rebalance every day the portfolio duplicating the trajectory of the payoffs promised by $\mathcal{A}$.

The previous arguments reveal in which sense we must interpret the commonplace that *in the presence of complete markets*, there are no links between risk-aversion of M. X and absence of arbitrage opportunities: indeed, during the derivation of (2.3), we did not make any assumptions on the utility function of M. X. Chapter 4 deepens these issues in the framwork of continuous time models.

2.5.2 Absence of arbitrage opportunities and Arrow-Debreu economies

In this subsection, we provide links between the absence of arbitrage opportunities and the possibility to build up economies displaying the same properties of the Arrow-Debreu economies of the previous chapter. We will see that the essential connection will be made with the help of the complete markets assumption. Even in the presence of incomplete markets, however, we can lay out the foundations for a theory that is described in chapter 6.

Here we consider a multistate economy but for simplicity we consider only one commodity. The extension to many commodities is in the last section of the present chapter. The case of an infinite state space is (very partially) covered in chapter 4.

Let $v_i(\omega_s)$ be the payoff of asset i in the state ω_s, $i = 1, \cdots, m$ and $s = 1, \cdots, d$. Consider the payoff matrix:

$$V \equiv \begin{pmatrix} v_1(\omega_1) & & v_m(\omega_1) \\ & \ddots & \\ v_1(\omega_d) & & v_m(\omega_d) \end{pmatrix}.$$

Let $v_{si} \equiv v_i(\omega_s)$, $v_{s,\cdot} \equiv (v_{s1}, v_{s2}, \cdots, v_{sm})$, $v_{\cdot,i} \equiv (v_{1i}, v_{2i}, \cdots, v_{di})'$. We suppose that rank$(V) = m \leq d$.

The budget constraint of each agent has the form:

$$\begin{cases} c_0 - w_0 = -q\theta = -\sum_{i=1}^{m} q_i\theta_i \\ c_s - w_s = v_{s\cdot}\theta = \sum_{i=1}^{m} v_{si}\theta_i, \;\; s = 1, \cdots, d \end{cases}$$

Let $x^1 = (x_1, \cdots, x_d)'$. The second constraint can be written as:

$$c^1 - w^1 = V\theta.$$

Inexistence of the Land of Cockaigne.

An arbitrage opportunity is a portfolio that has a negative value at the first period, and a positive value in at least one state in the second period, or a positive value in all states and a nonpositive value in the first period.

Notation: $\forall x \in \mathbb{R}^m$, $x > 0$ means that at least one component of x is strictly positive while the other components of x are nonnegative. $x >> 0$ means that all components of x are strictly positive. [**Insert here further notes**]

Definition 2.10. An *arbitrage opportunity* is a strategy θ which yields[1] either $V\theta \geq 0$ with an initial investment $q\theta < 0$, or a strategy θ which yields[2] $V\theta > 0$ with an initial investment $q\theta \leq 0$.

An arbitrage opportunity can not exist in a competitive equilibrium, because the agents' program would not be well defined in this case (cf. theorem 2.4 below).

Introduce the matrix $(d+1) \times m$

$$W = \begin{pmatrix} -q \\ V \end{pmatrix},$$

the vector subspace of $\mathbb{R}^{d+1}$

$$\langle W \rangle = \left\{ z \in \mathbb{R}^{d+1} : z = W\theta, \ \theta \in \mathbb{R}^m \right\},$$

and the null space of $\langle W \rangle$

$$\langle W \rangle^{\perp} = \left\{ x \in \mathbb{R}^{d+1} : xW = \mathbf{0}_m \right\}.$$

The economic interpretation of $\langle W \rangle$ is that of the excess demand space (in all states of nature) generated by "income transfers", i.e. by transfers generated by asset investments.

Naturally, $\langle W \rangle^{\perp}$ and $\langle W \rangle$ are orthogonal in the sense that $\langle W \rangle^{\perp}$ can also be rewritten as $\langle W \rangle^{\perp} = \left\{ x \in \mathbb{R}^{d+1} : xz = \mathbf{0}_m, \ z \in \langle W \rangle \right\}$.

The assumption that there are no arbitrage opportunities is equivalent to the following condition:

$$\langle W \rangle \bigcap \mathbb{R}^{d+1}_{+} = \{0\}. \tag{2.12}$$

The interpertation of condition (2.4) is that in the absence of arbitrage opportunities, $\nexists \theta : W\theta > 0$, i.e. $\nexists$ portfolios generating income transfers that are nonnegative and strictly positive in at least one state. Hence, $\langle W \rangle$ and the positive orthant can only have 0 as a common point (that can be obtained e.g. with $\theta = 0$).

The following result gives a first characterization of the financial asset "forced" prices.

Theorem 2.11. *There are no arbitrage opportunities if and only if there exists a* $\phi \in \mathbb{R}^d_{++}$: $q = \phi' V$. *If* $m = d$, ϕ *is unique, and if* $m < d$, $\dim \left(\phi \in \mathbb{R}^d_{++} : q = \phi' V \right) = d - m$.

Proof. In the appendix.

By pre-multiplying the second constraint by $\phi^{\top}$,[3]

$$\phi^{\top}(c^1 - w^1) = \phi^{\top} V\theta = q\theta = -(c_0 - w_0),$$

where the second equality is due to theorem 2.2, and the third equality is due to the first period constraint. Hence, we have shown that in the absence of arbitrage opportunities, each agent

[1] $V\theta \geq 0$ means that $[V\theta]_j \geq 0$, $j = 1, \cdots, d$, i.e. it allows for $[V\theta]_j = 0$, $j = 1, \cdots, d$.

[2] $V\theta > 0$ means $[V\theta]_j \geq 0$, $j = 1, \cdots, d$, with at least one j for which $[V\theta]_j > 0$.

[3] When markets are complete, we could also pre-multiply the second period constraint by V^{-1}: $V^{-1}(c^1 - w^1) = \theta$, replace this into the first period constraint, and obtain: $0 = c_0 - w_0 + qV^{-1}(c^1 - w^1)$, which is another way to state the conditions of the absence of arbitrage opportunities.

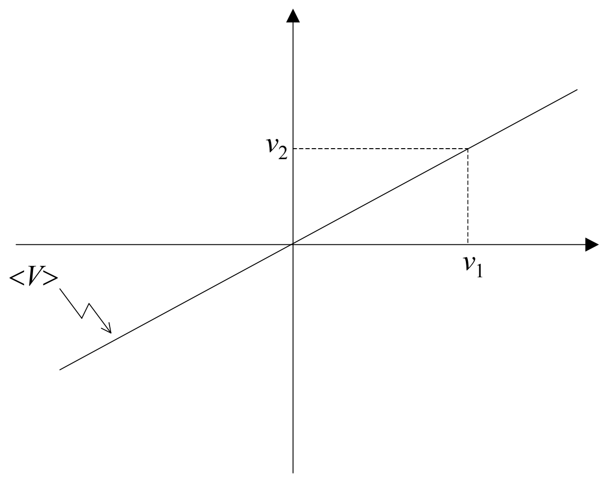


FIGURE 2.2. Incomplete markets, $d = 2$, $m = 1$.

has access to the following budget constraint:

$$0 = c_0 - w_0 + \phi^\top \left(c^1 - w^1\right) = c_0 - w_0 + \sum_{s=1}^{d} \phi_s \left(c_s - w_s\right), \text{ with } \left(c^1 - w^1\right) \in \langle V \rangle, \qquad (2.13)$$

The previous relation justifies why ϕ is commonly referred to as the vector containing the d *state prices*. Indeed, the initial two-period model was reconducted to a static Arrow-Debreu type model, where the price of a commodity in state s is ϕ_s. Here, such a ϕ_s can be re-intepreted as $\phi_s = \frac{p_s}{p_0}$. Such a property propagates to the multiperiod models, and even to the infinite horizon models under some regularity conditions spelled out in chapter 4. Furthermore, theorem 2.2 states that there exists one such ϕs when markets are complete.

It is instructive to deepen the interpretation of (2.5). First, $\langle V \rangle$ represents the subspace of excess demands to which agent have access in the second period, and this subspace is generated by the payoffs obtained by the portfolio choices at the first period:

$$\langle V \rangle = \left\{ e \in \mathbb{R}^d : e = V\theta,\ \theta \in \mathbb{R}^m \right\}.$$

As an example, consider $d = 2$ and $m = 1$. In this case, $\langle V \rangle = \{e \in \mathbb{R}^2 : e = V\theta,\ \theta \in \mathbb{R}\}$, with $V = V_1$, where $V_1 = \binom{v_1}{v_2}$ say, and $\dim \langle V \rangle = 1$. Figure 2.1 illustrates the situation.

Now suppose to open a new market for a second financial asset $V_2 = \binom{v_3}{v_4}$, so that $m = 2$, $V = \begin{pmatrix} v_1 & v_3 \\ v_2 & v_4 \end{pmatrix}$. In this case, $\langle V \rangle = \left\{ e \in \mathbb{R}^2 : e = \binom{\theta_1 v_1 + \theta_2 v_3}{\theta_1 v_2 + \theta_2 v_4}, \theta \in \mathbb{R}^2 \right\}$, and $\langle V \rangle = \mathbb{R}^2$. Now we can generate any excess demand vector in $\mathbb{R}^2$, and we go back to the case considered in the previous chapter: it suffices to multiply vector V_1 by θ_1 and vector V_2 by θ_2. As an example, if one wishes to generate vector V_4 in figure 2.2, θ_1 and θ_2 must be chosen in such a way that $\theta_1 > 1$ and $\theta_2 < 1$ (the exact values of θ_1 and θ_2 are obtained by solving a linear system). In figure 2.2, vector V_3 is obtained by taking $\theta_1 = \theta_2 = 1$. Generally, when markets are incomplete, $\langle V \rangle$ is only a subspace of $\mathbb{R}^d$, and agents have access to a choice space that is more constraint than the choice space of the complete markets case in which $\langle V \rangle = \mathbb{R}^d$.

The point of the representation (2.5) of the budget constraint is that it is relatively easy on a technical standpoint to produce existence results in the case of incomplete markets with the budget written in that way. Formulating problems with only one constraint turns out to be useful also in the case of complete markets. When we propagate the unifying budget property to models

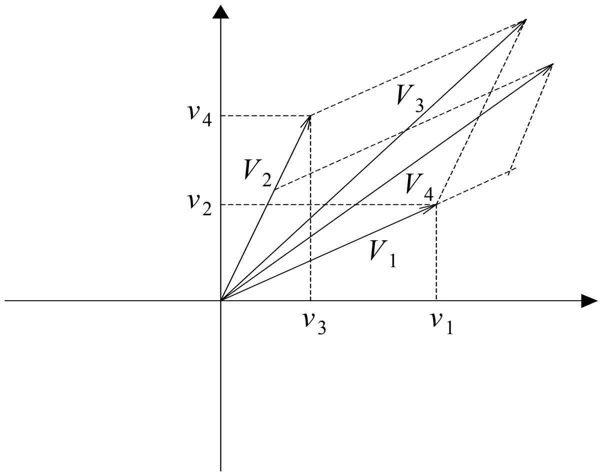


FIGURE 2.3. Complete markets, $\langle V \rangle = \mathbb{R}^2$.

with an infinity of commodities (as in many models arising in macroeconomics and finance), we can also study in a more elegant way optimality and related issues. In the case of complete markets, the next section also explains how to cast such a constraint in an expectation format under a certain probability measure. The incomplete markets case is more delicate because it implies the existence of an infinity of state prices (see chapter 5): in mathematical finance, some authors made reference to the so-called "minimax" equivalent martingale probability measure. One can implement fine duality results, but the infinite state spece has not entirely be solved (see chapter 6 for further details).

Our objective now is to show a very fundamental result, that we call here a result on the "viability of the model". Let:

$$\begin{array}{l} \left(\hat{c}_{0j}, \hat{c}_j^1\right) \in \arg\max_{c_{0j}, c_j^1} \left\{ u_j\left(c_{0j}\right) + \beta_j E\left[\nu_j(c_j^1)\right]\right\}, \quad c_j^1 \equiv \left(c_{sj}\right)_{s=1}^d \\ \text{s.t.} \left\{ \begin{array}{rcl} c_{0j} - w_{0j} & = & -q\theta_j \\ c_j^1 - w_j^1 & = & V\theta_j \end{array} \right. \end{array} \tag{2.14}$$

THEOREM 2.12. *Program* (2.6) *has a solution if and only if there are no arbitrage opportunities.*

PROOF. Let us suppose on the contrary that the previous program has a solution $\hat{c}_{0j}, \hat{c}_j^1, \hat{\theta}_j$ but that there exists a $\theta : W\theta > 0$. The program constraint can also be written as: $\hat{c}_j = w_j + W\hat{\theta}_j$, where the notation should be straight forward. We may define a strategy $\theta_j = \hat{\theta}_j + \theta$ such that $c_j = w_j + W(\hat{\theta}_j + \theta) = \hat{c}_j + W\theta > \hat{c}_j$, which contradicts the optimality of $\hat{c}_j$.

For the converse (absence of arbitrage opportunities $\Rightarrow \exists$ solution of the program), the absence of arbitrage opportunities implies that $\exists \phi \in \mathbb{R}_{++}^d : q = \phi' V$, whence the unified budget constraint (2.5) for a given ϕ. Such an unified constraint is clearly a closed subset of a compact set of the form B of the previous chapter (in fact, it is B restricted to $\langle V \rangle$), and thus it is a compact set, and every continuous function attains its maximum on a compact set. $\|$

DEFINITION 2.13. An *equilibrium* is allocations and prices $\{(\hat{c}_{0j})_{j=1}^{n}, ((\hat{c}_{sj})_{j=1}^{n})_{s=1}^{d}, (\hat{q}_i)_{i=1}^{m} \in \mathbb{R}_{+}^{n} \times \mathbb{R}_{+}^{nd} \times \mathbb{R}_{+}^{d}\}$, where allocations are solutions of program (2.6) and satisfy:

$$0=\sum_{j=1}^{n}\left(\hat{c}_{0j}-w_{0j}\right); \qquad 0=\sum_{j=1}^{n}\left(\hat{c}_{sj}-w_{sj}\right) \ (s=1,\cdots,d); \qquad 0=\sum_{j=1}^{n}\theta_{ij} \quad (i=1,\cdots,d).$$

2.6 Equivalent martingale measures and equilibrium

We are going to build up demand functions expressed in terms of the stochastic discount factor and then find the equilibrium, if any, by looking for the stochastic discount factor that clears the commodity markets. This also implies the equilibrium on the financial market by virtue of a Walras' like law. Formally, the aggregation of the constraints of the second period leads, in equilibrium, to:

$$\sum_{j=1}^{n}(c_j^1(m)-w_j^1)=V\sum_{j=1}^{n}\theta_{ij}(m)$$

By simplicity, we suppose that preferences are of the form $u_j(c_{0j})+\beta_j E[\nu_j(c_j^1)]$, where $x_j^1=(x_{sj})_{s=1}^{d}$, and require further that $u_j'(x)>0$, $u_j''(x)<0$ $\forall x>0$ and $\lim_{x\to 0}u_j'(x)=\infty$, $\lim_{x\to\infty}u_j'(x)=0$.

2.6.1 The rational expectations assumption

Lucas, Radner, Green. Every agent *correctly* anticipates the equilibrium price in each state of nature.

[Consider for example the models with asymmetric information that we will see later in these lectures. At some point we will have to compute, $E\left(\tilde{v}\,|\,p\left(\tilde{y}\right)=p\right)$. That is, the equilibrium is a pricing function which takes some values p depending on the state of nature. In this kind of models, $\lambda\theta_I\left(p\left(\tilde{y}\right),\tilde{y}\right)+\left(1-\lambda\right)\theta_U\left(p\left(\tilde{y}\right),\tilde{y}\right)+\tilde{y}=0$, and we look for a solution $p\left(\tilde{y}\right)$ satisfying this equation.]

2.6.2 Stochastic discount factors

Theorem 2.2 states that in the absence of arbitrage opportunities,

$$q_i=\phi^{\top}v_{\cdot,i}=\sum_{s=1}^{d}\phi_s v_{s,i},\ i=1,\cdots,m. \tag{2.15}$$

As in the previous section, assume that the first asset is safe, or $v_{s,1}=1$ $\forall s$. We have:

$$q_1\equiv\frac{1}{1+r}=\sum_{s=1}^{d}\phi_s. \tag{2.16}$$

Therefore, we a have a second way to interpret the state prices: here, the states of nature are exhaustive and mutually exclusive, and ϕ_s can thus be interpreted as the price to be paid today for the (sure) obtention of one unit of numéraire tomorrow in state s. Indeed, the previous relation shows that by buying in $t=0$ all these rights, one in fact is buying a pure discount bond.

Relation (2.8) suggests to introduce the following object:

$$P_s^* = (1+r)\phi_s,$$

which verifies, by construction,

$$\sum_{s=1}^{d} P_s^* = 1.$$

Therefore, we can intrepret $P^* \equiv (P_s^*)_{s=1}^d$ as a probability distribution. By replacing P^* in (2.7) we get:

$$q_i = \frac{1}{1+r}\sum_{s=1}^{d} P_s^* v_{s,i} = \frac{1}{1+r}E^{P^*}(v_{\cdot,i}),\ i = 1, \cdots, m. \tag{2.17}$$

Such a result will be linked to the so-called "martingale property" of $\left(\frac{q}{q_1}\right)_i$ under probability measure P^* in a dynamic context. Such a property justifies the usual reference to measure P^* as the *risk-neutral* probability measure. Similarly, by replacing P^* in (2.5) we get:

$$0 = c_0 - w_0 + \sum_{s=1}^{d}\phi_s(c_s - w_s) = c_0 - w_0 + \frac{1}{1+r}\sum_{s=1}^{d} P_s^*(c_s - w_s) = c_0 - w_0 + \frac{1}{1+r}E^{P^*}\left(c^1 - w^1\right).$$

It is useful to rewrite the constraint by only making use of the objective probability P: by so doing, we will easily associate the resulting constraint with programs in which one maximizes expected utility functions under P in multiperiod models + Debreu's evaluation equilibria and Pareto optima with an infinite-dimensional commodity space.

In this respect, we introduce η by the relation:

$$P_s^* = \eta_s P_s,\ s = 1, \cdots, d. \tag{2.18}$$

What is the meaning of η? P^* and P are *equivalent* measures, which means that they attribute the same weight to the null (negligeable) sets. By the arguments presented above, there are no arbitrage opportunities if and only if there exists a set of *Arrow-Debreu state prices* ϕ, or if and only if there exists P^*, but there is only one η in correspondence with a given P^*, whence: η is unique if and only if $m = d$.

By using (2.10) we get:

$$\frac{1}{1+r}E^{P^*}\left(c^1 - w^1\right) = \sum_{s=1}^{d}\frac{1}{1+r}P_s^*(c_s - w_s) = \sum_{s=1}^{d}\frac{1}{1+r}\eta_s P_s(c_s - w_s) = E\left[\frac{1}{1+r}\eta\left(c^1 - w^1\right)\right].$$

We call

$$m_s \equiv (1+r)^{-1}\eta_s$$

the *stochastic discount factor.*

By using the definition of the stochastic discount factor, we can rewrite the budget constraint as:

$$0 = c_0 - w_0 + E\left[m \cdot \left(c^1 - w^1\right)\right].$$

Similarly, by replacing m in (2.9),

$$q_i = \frac{1}{1+r}E^{P^*}(v_{\cdot,i}) = E(m \cdot v_{\cdot,i}),\ i = 1, \cdots, m. \tag{2.19}$$

Finally, note that the previous ways of writing the budget constraint must not hide the fact that the unknown of the problem is always ϕ:

$$m_s = (1+r)^{-1}\eta_s = (1+r)^{-1}\frac{P_s^*}{P_s} = \frac{\phi_s}{P_s},$$

which can not be determined without the previous solution of a general equilibrium model.

By determining m, we determine at the same time prices and equilibrium allocations. This is explained in the next subsection.

2.6.3 Equilibrium and optimality

In the absence of arbitrage opportunities, the program is

$$\max_{(c_0,c^1)} \left\{u_j(c_{0j}) + \beta_j \cdot E[\nu_j(c_j^1)]\right\} \text{ s.t. } 0 = c_{0j} - w_{0j} + E\left[m\cdot(c_j^1 - w_j^1)\right].$$

First order conditions are,

$$\begin{array}{rcl} u_j'(\hat{c}_{0j}) & = & \lambda_j \\ \beta_j \nu_j'(\hat{c}_{sj}) & = & \lambda_j m_s,\ s = 1,\cdots,d. \end{array}$$

where λ_j is a Lagrange multiplier. In an economy with only one agent, these conditions allow one to identify immediately the kernel of the model:

$$m_s = \beta\frac{\nu'(w_s)}{u'(w_0)}.$$

The economic interpretation is very simple. Notice that in the autarchic state,

$$-\left.\frac{dc_0}{dc_s}\right|_{c_0=w_0,c_s=w_s} = \beta\frac{\nu'(w_s)}{u'(w_0)}P_s = m_s P_s = \phi_s$$

represents the consumption to which the agents is disposed to give up at $t=0$ in order to receive additional consumption at $t=1$ in state s. This is the kernel for which the agent is happy not to buy or to sell financial assets. In this case, we would have:

$$P_s^* = \eta_s P_s = (1+r)\, m_s P_s = (1+r)\,\beta\frac{\nu'(w_s)}{u'(w_0)}P_s,$$

and it is easily checked (using the first order conditions, and the pure discount bond evaluation formula: $\frac{1}{1+r} = E(m)$) that $1 = (1+r)E(m) = (1+r)\sum_s m_s P_s = \sum_s P_s^*$. In other terms,

$$\sum_{s=1}^d P_s^* = (1+r)\,\beta\sum_{s=1}^d \frac{\nu'(w_s)}{u'(w_0)}P_s = 1.$$

There is another property associated with the change of measure $\frac{P_s^*}{P_s}$, namely,

$$\frac{P_s^*}{P_s} = m_s(1+r) = m_s\left\{\beta E\left[\frac{\nu'(w_s)}{u'(w_0)}\right]\right\}^{-1} = m_s\beta^{-1}\frac{u'(w_0)}{E\left[\nu'(w_s)\right]} = \frac{\nu'(w_s)}{E\left[\nu'(w_s)\right]},$$

where the second equality follows by the pure discount bond evaluation formula: $\frac{1}{1+r} = E(m)$.

In the multi-agent case, the situation is similar as soon as markets are complete. Indeed, consider the first order conditions of every agent,

$$\beta_j \frac{\nu_j'(\hat{c}_{sj})}{u_j'(\hat{c}_{0j})} = m_s, \quad s = 1, \cdots, d, \quad j = 1, \cdots, n.$$

The previous relationship clearly reveals that *as soon as markets are complete*, individuals *must* have the same marginal rate of substitution in equilibrium. This is so because ϕ is unique as soon as markets are complete (theorem 2.?), from which it follows the unicity of $m_s = \frac{\phi_s}{P_s}$ and hence the independence of $\beta_j \frac{\nu_j'(\hat{c}_{sj})}{u_j'(\hat{c}_{0j})}$ on j. All individuals must conform to it at the equilibrium and the equilibrium allocation is a Pareto optimum. One can then show the existence of a representative agent. The situation is somehow different when markets are incomplete, in which case the equalization of marginal rates of substitution can only take place in correspondence with a set of endowments distribution of measure zero. In this case, the situation is usually described as the one with *constrained Pareto optima* (constrained by nature!); unfortunately, it also turns out that there do not even exist constrained Pareto optima in the multiperiod models with incomplete markets (with the exception of endowments distributions with zero measure).

As regards the computation of the equilibrium, the first order conditions can be rewritten as:

$$\begin{cases} \hat{c}_{0j} &= I_j(\lambda_j) \\ \hat{c}_{sj} &= H_j\left(\beta_j^{-1}\lambda_j m_s\right) \end{cases} \tag{2.20}$$

where I_j and H_j denote the inverse functions of u_j' and ν_j', respectively. Such functions inherit the same properties of u_j' and ν_j'. By replacing these functions into the constraint,

$$\begin{aligned} 0 &= \hat{c}_{0j} - w_{0j} + E\left[m \cdot (\hat{c}_j^1 - w_j^1)\right] \\ &= I_j(\lambda_j) - w_{0j} + \sum_{s=1}^{d} P_s\left[m_s \cdot \left(H_j(\beta_j^{-1}\lambda_j m_s) - w_{sj}\right)\right], \end{aligned}$$

which allows one to define the function

$$z_j(\lambda_j) \equiv I_j(\lambda_j) + E\left[m H_j(\beta_j^{-1}\lambda_j m)\right] = w_{0j} + E\left(m \cdot w_j^1\right). \tag{2.21}$$

We see that $\lim_{x\to 0} z(x) = \infty$, $\lim_{x\to\infty} z(x) = 0$ and

$$z_j'(x) \equiv I_j'(x) + E\left[\beta_j^{-1} m^2 H_j'(\beta_j^{-1} m x)\right] < 0, \ \forall x \in \mathbb{R}_+.$$

Therefore, relation (2.12) determines an unique solution for λ_j:

$$\lambda_j \equiv \Lambda_j\left[w_{0j} + E(m \cdot w_j^1)\right],$$

where $\Lambda(\cdot)$ denotes the inverse function of z. By replacing back into (2.11) we finally get the solution of the program:

$$\begin{cases} \hat{c}_{0j} &= I_j\left(\Lambda_j\left(w_{0j} + E(m \cdot w_j^1)\right)\right) \\ \hat{c}_{sj} &= H_j\left(\beta_j^{-1} m_s \Lambda_j\left(w_{0j} + E(m \cdot w_j^1)\right)\right) \end{cases}$$

It remains to compute the general equilibrium. The kernel m must be determined. This means that we have d unknowns (m_s, $s = 1, \cdots, d$). We have $d+1$ equilibrium conditions (on the $d+1$

markets). By the Walras' law only d of these are independent. To fix ideas, we only consider the equilibrium conditions on the d markets in the second period:

$$g_s\left(m_s;(m_{s'})_{s'\neq s}\right)\equiv\sum_{j=1}^{n}H_j\left(\beta_j^{-1}m_s\Lambda_j\left(w_{0j}+E(m\cdot w_j^1)\right)\right)=\sum_{j=1}^{n}w_{sj}\equiv w_s,\ s=1,\cdots,d,$$

with which one determines the kernel $(m_s)_{s=1}^{d}$ with which it is possible to compute prices and equilibrium allocations. Further, once that one computes the optimal $\hat{c}_0,\hat{c}_s,\ s=1,\cdots,d$, one can find back the $\hat{\theta}$ that generated them by using $\hat{\theta}=V^{-1}(\hat{c}^1-w^1)$.

2.6.4 Existence

Trivial if $m=d$.

2.7 Consumption-based CAPM

Consider the fundamental pricing equation (2.19). It states that for every asset with gross return $\tilde{R}\equiv q^{-1}\cdot payoff$,

$$1=E(m\cdot\tilde{R}),\tag{2.22}$$

where m is some pricing kernel. We've just seen that in a complete markets economy, equilibrium leads to the following identification of the pricing kernel,

$$m_s=\beta\frac{\nu'(w_s)}{u'(w_0)}.$$

For a riskless asset, $1=E(m\cdot R)$. By combining this equality with eq. (2.22) leaves $E[m\cdot(\tilde{R}-R)]=0$. By rearranging terms,

$$E(\tilde{R})=R-\frac{cov(\nu'(w^+),\tilde{R})}{E\left[\nu'(w^+)\right]}.\tag{2.23}$$

2.7.1 The beta relationship

Suppose there is a $\tilde{R}_m$ such that

$$\tilde{R}_m=-\gamma^{-1}\cdot\nu'\left(w_s\right),\quad \text{all } s.$$

In this case,

$$E(\tilde{R})=R+\frac{\gamma\cdot cov(\tilde{R}_m,\tilde{R})}{E\left[v'(w^+)\right]}\quad\text{and}\quad E(\tilde{R}_m)=R+\frac{\gamma\cdot var(\tilde{R}_m)}{E\left[\nu'(w^+)\right]}.$$

These relationships can be combined to yield,

$$E(\tilde{R})-R=\beta\cdot[E(\tilde{R}_m)-R];\quad \beta\equiv\frac{cov(\tilde{R}_m,\tilde{R})}{var(\tilde{R}_m)}.$$

2.7.2 The risk-premium

Eq. (2.23) can be rewritten as,

$$E(\tilde{R}) - R = -\frac{cov(m,\tilde{R})}{E(m)} = -R \cdot cov(m,\tilde{R}). \tag{2.24}$$

The economic interpretation of this equation is that the risk-premium to invest in the asset is high for securities which pay high returns when consumption is high (i.e. when we don't need high returns) and low returns when consumption is low (i.e. when we need high returns). This simple remark can be shown to work very simply with a quadratic utility function $v(x) = ax - \frac{1}{2}bx^2$. In this case we have

$$E(\tilde{R}) - R = \frac{\beta b}{a - bw} R \cdot cov(w^+, \tilde{R}).$$

2.7.3 CCAPM & CAPM

Next, let $\tilde{R}^p$ be the portfolio (factor) return which is the most highly correlated with the pricing kernel m. We have,

$$E(\tilde{R}^p) - R = -R \cdot cov(m, \tilde{R}^p). \tag{2.25}$$

Using (2.24) and (2.25),

$$\frac{E(\tilde{R}) - R}{E(\tilde{R}^p) - R} = \frac{cov(m,\tilde{R})}{cov(m,\tilde{R}^p)},$$

and by rearranging terms,

$$E(\tilde{R}) - R = \frac{\beta_{\tilde{R},m}}{\beta_{\tilde{R}^p,m}}[E(\tilde{R}^p) - R] \quad [\textbf{CCAPM}].$$

If $\tilde{R}^p$ is perfectly correlated with m, i.e. if there exists $\gamma : \tilde{R}^p = -\gamma m$, then

$$\beta_{\tilde{R},m} = -\gamma \frac{cov(\tilde{R}^p, \tilde{R})}{var(\tilde{R}^p)} \quad \text{and} \quad \beta_{\tilde{R}^p,m} = -\gamma$$

and then

$$E(\tilde{R}) - R = \beta_{\tilde{R},\tilde{R}^p}[E(\tilde{R}^p) - R] \quad [\textbf{CAPM}].$$

As we will se in chapter 5, this is not the only way to obtain the CAPM. The CAPM can be obtained also with the so-called "maximum correlation portfolio", i.e. the portfolio that is the most highly correlated with the pricing kernel m.

2.8 Unified budget constraints in infinite horizons models with complete markets

We consider d states of the nature and $m = d$ Arrow securities. We write a unified budget constraint. The result we obtain is useful in applied macroeconomics and finance. It can also be used to show Pareto optimality as in Debreu (1954)[4] (valuation equilibria).

[4]Debreu, G. (1954): "Valuation Equilibrium and Pareto Optimum," *Proceed. Nat. Ac. Sciences*, 40, 588-592.

We have,

$$\begin{cases} p_0\left(c_0 - w_0\right) = -q^{(0)}\theta^{(0)} = -\sum_{i=1}^{m} q_i^{(0)}\theta_i^{(0)} \\ p_s^1\left(c_s^1 - w_s^1\right) = \theta_s^{(0)}, \ s = 1, \cdots, d \end{cases}$$

or,

$$p_0(c_0 - w_0) + \sum_{i=1}^{m} q_i^{(0)}\left[p_i^1\left(c_i^1 - w_i^1\right)\right] = 0.$$

The previous relation holds in a two-periods economy. In a multiperiod economy, in the second period (as in the following periods) agents save indefinitively for the future. In the appendix, we show that,

$$0 = E\left[\sum_{t=0}^{\infty} m_{0,t} \cdot p^t\left(c^t - w^t\right)\right], \tag{2.26}$$

where $m_{0,t}$ are the state prices. From the perspective of time 0, at time t there exist d^t states of nature and thus d^t possible prices.

2.9 Incomplete markets: the finite state-space case

We describe one immediate difficulty arising in an economy with incomplete markets structure. Suppose that ϕ' is an equilibrium state price. Then, all elements of the set

$$\Phi = \left\{\bar{\phi} \in \mathbb{R}_{++}^d : \left(\bar{\phi}' - \phi'\right)V = 0\right\}$$

are also equilibrium state prices. This means that there exists an infinity of equilibrium state prices that are consistent with absence of arbitrage opportunities.[5] In other terms, when markets are incomplete there is an infinity of equilibrium state prices guaranteeing the same, observable assets price vector q (since by definition $\bar{\phi}'V = \phi'V = q$): the "observation" of the current prices in q is consistent with an infinity of admissible state prices. Such a phenomenon is often encountered in most models of finance in which, for a given q_0, there exists an infinity of risk-neutral probability measures P^* defined as $P^* : q_0 = \frac{1}{1+r}E^{P^*}(q_1 + d_1)$.

2.9.1 Nominal assets and real indeterminacy of the equilibrium

The equilibrium is a set of prices $(\hat{p}, \hat{q}) \in \mathbb{R}_{++}^{m\cdot(d+1)} \times \mathbb{R}_{++}^a$ such that:

$$0 = \sum_{j=1}^{n} e_{0j}\left(\hat{p}, \hat{q}\right); \qquad 0 = \sum_{j=1}^{n} e_{1j}\left(\hat{p}, \hat{q}\right); \qquad 0 = \sum_{j=1}^{n} \theta_j\left(\hat{p}, \hat{q}\right),$$

where the previous functions are the results of optimal plans of the agents. This system has $m \cdot (d+1) + a$ equations and $m \cdot (d+1) + a$ unknowns, where $a \leq d$. Let us aggregate the constraints of the agents:

$$p_0 \sum_{j=1}^{n} e_{0j} = -q \sum_{j=1}^{n} \theta_j; \qquad p_1 \square \sum_{j=1}^{n} e_{1j} = B \sum_{j=1}^{n} \theta_j$$

[5] This is so because $\dim \ker(V) = d - a$.

If the financial markets clearing condition is satisfied, i.e. $\sum_{j=1}^{n}\theta_j = 0$, then:

$$\begin{cases} 0 = p_0 \sum_{j=1}^{n} e_{0j} \equiv p_0 e_0 = \sum_{\ell=1}^{m} p_0^{(\ell)} e_0^{(\ell)} \\ \mathbf{0}_d = p_1 \square \sum_{j=1}^{n} e_{1j} \equiv p_1 \square e_1 = \left(\sum_{\ell=1}^{m} p_1^{(\ell)}(\omega_1) e_1^{(\ell)}(\omega_1), \cdots, \sum_{\ell=1}^{m} p_1^{(\ell)}(\omega_d) e_1^{(\ell)}(\omega_d) \right)' \end{cases}$$

This means that each state of nature has a redundant equation. Therefore, there exist $d+1$ redundant equations in total. As a result of this, our system has less independent equations (precisely, $m \cdot (d+1) - 1$) than unknowns (precisely, $m \cdot (d+1) + d$), i.e., an indeterminacy degree equal to $d+1$.

The previous indeterminacy result is independent on the markets structure (i.e., complete vs. incomplete), and in fact it reproduces by means of a different argument a result obtained in chapter 2, section 2.5. In any case, such a result is not really an indeterminacy result when markets are complete and when we assume that agents organize themselves and concentrate the exchanges at the beginning of the economy: in this case, onle the (suitably normalized) Arrow-Debreu state prices would matter for agents.

As pointed out in chapter 2 (section 2.5), the dimension of such an indeterminacy can be reduced to $d-1$ because we can use two additional homogeneity relationships. The first one is obtained by noticing that the budget constraint of each agent,

$$\begin{cases} p_0 e_{0j} = -q\theta_j \\ p_1 \square e_{1j} = B\theta_j \end{cases}$$

still remains the same when one multiplies the first equation by a positive constant. This essentially means that if $(\hat{p}_0, \hat{p}_1, \hat{q})$ is an equilibrium, $(\lambda\hat{p}_0, \hat{p}_1, \lambda\hat{q})$ is also an equilibrium. The second homogeneity relationship is obtained 1) by multiplying the spot prices of the second period by a positive constant, and 2) by increasing the agents' purchasing power by dividing each element of the asset prices vector by the same constant:

$$\begin{cases} p_0 e_{0j} = -\frac{q}{\lambda}\lambda\theta_j \\ \lambda p_1 \square e_{1j} = B\lambda\theta_j \end{cases}$$

Therefore, if $(\hat{p}_0, \hat{p}_1, \hat{q})$ is an equilibrium, $\left(\hat{p}_0, \lambda\hat{p}_1, \frac{\hat{q}}{\lambda}\right)$ is also an equilibrium.

2.9.2 Nonneutrality of money

The previous indeterminacy results are due to the fact that the considered structure of contracts was of the *nominal* type: the payoffs matrix had elements expressed in terms of one *unité de compte* which was not made precise among other things ... One immediately sees that this would not be anymore the case if one considered contracts of the *real* type, i.e. contracts the payoffs of which are directly expressed in terms of the goods. Indeed, in this case the agents constraints would be:

$$\begin{cases} p_0 e_{0j} = -q\theta_j \\ p_1(\omega_s) e_{1j}(\omega_s) = p_1(\omega_s) A_s \theta_j, \ s = 1, \cdots, d \end{cases}$$

where $A_s = [A_s^1, \cdots, A_s^a]$ is the $m \times a$ matrix the real payoffs. The previous constraint immediately reveals how to "recover" $d+1$ homogeneity relationships: for each strictly positive vector $\lambda = (\lambda_0, \lambda_1 \cdots, \lambda_d)$, if $(\hat{p}_0, q, p_1(\omega_1), \cdots, p_1(\omega_s), \cdots, p_1(\omega_d))$ is an equilibrium, then

$(\lambda_0\hat{p}_0, \lambda_0 q, p_1(\omega_1), \cdots, p_1(\omega_s), \cdots, p_1(\omega_d))$ is also an equilibrium, as is also naturally the case of $(\hat{p}_0, q, p_1(\omega_1), \cdots, \lambda_s p_1(\omega_s), \cdots, p_1(\omega_d))$ for λ_s, $s = 1, \cdots, d$.

As is clear, the distinction between nominal and real assets has a very precise sense when one considers a multi-commodity economy. Even in this case, however, such a distinctions is not very interesting without a suitable introduction of a *unité de compte*. Such considerations motivated Magill and Quinzii[6] to propose an elegant way to solve for indetermincay while still remaining in a framework with nominal assets. They simply propose to introduce money as a mean of exchange via the scheme inspiring the construction of the model in section 1.6.3. The indeterminacy can then be resolved by "fixing" the prices via the $d + 1$ equations defining the money market equilibrium in all states of nature:

$$M_s = p_s \cdot \sum_{j=1}^{n} w_{sj}, \ s = 0, 1, \cdots, d.$$

Magill and Quinzii (1992) showed that the monetary policy $(M_s)_{s=0}^{d}$ is generically nonneutral. A heuristic reason explaining such a result is the following one: in this framework, money is "selecting" equilibria which would otherwise be indetermined; but there exists a continuum of equilibria which one can thusly generated. Therefore, money is really determining equilibrium allocations.

2.10 Broader definitions of risk - Rothschild and Stiglitz theory

The papers are Rothschild and Stiglitz (1970, 1971).

First, we introduce the notion of stochastic dominance:

Definition 2.14 (Second-order stochastic dominance). *$\tilde{x}_2$ dominates $\tilde{x}_1$ if for each Von-Neumann/Morgenstein utility function u satisfying $u' \geq 0$, we have also that $E[u(\tilde{x}_2)] \geq E[u(\tilde{x}_1)]$.*

We have:

Theorem 2.15. *The following statements are equivalent:*

a) *$\tilde{x}_2$ dominates $\tilde{x}_1$, or $E[u(\tilde{x}_2)] \geq E[u(\tilde{x}_1)]$;*

b) *$\exists$ r.v. $\epsilon > 0 : \tilde{x}_2 = \tilde{x}_1 + \epsilon$;*

c) *$\forall x > 0, \ F_1(x) \geq F_2(x)$.*

Proof. We work on a compact support $[a, b]$. A more general proof can be provided but it would add nothing really important on an economic standpoint.

First we show that *b*) ⇒ *c*). We have: $\forall t_0 \in [a, b]$, $F_1(t_0) \equiv \Pr(\tilde{x}_1 \leq t_0) = \Pr(\tilde{x}_2 \leq t_0 + \epsilon) \geq \Pr(\tilde{x}_2 \leq t_0) \equiv F_2(t_0)$. Next, we show that *c*) ⇒ *a*). By integrating by parts we get,

$$E[u(x)] = \int_a^b u(x)dF(x) = u(b) - \int_a^b u'(x)F(x)dx,$$

[6] Cf. chapter 1, section 1.6.3, for the references.

where we have used the fact that: $F(a) = 0$ and $F(b) = 1$.
Therefore,

$$E[u(\tilde{x}_2)] - E[u(\tilde{x}_1)] = \int_a^b u'(x)[F_1(x) - F_2(x)]\,dx.$$

Finally, it's easy to show that $a) \Rightarrow b)$, and we're done. ‖

Next, we turn to the definition of "increasing risk":

Definition 2.16. *$\tilde{x}_1$ is more risky than $\tilde{x}_2$ if, for each function u satisfying $u'' < 0$, we have also that $E[u(\tilde{x}_1)] \leq E[u(\tilde{x}_2)]$ for $\tilde{x}_1$ and $\tilde{x}_2$ having the same mean.*

Such a definition holds independently of the sign of u'. Furthermore, if $var(\tilde{x}_1) > var(\tilde{x}_2)$, $\tilde{x}_1$ is not necessarily more risky than $\tilde{x}_2$ in the sense of the previous definition. The standard counterexample is the following one. Let $\tilde{x}_2 = 1$ with prob. 0.8, and 100 with prob. 0.2; let $\tilde{x}_1 = 10$ with prob. 0.99, and 1090 with prob. 0.01. We have that $E(\tilde{x}_1) = E(\tilde{x}_2) = 20.8$, but $var(\tilde{x}_1) = 11762.204$ and $var(\tilde{x}_2) = 1647.368$. Take, however, $u(x) = \log x$. We have $E(\log(\tilde{x}_1)) = 2.35 > E(\log(\tilde{x}_2)) = 0.92$. Preferences matter when defining risk! It is also easily seen that in this particular example, the distribution function F_1 of $\tilde{x}_1$ "intersects" F_2,[7] and this is in contradition with the following theorem.

Theorem 2.17. *The following statements are equivalent:*

a) *$\tilde{x}_1$ is more risky than $\tilde{x}_2$;*

b) $\forall t$, $\int_{-\infty}^{t}[F_1(x) - F_2(x)]\,dx \geq 0$ *($\tilde{x}_1$ has more weight in the tails than $\tilde{x}_2$);*

c) *$\tilde{x}_1$ is a mean preserving spread of $\tilde{x}_2$, i.e. $\exists$ r.v. $\epsilon : \tilde{x}_1 \overset{d}{=} \tilde{x}_2 + \epsilon$ and $E[\epsilon / \tilde{x}_2 = x_2] = 0$.*[8]

Proof. Let's begin with $c) \Rightarrow a)$. We have:

$$\begin{aligned} E[u(\tilde{x}_1)] &= E[u(\tilde{x}_2 + \epsilon)] \\ &= E[E(u(\tilde{x}_2 + \epsilon)|\,\tilde{x}_2 = x_2)] \\ &\leq E\{u(E(\tilde{x}_2 + \epsilon|\,\tilde{x}_2 = x_2))\} \\ &= E[u(E(\tilde{x}_2|\,\tilde{x}_2 = x_2))] \\ &= E[u(\tilde{x}_2)]. \end{aligned}$$

[7] Naturally, the intersection is only an ideal intersection given the discreteness of the state-space considered here. However, the example makes well the point.

[8] The symbol $\overset{d}{=}$ means that the two r.v. at the left and the right of it have the same distribution.

As regards $a) \Rightarrow b)$, we have that:

$$
\begin{aligned}
& E\left[u\left(\tilde{x}_1\right)\right]-E\left[u\left(\tilde{x}_2\right)\right] \\
= & \int_a^b u(x)\left[f_1(x)-f_2(x)\right] dx \\
= & u(x)\left[F_1(x)-F_2(x)\right]\Big|_a^b-\int_a^b u'(x)\left[F_1(x)-F_2(x)\right] dx \\
= & -\int_a^b u'(x)\left[F_1(x)-F_2(x)\right] dx \\
= & -\left[u'(x)\left[\bar{F}_1(x)-\bar{F}_2(x)\right]\Big|_a^b-\int_a^b u''(x)\left[\bar{F}_1(x)-\bar{F}_2(x)\right] dx\right] \\
= & \int_a^b u''(x)\left[\bar{F}_1(x)-\bar{F}_2(x)\right] dx-u'(b)\left[\bar{F}_1(b)-\bar{F}_2(b)\right],
\end{aligned}
$$

where $\bar{F}_i(x)=\int_a^x F_i(u)du$. Now, $\tilde{x}_1$ is more risky than $\tilde{x}_2$ means that $E\left[u\left(\tilde{x}_1\right)\right]<E\left[u\left(\tilde{x}_2\right)\right]$ for $u''<0$. By the previous relation, we must then have that $\bar{F}_1(x)>\bar{F}_2(x)$, and we are done.

Finally, see Rothschild and Stiglitz (1970) p. 238 for the proof of $b) \Rightarrow c)$. $\|$

2.11 Appendix 1

In this appendix we prove that the program in eq. (???) admits a unique maximum. Suppose on the contrary the existence of two maxima:

$$\bar{c} = (\bar{c}_{1j}, \cdots, \bar{c}_{mj}) \text{ and } \bar{\bar{c}} = (\bar{\bar{c}}_{1j}, \cdots, \bar{\bar{c}}_{mj}).$$

Naturally, these maxima would verify $u_j(\bar{c}) = u_j(\bar{\bar{c}})$ with $\sum_{i=1}^m p_i \bar{c}_{ij} = \sum_{i=1}^m p_i \bar{\bar{c}}_{ij} = R_j$.To validate the last claim suppose on the contrary that $\sum_{i=1}^m p_i \bar{\bar{c}}_{ij} < R_j$. In this case the consumption bundle

$$\bar{\bar{\bar{c}}} = (\bar{\bar{c}}_{1j} + \varepsilon, \cdots, \bar{\bar{c}}_{mj}), \ \varepsilon > 0,$$

would be preferred to $\bar{\bar{c}}$ (by assumption (A1)) and, at the same time, it would hold that[9]

$$\sum_{i=1}^m p_i \bar{\bar{\bar{c}}}_{ij} = \varepsilon p_1 + \sum_{i=1}^m p_i \bar{\bar{c}}_{ij} < R_j, \text{ for sufficiently small } \varepsilon.$$

Hence, $\bar{\bar{\bar{c}}}$ would be a solution of program (1.1), thus contradicting the optimality of $\bar{\bar{c}}$. The conclusion is that the existence of two optima implies a full use of resources. Next, consider a point y lying between $\bar{c}$ and $\bar{\bar{c}}$, viz $y = \alpha\bar{c} + (1-\alpha)\bar{\bar{c}}$, $\alpha \in (0,1)$. By assumption (A3),

$$u_j(y) = u_j\left(\alpha\bar{c} + (1-\alpha)\bar{\bar{c}}\right) > u_j(\bar{c}) = u_j(\bar{\bar{c}}).$$

Furthermore,

$$\sum_{i=1}^m p_i y_i = \sum_{i=1}^m p_i\left(\alpha\bar{c}_{ij} + (1-\alpha)\bar{\bar{c}}_{ij}\right) = \alpha\sum_{i=1}^m p_i\bar{c}_{ij} + \textstyle\sum_{i=1}^m \bar{\bar{c}}_{ij} - \alpha\sum_{i=1}^m \bar{\bar{c}}_{ij} = \alpha R_j + R_j - \alpha R_j = R_j.$$

Hence, $y \in B_j(p)$ and is also strictly preferred to $\bar{c}$ and $\bar{\bar{c}}$, which means that $\bar{c}$ and $\bar{\bar{c}}$ are not optima, as initially conjectured. This establishes uniqueness of the solution of the program (1.1), as claimed before.

[9] $A \equiv \sum_{i=1}^m p_i \bar{\bar{c}}_{ij}$. $A < R_j \Rightarrow \exists\varepsilon > 0 : A + \varepsilon p_1 < R_j$. E.g., $\varepsilon p_1 = R_j - A - \eta$, $\eta > 0$. The condition is then: $\exists\eta > 0 : R_j - A > \eta$.

2.12 Appendix 2: Separation of two convex sets

In addition to be useful for the Pareto's optimality issues treated in this chapter, the following theorem will be used to prove the fundamental theorem of finance given for the discrete space case of the following chapter.

THEOREM 2.A1 (Minkowski's separation theorem). *Let A and B be two non-empty convex subsets of $\mathbb{R}^d$. If A is closed, B is compact and $A \bigcap B = \emptyset$, then there exists a $\phi \in \mathbb{R}^d$ and two real numbers d_1, d_2 such that:*

$$a^\top \phi \leq d_1 < d_2 \leq b^\top \phi, \quad \forall a \in A, \, \forall b \in B.$$

2.13 Appendix 3: Proof of theorem 2.11

Relation (2.4) holds for any compact subset of $\mathbb{R}_+^{d+1}$, and therefore it also holds when it is restricted to the unit simplex in $\mathbb{R}_+^{d+1}$:

$$\langle W \rangle \bigcap \mathcal{S}^d = \{0\} .$$

By the Minkowski's separation theorem given in the appendix of chapter 1, $\exists \tilde{\phi} \in \mathbb{R}^{d+1} : w^\top \tilde{\phi} \leq d_1 < d_2 \leq \sigma^\top \tilde{\phi}$, $w \in \langle W \rangle$, $\sigma \in \mathcal{S}^d$. By walking along the simplex boundaries, one finds that $d_1 < \tilde{\phi}_s$, $s = 1, \cdots, d$. On the other hand, $0 \in \langle W \rangle$, which reveals that $d_1 \geq 0$, and $\tilde{\phi} \in \mathbb{R}_{++}^{d+1}$. Next we show that $w^\top \tilde{\phi} = 0$. Assume the contrary, i.e. $\exists w_* \in \langle W \rangle$ that satisfies at the same time $w_*^\top \tilde{\phi} \neq 0$. In this case, there would be a real number ϵ with $\text{sign}(\epsilon) = \text{sign}(w_*^\top \tilde{\phi})$ such that $\epsilon w_* \in \langle W \rangle$ and $\epsilon w_*^\top \tilde{\phi} > d_2$, a contradiction. Therefore, we have $0 = \tilde{\phi}^\top W \theta = (\tilde{\phi}^\top (-q \ V)^\top) \theta = (-\tilde{\phi}_0 q + \tilde{\phi}_{(d)}^\top V)\theta$, $\forall \theta \in \mathbb{R}^m$, where $\tilde{\phi}_{(d)}$ contains the last d components of $\tilde{\phi}$. Whence $q = \phi^\top V$, where $\phi^\top = \left(\frac{\tilde{\phi}_1}{\tilde{\phi}_0}, \cdots, \frac{\tilde{\phi}_d}{\tilde{\phi}_0} \right)$.

The proof of the converse is immediate (hint: multiply by θ): **shown in further notes**.

The proof of the second part is the following one. We have that "each point of $\mathbb{R}^{d+1}$ is equal to each point of $\langle W \rangle$ plus each point of $\langle W \rangle^\perp$", or $\dim \langle W \rangle + \dim \langle W \rangle^\perp = d + 1$. Since $\dim \langle W \rangle = \text{rank}(W)$, $\dim \langle W \rangle^\perp = d + 1 - \dim \langle W \rangle$, and since $q = \phi^\top V$ in the absence of arbitrage opportunities, $\dim \langle W \rangle = \dim \langle V \rangle = m$, whence:

$$\dim \langle W \rangle^\perp = d - m + 1.$$

In other terms, before we showed that $\exists \tilde{\phi} : \tilde{\phi}^\top W = 0$, or $\tilde{\phi}^\top \in \langle W \rangle^\perp$. Whence $\dim \langle W \rangle^\perp \geq 1$ in the absence of arbitrage opportunities. The previous relation provides more information. Specifically, $\dim \langle W \rangle^\perp = 1$ if and only if $d = m$. In this case, $\dim\{\tilde{\phi} \in \mathbb{R}_+^{d+1} : \tilde{\phi}^\top W = 0\} = 1$, which means that the relation $-\tilde{\phi}_0 q + \tilde{\phi}_d^\top V = 0$ also holds true for $\tilde{\phi}^* = \tilde{\phi} \cdot \lambda$, for every positive scalar λ, but there are no other possible candidates. Therefore, $\phi^\top = \left(\frac{\tilde{\phi}_1}{\tilde{\phi}_0}, \cdots, \frac{\tilde{\phi}_d}{\tilde{\phi}_0} \right)$ is such that $\phi = \phi(\lambda)$, and then it is unique. By a similar reasoning, $\dim\{\tilde{\phi} \in \mathbb{R}_+^{d+1} : \tilde{\phi}^\top W = 0\} = d - m + 1 \Rightarrow \dim \left\{ \phi \in \mathbb{R}_{++}^d : q = \phi^\top V \right\} = d - m$. ‖

2.14 Appendix 4: Proof of eq. (2.26)

We make use of the following notation: $q_{s',s}^{(2)(\ell)}$ is the price at $t = 2$ in state s' if the state in $t = 1$ was s for the Arrow security promising 1 unit of numéraire in state ℓ at $t = 3$; $q_{s',s}^{(2)} = (q_{s',s}^{(2)(1)}, \cdots, q_{s',s}^{(2)(m)})$; $\theta_i^{(1)(s)}$ is the quantity acquired at $t = 1$ in state i of Arrow securities promising 1 unit of numéraire if s at $t = 2$; $p_{s,i}^2$ is the price of the good at $t = 2$ in state s if the previous state at $t = 1$ was i; $q^{(0)(i)}$ and $q_s^{(1)(i)}$ correspond to $q_{s',s}^{(2)(\ell)}$; $q^{(0)}$ and $q_s^{(1)}$ correspond to $q_{s',s}^{(2)}$.

The constraint becomes:

$$\begin{cases} p_0 (c_0 - w_0) = -q^{(0)}\theta^{(0)} = -\sum_{i=1}^{m} q^{(0)(i)}\theta^{(0)(i)} \\ p_s^1 \left(c_s^1 - w_s^1\right) = \theta^{(0)(s)} - q_s^{(1)}\theta_s^{(1)} = \theta^{(0)(s)} - \sum_{i=1}^{m} q_s^{(1)(i)}\theta_s^{(1)(i)}, \ s = 1, \cdots, d. \end{cases} \tag{2.27}$$

Here $q_s^{(1)(i)}$ is the price to be paid at time 1 and in state s, for an Arrow security giving 1 unit of numéraire if the state at time 2 is i.

By replacing the second equation of (3.9) in the first one:

$$p_0 (c_0 - w_0) = -\sum_{i=1}^{m} q^{(0)(i)} \left[p_i^1 \left(c_i^1 - w_i^1\right) + q_i^{(1)}\theta_i^{(1)}\right]$$

$\Longleftrightarrow$

$$\begin{aligned} 0 &= p_0 (c_0 - w_0) + \sum_{i=1}^{m} q^{(0)(i)} p_i^1 \left(c_i^1 - w_i^1\right) + \sum_{i=1}^{m} q^{(0)(i)} q_i^{(1)}\theta_i^{(1)} \\ &= p_0 (c_0 - w_0) + \sum_{i=1}^{m} q^{(0)(i)} p_i^1 \left(c_i^1 - w_i^1\right) + \sum_{i=1}^{m} q^{(0)(i)} \sum_{i=1}^{m} q_i^{(1)(j)}\theta_i^{(1)(j)} \\ &= p_0 (c_0 - w_0) + \sum_{i=1}^{m} q^{(0)(i)} p_i^1 \left(c_i^1 - w_i^1\right) + \sum_{i=1}^{m}\sum_{j=1}^{m} q^{(0)(i)} q_i^{(1)(j)}\theta_i^{(1)(j)} \end{aligned} \tag{2.28}$$

At time 2,

$$p_{s,i}^2 \left(c_{s,i}^2 - w_{s,i}^2\right) = \theta_i^{(1)(s)} - q_{s,i}^{(2)}\theta_{s,i}^{(2)} = \theta_i^{(1)(s)} - \sum_{\ell=1}^{m} q_{s,i}^{(2)(\ell)}\theta_{s,i}^{(2)(\ell)}, \ s = 1, \cdots, d. \tag{2.29}$$

Here $q_{s,i}^{(2)}$ is the price vector, to be paid at time 2 in state s if the previous state was i, for the Arrow securities expiring at time 3. The other symbols have a similar interpretation.

By plugging (3.11) into (3.10),

$$\begin{aligned} 0 &= p_0 (c_0 - w_0) + \sum_{i=1}^{m} q^{(0)(i)} p_i^1 \left(c_i^1 - w_i^1\right) + \sum_{i=1}^{m}\sum_{j=1}^{m} q^{(0)(i)} q_i^{(1)(j)} \left[p_{j,i}^2 \left(c_{j,i}^2 - w_{j,i}^2\right) + q_{j,i}^{(2)}\theta_{j,i}^{(2)}\right] \\ &= p_0 (c_0 - w_0) + \sum_{i=1}^{m} q^{(0)(i)} p_i^1 \left(c_i^1 - w_i^1\right) + \sum_{i=1}^{m}\sum_{j=1}^{m} q^{(0)(i)} q_i^{(1)(j)} p_{j,i}^2 (c_{j,i}^2 - w_{j,i}^2) \\ &+ \sum_{i=1}^{m}\sum_{j=1}^{m}\sum_{\ell=1}^{m} q^{(0)(i)} q_i^{(1)(j)} q_{j,i}^{(2)(\ell)}\theta_{j,i}^{(2)(\ell)} \end{aligned} \tag{2.30}$$

In the absence of arbitrage opportunities, $\exists \phi_{t+1,s'} \in \mathbb{R}_{++}^d$—the state prices vector for $t+1$ if the state in t is s'—such that:

$$q_{s',s}^{(t)(\ell)} = \phi_{t+1,s'}' \cdot e_\ell, \ \ell = 1, \cdots, m,$$

where $e_\ell \in \mathbb{R}^d_+$ and has all zeros except in the ℓ-th component which is 1. We may now re-write the previous relationship in terms of the kernel $m_{t+1,s'} = \left(m^{(\ell)}_{t+1,s'}\right)^d_{\ell=1}$ and the probability distribution $P_{t+1,s'} = \left(P^{(\ell)}_{t+1,s'}\right)^d_{\ell=1}$ of the events in $t+1$ when the state in t is s':

$$q^{(t)(\ell)}_{s',s} = m^{(\ell)}_{t+1,s'} \cdot P^{(\ell)}_{t+1,s'},\ \ell = 1, \cdots, m.$$

By replacing in (3.12), and imposing the transversality condition:

$$\sum_{\ell_1=1}^{m} \sum_{\ell_2=1}^{m} \sum_{\ell_3=1}^{m} \sum_{\ell_4=1}^{m} \cdots \sum_{\ell_t=1}^{m} \cdots q^{(0)(\ell_1)} q^{(1)(\ell_2)}_{\ell_1} q^{(2)(\ell_3)}_{\ell_2,\ell_1} q^{(3)(\ell_4)}_{\ell_3,\ell_2} \cdots q^{(t-1)(\ell_t)}_{\ell_{t-1},\ell_{t-2}} \cdots \underset{t\to\infty}{\to} 0,$$

we get eq. (2.26). $\|$

2.15 Appendix 5: The multicommodity case

The multicommodity case is more interesting, but at the same time it is more delicate to deal with when markets are incomplete. There can even be conceptual difficulties. While standard regularity conditions ensure the existence of an equilibrium in the static and complete markets case, only "generic" existence results are available for the incoplete markets cases. Hart (1974) built up well-chosen examples in which there exist sets of endowments distributions for which no equilibrium can exist. However, Duffie and Shafer (1985) showed that such sets have zero measure, and this justifies the previous terminology ("generic" existence).

Here we only provide a derivation of the contraints. m_t commodities are exchanged in periode t ($t = 0, 1$). The states of nature in the second period are d, and the number of exchanged assets is a. The first period budget constraint is:

$$p_0 e_{0j} = -q\theta_j,\ e_{0j} \equiv c_{0j} - w_{0j}$$

where $p_0 = (p_0^{(1)}, \cdots, p_0^{(m_1)})$ is the first period price vector, $e_{0j} = (e_{0j}^{(1)}, \cdots, e_{0j}^{(m_1)})'$ is the first period excess demands vector, $q = (q_1, \cdots, q_a)$ is the financial asset price vector, and $\theta_j = (\theta_{1j}, \cdots, \theta_{aj})'$ is the vector of assets quantities that agent j buys at the first period.

The second period budget constraint is,

$$\underset{d\times d\cdot m_2}{E_1}\ p_1' = B \cdot \theta_j$$

where

$$\underset{d\times d\cdot m_2}{E_1} = \begin{pmatrix} \underset{1\times m_2}{e_1(\omega_1)} & \underset{1\times m_2}{0} & \cdots & \underset{1\times m_2}{0} \\ \underset{1\times m_2}{0} & \underset{1\times m_2}{e_1(\omega_2)} & \cdots & \underset{1\times m_2}{0} \\ & & & \\ \underset{1\times m_2}{0} & \underset{1\times m_2}{0} & \cdots & \underset{1\times m_2}{e_1(\omega_d)} \end{pmatrix}$$

is the matrix of excess demands, $p_1 = (\underset{m_2\times 1}{p_1(\omega_1)}, \cdots, \underset{m_2\times 1}{p_1(\omega_d)})$ is the matrix of spot prices, and

$$\underset{d\times a}{B} = \begin{pmatrix} v_1(\omega_1) & & v_a(\omega_1) \\ & \ddots & \\ v_1(\omega_d) & & v_a(\omega_d) \end{pmatrix}$$

is the payoffs matrix.

We can rewrite the second period constraint as $p_1 \Box e_{1j} = B \cdot \theta_j$, where e_{1j} is defined similarly as e_{0j}, and $p_1 \Box e_{1j} \equiv (p_1(\omega_1)e_{1j}(\omega_1), \cdots, p_1(\omega_d)e_{1j}(\omega_d))'$. The budget constraints are then:

$$\begin{cases} p_0 e_{0j} = -q\theta_j \\ p_1 \Box e_{1j} = B\theta_j \end{cases}$$

Now suppose that markets are complete, i.e., $a = d$ and B is inversible. The second constraint is then: $\theta_j = B^{-1} p_1 \Box e_{1j}$. Consider without loss of generality Arrow securities, or $B = I$. We have

$\theta_j = p_1 \Box e_{1j}$, and by replacing into the first constraint,

$$
\begin{aligned}
0 &= p_0 e_{0j} + q\theta_j \\
&= p_0 e_{0j} + q p_1 \Box e_{1j} \\
&= p_0 e_{0j} + q \cdot (p_1(\omega_1) e_{1j}(\omega_1), \cdots, p_1(\omega_d) e_{1j}(\omega_d))' \\
&= p_0 e_{0j} + \sum_{i=1}^{d} q_i \cdot p_1(\omega_i) e_{1j}(\omega_i) \\
&= \sum_{h=1}^{m_1} p_0^{(h)} e_{0j}^{(h)} + \sum_{i=1}^{d} q_i \cdot \sum_{\ell=1}^{m_2} p_1^{(\ell)}(\omega_i) e_{1j}^{\ell}(\omega_i) \\
&= \sum_{h=1}^{m_1} p_0^{(h)} e_{0j}^{(h)} + \sum_{i=1}^{d} \sum_{\ell=1}^{m_2} \tilde{p}_1^{(\ell)}(\omega_i) e_{1j}^{\ell}(\omega_i)
\end{aligned}
$$

where $\tilde{p}_1^{(\ell)}(\omega_i) \equiv q_i \cdot p_1^{(\ell)}(\omega_i)$. The price to be paid today for the obtention of a good ℓ in state i is equal to the price of an Arrow asset written for state i multiplied by the spot price $\tilde{p}_1^{(\ell)}(\omega_i)$ of this good in this state; here the Arrow-Debreu state price is $\tilde{p}_1^{(\ell)}(\omega_i)$. The general equilibrium can be analyzed by making reference to such state prices. From now on, we simplify and set $m_1 = m_2 \equiv m$. Then we are left with determining $m(d+1)$ equilibrium prices, i.e. $p_0 = (p_0^{(1)}, \cdots, p_0^{(m)})$, $\tilde{p}_1(\omega_1) = (\tilde{p}_1^{(1)}(\omega_1), \cdots, \tilde{p}_1^{(m)}(\omega_1))$, $\cdots$, $\tilde{p}_1(\omega_d) = (\tilde{p}_1^{(1)}(\omega_d), \cdots, \tilde{p}_1^{(m)}(\omega_d))$. By exactly the same arguments of the previous chapter, there exists one degree of indeterminacy. Therefore, there are only $m(d+1)-1$ relations that can determine the $m(d+1)$ prices.[10] On the other hand, in the initial economy we have to determine $m(d+1)+d$ prices $(\hat{p}, \hat{q}) \in \mathbb{R}_{++}^{m\cdot(d+1)} \times \mathbb{R}_{++}^{d}$ which are solution of the system:

$$
\begin{cases}
\sum_{j=1}^{n} e_{0j}(\hat{p}, \hat{q}) = 0 \\
\sum_{j=1}^{n} e_{1j}(\hat{p}, \hat{q}) = 0 \\
\sum_{j=1}^{n} \theta_j(\hat{p}, \hat{q}) = 0
\end{cases}
$$

where the previous functions are obtained as solutions of the agents' programs.

When we solve for Arrow-Debreu prices, in a second step we have to determine $m(d+1)+d$ prices starting from the knowledge of $m(d+1)-1$ relations defining the Arrow-Debreu prices, which implies a price indeterminacy of the initial economy equal to $d+1$. In fact, it is possible to show that the degree of indeterminacy is only $d-1$.

[10] Price normalisation can be done by letting one of the first period commodities be the numéraire. This is also the common practice in many financial models, in which there is typically one good the price of which at the beginning of the economy is the point of reference.

3
Infinite horizon economies

3.1 Introduction

We consider the mechanics of asset price formation in infinite horizon economies with no frictions. In one class of models, there are agents living forever. These agents have access to a complete markets setting. Their plans are thus equivalent to a single, static plan. In a second class of models, there are overlapping generations of agents.

3.2 Recursive formulations of intertemporal plans

The deterministic case. A "representative" agent solves the following problem:

$$\begin{cases} V_t(w_t, R_{t+1}) & = & \max_{\{c\}} \sum_{i=0}^{\infty} \beta^i u(c_{t+i}) \\ \text{s.t. } w_{t+1} & = & (w_t - c_t) R_{t+1}, \quad \{R_t\}_{t=0}^{\infty} \text{ given} \end{cases}$$

and to simplify notation, we will set $V_t(w_t) \equiv V_t(w_t, R_{t+1})$.

The previous problem can be reformulated in a recursive format:

$$V_t(w_t) = \max_{\{c\}} [u(c_t) + \beta V_{t+1}(w_{t+1})] \text{ s.t. } w_{t+1} = (w_t - c_t) R_{t+1}.$$

We only consider the stationary case in which $V_t(\cdot) = V_{t+1}(\cdot) \equiv V(\cdot)$. The first order condition for c is,

$$0 = u'(c_t) + \beta V'(w_{t+1}) \frac{dw_{t+1}}{dc_t} = u'(c_t) - \beta V'(w_{t+1}) R_{t+1}.$$

Let $c_t^* = c^*(w_t) \equiv c^*(w_t, R_{t+1})$ be the consumption policy function. The value function and the previous first order conditions can be written as:

$$\begin{aligned} V(w_t) &= u(c^*(w_t)) + \beta V((w_t - c^*(w_t)) R_{t+1}) \\ u'(c^*(w_t)) &= \beta V'((w_t - c^*(w_t)) R_{t+1}) R_{t+1} \end{aligned}$$

Next, differentiate the value function,

$$\begin{aligned} V'(w_t) &= u'\left(c^*(w_t)\right)\frac{dc^*}{dw_t}(w_t) + \left[\beta V'\left(\left(w_t - c^*(w_t)\right)R_{t+1}\right)R_{t+1}\right]\left[1 - \frac{dc^*(w_t)}{dw_t}\right] \\ &= u'\left(c^*(w_t)\right)\frac{dc^*}{dw_t}(w_t) + u'\left(c^*(w_t)\right)\left[1 - \frac{dc^*(w_t)}{dw_t}\right] \\ &= u'\left(c^*(w_t)\right), \end{aligned}$$

where we have used the optimality condition that $u'\left(c^*(w_t)\right) = \beta V'\left(\left(w_t - c^*(w_t)\right)R_{t+1}\right)R_{t+1}$. Therefore, $V'(w_{t+1}) = u'\left(c^*(w_{t+1})\right)$ too, and by substituting back into the first order condition,

$$\beta\frac{u'\left(c^*(w_{t+1})\right)}{u'\left(c^*(w_t)\right)} = \frac{1}{R_{t+1}}.$$

The general idea that helps solving problems such as the previous ones is to consider the value function as depending on "bequests" determined by past choices. In this introductory example, the value function is a function of wealth w. In problems considered below, our value function will be a function of past "capital" k.

One simple example. Let $u = \log c$. In this case, $V'(w) = c(w)^{-1}$ by the envelope theorem. Let's conjecture that $V_t(w) = a_t + b\log w$. If the conjecture is true, it must be the case that $c(w) = b^{-1}w$. But then,

$$\frac{1}{R_{t+1}} = \beta\frac{u'\left(c^*(w_{t+1})\right)}{u'\left(c^*(w_t)\right)} = \beta\frac{c^*(w_t)}{c^*(w_{t+1})} = \beta\frac{w_t}{w_{t+1}},$$

or,

$$\begin{cases} w_{t+1} = \beta w_t R_{t+1} \\ w_{t+1} = \left(w_t - c(w_t)\right)R_{t+1} \end{cases}$$

where the second equation is the budget constraint. Solving the previous two equations in terms of c leaves, $c^*(w_t) = (1-\beta)w_t$. In other terms, $b = (1-\beta)^{-1}$.[1]

3.3 The Lucas' model

The paper is Lucas (1978).[2]

3.3.1 Asset pricing and marginalism

Suppose that at time t you give up to a small quantity of consumption equal to Δc_t. The (current) corresponding utility reduction is then equal to $\beta^t u'(c_t)\Delta c_t$. But by investing Δc_t in a safe asset, you can dispose of $\Delta c_{t+1} = R_{t+1}\Delta c_t$ additional units of consumption at time $t+1$, to which it will correspond an expected (current) utility gain equal to $\beta^{t+1}E\left(u'(c_{t+1})\Delta c_{t+1}\right)$. If

[1] To pin down the coefficient a_t, thus confirming our initial conjecture that $V_t(w) = a_t + b\log w$, we simply use the definition of the value function, $V_t(w_t) \equiv u\left(c^*(w_t)\right) + \beta V_{t+1}(w_{t+1})$. By plugging $V_t(w) = a_t + b\log w$ and $c^*(w) = (1-\beta)^{-1}w$ into the previous definition leaves, $a_t = \log(1-\beta) + \beta a_{t+1} + \frac{\beta}{1-\beta}\log(\beta R_{t+1})$. Hence the "stationary case" is obtained only when R is constant - something which happens in equilibrium. For example, if comsumption c is equal to some (constant) output Z, then in equilibrium for all t, $R_t = \beta^{-1}$ and $a_t = (1-\beta)^{-1}\log(1-\beta)$.

[2] Lucas, R.E. Jr. (1978): "Asset Prices in an Exchange Economy," *Econometrica*, 46, 1429-1445.

c_t and c_{t+1} are part of an optimal consumption plan, there would not be economic incentives to implement such intertemporal consumption transfers. Therefore, along an optimal consumption plan, any utility reductions and gains of the kind considered before should be identical:

$$u'(c_t) = \beta E\left[u'(c_{t+1})R_{t+1}\right].$$

This relationship generalizes relationships derived in the previous sections to the uncertainty case.

Next suppose that Δc_t may now be invested in a risky asset whose price is q_t at time t. You can buy $\Delta c_t / q_t$ units of such an asset that can subsequently be sold at time $t+1$ at the unit price q_{t+1} to finance a (random) consumption equal to $\Delta c_{t+1} = (\Delta c_t / q_t) \cdot (q_{t+1} + D_{t+1})$ at time $t+1$, where D_{t+1} is the divident paid by the asset. The utility reduction as of time t is $\beta^t u'(c_t)\Delta c_t$, and the expected utility gain as of time $t+1$ is $\beta^{t+1} E\left\{u'(c_{t+1})\Delta c_{t+1}\right\}$. As before, you should not be incited to any intertemporal transfers of this kind if you already are along an optimal consumption path. But such an incentive does not exist if and only if utility reduction and the expected utility gains are the same. Such a condition gives:

$$u'(c_t) = \beta E\left[u'(c_{t+1})\frac{q_{t+1} + D_{t+1}}{q_t}\right]. \tag{3.1}$$

3.3.2 Model

The only source of risk is given by future realization of the dividend vector $D = (D_1, \cdots, D_m)$. We suppose that this is a Markov process and note with $P(D^+/D)$ its conditional distribution function. A representative agent solves the following program:

$$V(\theta_t) = \max_{\{c,\theta\}} E\left[\sum_{i=0}^{\infty} \beta^i u(c_{t+i}) \middle| \mathcal{F}_t\right] \qquad \text{s.t. } c_t + q_t\theta_{t+1} = (q_t + D_t)\,\theta_t$$

where $\theta_{t+1} \in \mathbb{R}^m$ is $\mathcal{F}_t$-measurable, i.e. θ_{t+1} must be chosen at time t. Note the similarity with the construction made in chapter 2: θ is self-financing, i.e. $c_{t-1} + q_{t-1}\theta_t = q_{t-1}\theta_{t-1} + D_{t-1}\theta_{t-1}$, and defining the strategy value as $V_t \equiv q_t\theta_t$, we get $\Delta w_t = q_t\theta_t - q_{t-1}\theta_{t-1} = \Delta q_t\theta_t - c_{t-1} + D_{t-1}\theta_{t-1}$, which is exactly what found in chapter 2, section 2.? (ignoring dividends and intermediate consumption).

The Bellman's equation is:

$$V(\theta_t, D_t) = \max_{\{c_t, \theta_{t+1}\}} E\left[u(c_t) + \beta V(\theta_{t+1}, D_{t+1}) \middle| \mathcal{F}_t\right] \text{ s.t. } c_t + q_t\theta_{t+1} = (q_t + D_t)\,\theta_t$$

or,

$$V(\theta, D) = \max_{\theta^+} E\left[u\left((q+D)\theta - q\theta^+\right) + \beta V(\theta^+, D^+)\right].$$

Naturally, $\theta_i^+ = \theta_i^+(\theta, D)$, and the first order condition for θ_i^+ gives:

$$0 = E\left[-u'(c)q_i + \beta\frac{\partial V(\theta^+, D^+)}{\partial \theta_i^+}\right],$$

By differentiating the value function one gets:

$$\frac{\partial V(\theta, D)}{\partial \theta_i} = E\left[u'(c)\left(q_i + D_i - \sum\nolimits_j q_j \frac{\partial \theta_j^+}{\partial \theta_i}\right) + \beta \sum\nolimits_j \frac{\partial V(\theta^+, D^+)}{\partial \theta_j^+}\frac{\partial \theta_j^+}{\partial \theta_i}\right].$$

By replacing the first order condition into the last relationship one gets:

$$\frac{\partial V(\theta, D)}{\partial \theta_i} = u'(c)(q_i + D_i).$$

By substituting back into the first order condition,

$$0 = E\left[-u'(c)q_i + \beta u'(c^+)\left(q_i^+ + D_i^+\right)\right].$$

This is eq. (3.13) derived in the previous subsection.

3.3.3 Rational expectations equilibria

The clearing conditions on the financial market are:

$$\theta_t = \mathbf{1}_m \text{ and } \theta_t^{(0)} = 0 \ \forall t \geq 0,$$

where $\theta^{(0)}$ is the quantity of internal money acquired by the representative agent. It is not hard to show—by using the budget constraint of the representative agent—that such a condition implies the equilibrium on the real market:[3]

$$c = \sum_{i=1}^{m} D_i \equiv Z. \tag{3.2}$$

An *equilibrium* is a sequence of prices $(q_t)_{t=0}^{\infty}$ satisfying (3.9) and (3.10).
By replacing the previous equilibrium condition into relation (3.9) leaves:

$$u'(Z)q_i = \beta E\left[u'\left(Z^+\right)\left(q_i^+ + D_i^+\right)\right] = \beta \int u'\left(Z^+\right)\left(q_i^+ + D_i^+\right) dP\left(D^+ \middle| D\right). \tag{3.3}$$

The rational expectations assumption here consists in supposing the existence of a price function of the form:

$$q_i = q_i(D). \tag{3.4}$$

A *rational expectations equilibrium* is a sequence of prices $(q_t)_{t=0}^{\infty}$ satisfying (3.11) and (3.12).
By replacing $q_i(D)$ into (3.11) leaves:

$$u'(Z)q_i(D) = \beta \int u'\left(Z^+\right)\left[q_i(D^+) + D_i^+\right] dP\left(D^+ \middle| D\right). \tag{3.5}$$

This is a functional equation in $q_i(\cdot)$ that we are going to study by focusing, first, on the i.i.d. case: $P(D^+|\, D) = P(D^+)$.

[3] If we had to consider n agents, each of them would have a constraint of the form $c_t^{(j)} + q_t\theta_{t+1}^{(j)} = (q_t + \zeta_t)\theta_t^{(j)}$, $j = 1, ..., n$, and the financial market clearing conditions would then be $\sum_{j=1}^{n} \theta_t^{(j)} = \mathbf{1}_m$ and $\sum_{j=1}^{n} \theta_t^{(0),j} = 0 \ \forall t \geq 0$, which would imply that $\sum_{j=1}^{n} c_t^{(j)} = \sum_{i=1}^{m} \zeta_t^{(i)}$.

3.3.3.1 IID shocks

Eq. (3.17) simplifies:

$$u'(Z)q_i(D) = \beta \int u'(Z^+) \left[q_i(D^+) + D_i^+\right] dP(D^+). \tag{3.6}$$

The r.h.s. of the previous equation does not depend on D. Therefore $u'(Z)q_i(D)$ is a constant:

$$u'(Z)q_i(D) \equiv \kappa_i \ \text{(say)}.$$

By substituting κ back into (3.18) we get:

$$\kappa_i = \frac{\beta}{1-\beta} \int u'(Z^+)D_i^+ dP(D^+). \tag{3.7}$$

The solution for $q_i(D)$ is then:

$$q_i(D) = \frac{\kappa_i}{u'(Z)},$$

where κ is given by (3.19).

The price-dividend elasticity is then:

$$\frac{D_i}{q_i}\frac{\partial q_i(D)}{\partial D_i} = -D_i\frac{u''(Z)}{u'(Z)}.$$

If one further simplifies and sets $m = 1$, the previous relationhip reveals that the price-dividend elasticity is just equal to a risk-aversion index:

$$\frac{D}{q}q'(D) = -D\frac{u''}{u'}(D).$$

The object $-xu''(x)/\,u'(x)$ is usually referred to as "relative risk aversion" (henceforth, RRA). Furthermore, if the RRA is constant (CRRA) and equal to η, we obtain by integrating that apart from an unimportant constant, $u(x) = x^{1-\eta}/(1-\eta)$, $\eta \in (0,\infty)\backslash\{1\}$ and, for $\eta = 1$, $u(x) = \log x$. Figure 3.1 depicts the graph of the price function in the CRRA case:

$$q(D) = \kappa_\eta \cdot D^\eta,\ \kappa_\eta \equiv \frac{\beta}{1-\beta}\int x^{1-\eta} dP(x).$$

Clearly the higher η the higher the CRRA is. When agents are risk-neutral ($\eta \to 0$), and the asset price is constant and equal to $\beta(1-\beta)^{-1} \cdot E(D)$.

3.3.3.2 Dependent shocks

Define $g_i(D) \equiv u'(Z)q_i(D)$ and $h_i(D) \equiv \beta \int u'(Z^+)D_i^+ dP(D^+|\, D)$. In terms of these new functions, eq. (3.17) is:

$$g_i(D) = h_i(D) + \beta \int g_i(D^+)dP(D^+\big|\, D).$$

The celebrated Blackwell's theorem can now be used to show that there exists one and only one solution g_i to this functional equation.

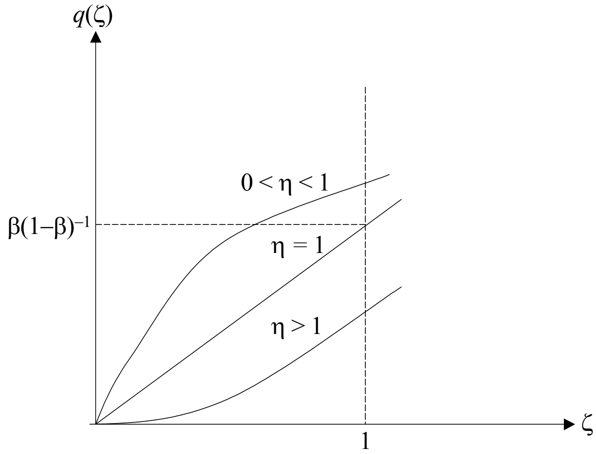


FIGURE 3.1. The asset pricing function $q(D)$ arising when $\beta < \frac{1}{2}$, $\eta \in (0, \infty)$ and $\frac{\partial}{\partial \eta} \kappa\eta < 0$.

THEOREM 3.1. *Let $\mathcal{B}(X)$ the Banach space of continuous bounded real functions on $X \subseteq \mathbb{R}^n$ endowed with the norm $\|f\| = \sup_X |f|$, $f \in \mathcal{B}(X)$. Introduce an operator $T : \mathcal{B}(X) \mapsto \mathcal{B}(X)$ with the following properties*:

i) T is monotone: $\forall x \in X$ *and* $f_1, f_2 \in \mathcal{B}(X)$, $f_1(x) \leq f_2(x) \Longleftrightarrow T[f_1](x) \leq T[f_2](x)$;

ii) $\forall x \in X$ *and* $c \geq 0$, $\exists \beta \in (0, 1) : T[f + c](x) \leq T[f](x) + \beta c$.

Then, T is a β-contraction and, $\forall f_0 \in \mathcal{B}(X)$, it has a unique fixed point $\lim_{\tau \to \infty} T^\tau [f_0] = f = T[f]$.

Motivated by the preceding theorem, we introduce the operator

$$T[g_i](D) = h_i(D) + \beta \int g_i(D^+) dP(D^+ | D).$$

The existence of g_i can thus be reconducted to finding a fixed point of $T : g_i = T[g_i]$. It is easily checked that conditions i) and ii) of theorem 3.1 hold here. It remains to be shown that $T : \mathcal{B}(D) \mapsto \mathcal{B}(D)$. It is sufficient to show that $h_i \in \mathcal{B}(D)$. A sufficient condition given by Lucas (1978) is that u is bounded and bounded away by a constant $\bar{u}$. In this case, concavity of u implies that $\forall x$, $0 = u(0) \leq u(x) + u'(x)(-x) \leq \bar{u} - xu'(x) \Longrightarrow \forall x$, $xu'(x) \leq \bar{u}$. Whence $h_i(D) \leq \beta\bar{u}$. It is then possible to show that the solution is in $\mathcal{B}(D)$, which implies that $T : \mathcal{B}(D) \mapsto \mathcal{B}(D)$.

3.3.4 Arrow-Debreu state prices

We have,

$$q = E\left[m(q^+ + D^+)\right], \; m \equiv \beta \frac{u'(D^+)}{u'(D)}.$$

It is easy to show that,

$$\frac{dP^*}{dP}(D^+ | D) = \frac{u'(D^+)}{E\left[u'(D^+) | D\right]}.$$

To show this, let b denote the (shadow) price of a pure discount bond expiring the next period. Note that $b = E(m) = \int m dP \equiv \int R^{-1} dP^* = E^{P^*}(R^{-1})$, whence:

$$\frac{dP^*}{dP}(D^+ | D) = mR \equiv mb^{-1} = m\left[\beta E\left(\frac{u'(D^+)}{u'(D)}\bigg| D\right)\right]^{-1} = m\beta^{-1}\frac{u'(D)}{E\left[u'(D^+)| D\right]} = \frac{u'(D^+)}{E\left[u'(D^+)| D\right]}.$$

In this model, the Arrow-Debreu state-price density is given by

$$d\tilde{P}^*(D^+ | D) = dP^*(D^+ | D)R^{-1} = dP^*(D^+ | D)b.$$

This is the price to pay, in state D, to obtain one unit of the good the next period in state D^+. To check this, note that

$$\int d\tilde{P}^*(D^+ | D) = \int dP^*(D^+ | D)b = b\int dP^*(D^+ | D) = b.$$

3.3.5 CCAPM & CAPM

Define the gross return $\tilde{R}$ as,

$$\tilde{R} \equiv \frac{q^+ + D^+}{q}.$$

Then all the considerations made in section 2.7 also holds here.

3.4 Production: foundational issues

In the previous sections, the dividend process was taken as exogenous. We wish to consider production-based economies in which firms maximize their value and set dividends endogenously. In the economies we wish to examine, production and capital accumulation are also endogenous. The present section reviews some foundational issues arising in economies with capital accumulation. The next section develops the asset pricing implications of these economies.

3.4.1 Decentralized economy

There is a continuum of identical firms in $(0, 1)$ that have access to capital and labour markets, and produce by means of the following technology:

$$(K, N) \mapsto Y(K, N),$$

where $Y_i(K, N) > 0$, $y_{ii}(K, N) < 0$, $\lim_{K\to 0+} Y_1(K, N) = \lim_{N\to 0+} Y_2(K, N) = \infty$, $\lim_{K\to\infty} Y_1(K, N) = \lim_{N\to\infty} Y_2(K, N) = 0$, and subscripts denote partial derivatives. Furthermore, we suppose that Y is homogeneous of degree one, i.e. for all scalars $\lambda > 0$, $Y(\lambda K, \lambda N) = \lambda Y(K, N)$. We then define per capital production $y(k) \equiv Y(K/N, 1)$, where $k \equiv K/N$ is per-capita capital, and suppose that N obeys the equation

$$N_t = (1 + n) N_{t-1},$$

where n is population growth. Firms reward capital and labor at $R = Y_1(K,N)$ and $w = Y_2(K,N) = w$. We have,[4]

$$\begin{aligned} R &= y'(k) \\ w &= y(k) - ky'(k) \end{aligned}$$

In each period t, there are also N_t identical consumers who live forever. The representative consumer is also the representative worker. To simplify the presentation at this stage, we assume that each consumer offers inelastically one unit of labor, and that $N_0 = 1$ and $n = 0$. Their action plans have the form $((c,s)_t)_{t=0}^{\infty}$, and correspond to sequences of consumption and saving. Their resource constraint is:

$$c_t + s_t = R_t s_{t-1} + w_t N_t, \quad N_t \equiv 1, \quad t = 1, 2, \cdots.$$

The previous equation means that at time $t-1$, each consumer saves s_{t-1} units of capital that lends to the firm. At time t, the consumer receives $R_t s_{t-1}$ back from the firm, where $R_t = y'(k_t)$ (R is a *gross* return on savings). Added to the wage receipts $w_t N_t$ obtained at t, the consumer then at time t consumes c_t and lends s_t to the firm.

At time zero we have:

$$c_0 + s_0 = V_0 \equiv Y_1(K_0, N_0)K_0 + w_0 N_0, \quad N_0 \equiv 1.$$

The previous equation means that at time t, the consumer is endowed with K_0 units of capital (the initial condition) which she invests to obtain $Y_1(K_0, N_0)K_0$ plus the wage recepits $w_0 N_0$.

As in the previous chapter, we wish to write down only one constraint. By iterating the constraints we get:

$$0 = c_0 + \sum_{t=1}^{T} \frac{c_t - w_t N_t}{\prod_{i=1}^{t} R_i} + \frac{s_T}{\prod_{i=1}^{T} R_i} - V_0,$$

and we impose a *transversality condition*

$$\lim_{T\to\infty} \frac{s_T}{\prod_{i=1}^{T} R_i} \to 0+.$$

Provided that $\lim_{T\to\infty} \sum_{t=1}^{T} \frac{c_t}{\prod_{i=1}^{t} R_i}$ exists, we obtain a representation similar to ones obtained in chapter 2:

$$0 = c_0 + \sum_{t=1}^{\infty} \frac{c_t - w_t N_t}{\prod_{i=1}^{t} R_i} - V_0. \tag{3.8}$$

The representative consumer's program can then be written as:

$$\max_c \sum_{t=1}^{\infty} \beta^t u(c_t),$$

under the previous constraint.[5] Here we shall suppose that u share the same properties of y.

[4] To derive these expressions, please note that $\frac{\partial}{\partial K} y(k) = \frac{y'(k)}{N}$. But we also have $\frac{\partial}{\partial K} y(k) = \frac{\partial}{\partial K}\left(\frac{1}{N} Y(K,N)\right) = \frac{1}{N} Y_1(K,N)$. And so by identifying $y'(k) = Y_1(K,N)$. As regards labor, $\frac{\partial}{\partial N} y(k) = -\frac{ky'(k)}{N}$. But we also have, $\frac{\partial}{\partial N} y(k) = \frac{\partial}{\partial N}\left(\frac{1}{N} Y(K,N)\right) = -\frac{y(k)}{N} + \frac{w}{N}$. And so by identifying $w = y(k) - ky'(k)$.

[5] Typically the respect of the transversality condition implies that the optimal consumption path is different from the consumption path resulting from a program without a transversality condition.

We wish to provide an interpretation of the transversality condition. The first order conditions of the representative agent are:

$$\beta^t u'(c_t) = \lambda \frac{1}{\prod_{i=1}^t R_i}, \tag{3.9}$$

where λ is a Lagrange multiplier. In addition, current savings equal next period capital at the equilibrium, or

$$k_{t+1} = s_t. \tag{3.10}$$

By plugging (3.2) and (3.4) into condition (3.1) we get:

$$\lim_{T\to\infty} \beta^T u'(c_T) k_{T+1} \to 0+. \tag{3.11}$$

Capital weighted by marginal utility discounted at the psychological interest rate is nil at the stationary state.

Relation (3.2) also implies that $\beta^{t+1}u(c_{t+1}) = \lambda \frac{1}{\prod_{i=1}^{t+1} R_i}$, which allows us to eliminate λ and obtain:

$$\beta \frac{u'(c_{t+1})}{u'(c_t)} = \frac{1}{R_{t+1}}.$$

Finally, by using the identity $k_{t+1} + c_t \equiv y(k_t)$, we get the following system of difference equations:

$$\left\{ \begin{array}{lll} k_{t+1} & = & y(k_t) - c_t \\ \beta \dfrac{u'(c_{t+1})}{u'(c_t)} & = & \dfrac{1}{y'(k_{t+1})} \end{array} \right. \tag{3.12}$$

In this economy, an equilibrium is a sequence $((\hat{c}, \hat{k})_t)_{t=0}^\infty$ solving system (3.5) and satisfying the transversality condition (3.4).

3.4.2 Centralized economy

The social planner problem is:

$$\left\{ \begin{array}{lll} V(k_0) & = & \max_{(c,k)_t} \sum_{t=0}^{\infty} \beta^t u(c_t) \\ \text{s.t. } k_{t+1} & = & y(k_t) - c_t, \quad k_0 \text{ given} \end{array} \right. \tag{3.13}$$

The solution of this program is in fact the solution of system (3.5) + the transversality condition (3.4). Here are three methods of deriving system (3.5).

- The first method consists in rewriting the program with an infinity of constraints:

  $$\max_{(c,k)_t} \sum_{t=0}^{\infty} \left[\beta^t u(c_t) + \lambda_t \left(k_{t+1} - y(k_t) + c_t \right) \right],$$

  where $(\lambda_t)_{t=0}^\infty$ is a sequence of Lagrange multipliers. The first order conditions are:

  $$\left\{ \begin{array}{lll} 0 & = & \beta^t u'(c_t) + \lambda_t \\ \lambda_{t-1} & = & \lambda_t y'(k_t) \end{array} \right.$$

  whence

  $$0 = \beta^t u'(c_t) + \lambda_{t+1} y'(k_{t+1}) = \beta^t u'(c_t) - \beta^{t+1} u'(c_{t+1}) y'(k_{t+1}).$$

- The second method consists in replacing the constraint $k_{t+1} = y(k_t) + c_t$ into $\sum_{t=0}^{\infty} \beta^t u(c_t)$,

$$\max_{(k)_t} \sum_{t=0}^{\infty} \beta^t u\left(y(k_t) - k_{t+1}\right);$$

The first order condition is:

$$0 = -\beta^{t-1} u'(c_{t-1}) - \beta^t u'(c_t) y'(k_t).$$

- The previous methods (as in fact the methods of the previous section) are heuristic. The rigorous approach consists in casting the problem in a recursive setting. In section 3.5 we explain heuristically that program (3.6) can equivalently be formulated in terms of the *Bellman's equation*:

$$V(k_t) = \max_c \left\{u(c_t) + \beta V(k_{t+1})\right\}, \quad k_{t+1} = y(k_t) - c_t.$$

The first order condition is:

$$0 = u'(c_t) + \beta V'(k_{t+1}) \frac{dk_{t+1}}{dc_t} = u'(c_t) - \beta V'(k_{t+1}) = u'(c_t) - \beta V'\left(y(k_t) - c_t\right).$$

Let us denote the optimal solution as $c_t^* = c^*(k_t)$, and express the value function and the previous first order condition in terms of $c^*(\cdot)$:

$$\begin{cases} V(k_t) = u\left(c^*(k_t)\right) + \beta V\left(y(k_t) - c^*(k_t)\right) \\ u'\left(c^*(k_t)\right) = \beta V'\left(y(k_t) - c^*(k_t)\right) \end{cases}$$

Let us differentiate the value function:

$$V'(k_t) = u'\left(c^*(k_t)\right) \frac{dc^*}{dk_t}(k_t) + \beta V'\left(y(k_t) - c^*(k_t)\right) \left[y'(k_t) - \frac{dc^*}{dk_t}(k_t)\right].$$

By replacing the first order condition into the previous condition,

$$V'(k_t) = u'\left(c^*(k_t)\right) \frac{dc^*}{dk_t}(k_t) + u'\left(c^*(k_t)\right) \left[y'(k_t) - \frac{dc^*}{dk_t}(k_t)\right] = u'\left(c^*(k_t)\right) y'(k_t)$$

This is an *envelope theorem*. By replacing such a result back into the first order condition,

$$u'\left(c^*(k_t)\right) = \beta V'\left(y(k_t) - c^*(k_t)\right) = \beta u'\left(c^*(k_{t+1})\right) y'(k_{t+1}).$$

3.4.3 Deterministic dynamics

The objective here is to study the dynamics behavior of system (3.5). Here we only provide details on such dynamics in a neighborood of the stationary state. Here, the stationary state is defined as the couple (c, k) solution of:

$$\begin{cases} c &= y(k) - k \\ \beta &= \frac{1}{y'(k)} \end{cases}$$

Rewrite system (3.5) as:

$$\begin{cases} k_{t+1} &= y(k_t) - c_t \\ u'(c_t) &= \beta u'(c_{t+1}) y'(k_{t+1}) \end{cases}$$

A first order Taylor expansion of the two sides of the previous system near (c, k) yields:

$$\begin{cases} k_{t+1} - k & = y'(k)\,(k_t - k) - (c_t - c) \\ u''(c)\,(c_t - c) & = \beta u''(c) y'(k)\,(c_{t+1} - c) + \beta u'(c) y''(k)\,(k_{t+1} - k) \end{cases}$$

By rearranging terms and taking account of the relationship $\beta y'(k) = 1$,

$$\begin{cases} \hat{k}_{t+1} & = y'(k)\hat{k}_t - \hat{c}_t \\ \hat{c}_{t+1} & = \hat{c}_t - \beta \frac{u'(c)}{u''(c)} y''(k) \hat{k}_{t+1} \end{cases}$$

where we have defined $\hat{x}_t \equiv x_t - x$. By plugging the first equation into the second equation and using again the relationship $\beta y'(k) = 1$, we eventually get the following linear system:

$$\begin{pmatrix} \hat{k}_{t+1} \\ \hat{c}_{t+1} \end{pmatrix} = A \begin{pmatrix} \hat{k}_t \\ \hat{c}_t \end{pmatrix} \tag{3.14}$$

where

$$A \equiv \begin{pmatrix} y'(k) & -1 \\ -\frac{u'(c)}{u''(c)} y''(k) & 1 + \beta \frac{u'(c)}{u''(c)} y''(k) \end{pmatrix}.$$

The solution of such a system can be obtained with standard tools that are presented in appendix 1 of the present chapter. Such a solution takes the following form:

$$\begin{cases} \hat{k}_t & = v_{11} \kappa_1 \lambda_1^t + v_{12} \kappa_2 \lambda_2^t \\ \hat{c}_t & = v_{21} \kappa_1 \lambda_1^t + v_{22} \kappa_2 \lambda_2^t \end{cases}$$

where κ_i are constants depending on the initial state, λ_i are the eigenvalues of A, and $\binom{v_{11}}{v_{21}}$, $\binom{v_{12}}{v_{22}}$ are the eigenvectors associated with λ_i. It is possible to show (see appendix 1) that $\lambda_1 \in (0, 1)$ and $\lambda_2 > 1$. The analytical proof provided in the appendix is important because it clearly illustrates how we have to modify the present neoclassical model in order to make indeterminacy arise: in particular, you will see that a critical step in the proof is based on the assumption that $y''(k) > 0$. The only case excluding an explosive behavior of the system (which would contradict that (a) (c, k) is a stationary point, as maintained previously, and (b) the optimality of the trajectories) thus emerges when we lock the initial state $(\hat{k}_0, \hat{c}_0)$ in such a way that $\kappa_2 = 0$; this reduces to solving the following system:

$$\begin{aligned} \hat{k}_0 & = v_{11} \kappa_1 \\ \hat{c}_0 & = v_{21} \kappa_1 \end{aligned}$$

which yields $\frac{\hat{c}_0}{\hat{k}_0} = \frac{v_{21}}{v_{11}}$. In fact it is possible to show the converse (i.e. $\frac{\hat{c}_0}{\hat{k}_0} = \frac{v_{21}}{v_{11}} \Rightarrow \kappa_2 = 0$, see appendix 1). Hence, the set of initial points ensuring stability must lie on the line $c_0 = c + \frac{v_{21}}{v_{11}}(k_0 - k)$.

Since k is a predetermined variable, there exists only one value of c_0 that ensures a nonexplosive behavior of $(k, c)_t$ near (k, c). In figure 3.1, k_* is defined as the solution of $1 = y'(k_*) \Leftrightarrow k_* = (y')^{-1}[1]$, and $k = (y')^{-1}[\beta^{-1}]$.

The linear approximation approach described above has not to be taken literally in applied work. Here we are going to provide an example in which the dynamics can be rather different

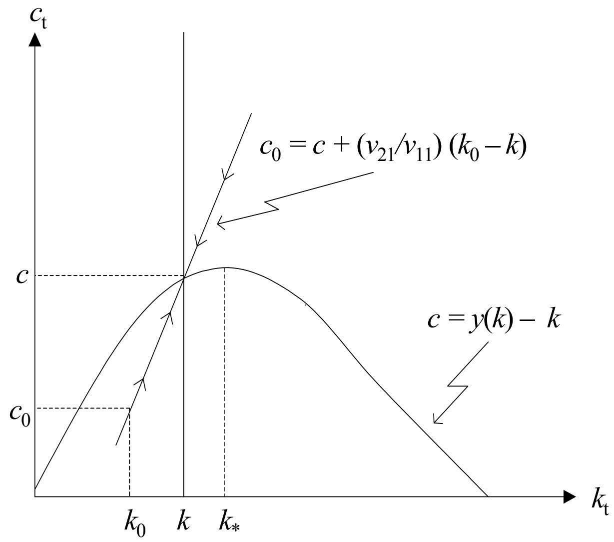


FIGURE 3.2.

when they start far from the stationary state. We take $y(k) = k^\gamma$, $u(c) = \log c$. By using the Bellman's equation approach described in the next section, one finds that the exact solution is:

$$\begin{cases} c_t &= (1-\beta\gamma)\,k_t^\gamma \\ k_{t+1} &= \beta\gamma k_t^\gamma \end{cases} \tag{3.15}$$

Figure 3.2 depicts the nonlinear manifold associated with the previous system. It also depicts its linear approximation obtained by linearizing the corresponding Euler equation. As an example, if $\beta = 0.99$ and $\gamma = 0.3$, we obtain that in order to be on the (linear) saddlepath it must be the case that

$$c_t = c + \frac{v_{21}}{v_{11}}\,(k_t - k)\,,$$

where $v_{21}/\,v_{11} \simeq 0.7101$, and

$$\begin{cases} c &= (1-\gamma\beta)\,k^\gamma \\ k &= \gamma\beta^{1/(1-\gamma)} \end{cases}$$

Clearly, having such a system is not enough to generate dynamics. It is also necessary to derive the dynamics of k_t. This is accomplished as follows:

$$\hat{k}_t = v_{11}\kappa_1\lambda_1^t = \lambda_1 v_{11}\kappa_1\lambda_1^{t-1} = \lambda_1\hat{k}_{t-1},$$

where $\lambda_1 = 0.3$. Hence we have:

$$\begin{cases} k_t &= k + \lambda_1\,(k_{t-1} - k) \\ c_t &= c + \dfrac{v_{21}}{v_{11}}\,(k_t - k) \end{cases}$$

Naturally, the previous linear system coincides with the one that can be obtained by linearizing the nonlinear system (3.?):

$$\begin{cases} k_{t+1} &= \beta\gamma k^\gamma + \gamma\cdot\beta\gamma k^{\gamma-1} = (k_t - k)\,k + \gamma\,(k_t - k) \\ c_t &= (1-\beta\gamma)\,k^\gamma + \gamma\,(1-\beta\gamma)\,k^{\gamma-1}\,(k_t-k) = c + \dfrac{1-\beta\gamma}{\beta}\,(k_t - k) \end{cases}$$

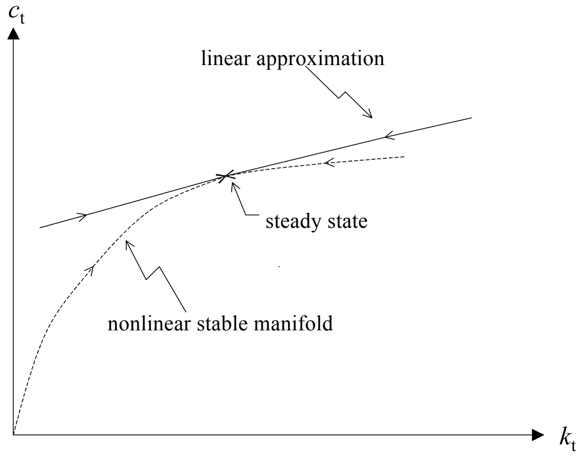


FIGURE 3.3.

where we used the solutions for (k, c). The claim follows because $\beta^{-1}(1 - \beta\gamma) \simeq 0.7101$.

A word on stability. It is often claimed (e.g. Azariadis p. 73) that there is a big deal of instability in systems such as this one because ∃! nonlinear stable manifold as in figure 3.2. The remaining portions of the state space lead to instability. Clearly the issue is true. However, it's also trivial. This is so because the "unstable" portions of the state-space are what they are because the transversality condition is not respected there, so one is forced to analyze the system by finding the manifold which also satisfies the transversality condition (i.e., the *stable* manifold). However, as shown for instance in Stockey and Lucas p. 135, when a problem like this one is solved via dynamic programming, the conditions for system's stability are *always* automatically satisfied: the sequence of capital and consumption is always well defined and convergent. This is so because dynamic programming is implying a transversality condition that is automatically fulfilled (otherwise the value function would be infinite or nil).

3.4.4 Stochastic economies

> "Real business cycle theory is the application of general equilibrium theory to the quantitative analysis of business cycle fluctuations."Edward Prescott, 1991, p. 3, "Real Business Cycle Theory: What have we learned ?," *Revista de Analisis Economico*, Vol. 6, No. 2, November, pp. 3-19.

> "The Kydland and Prescott model is a complete markets set-up, in which equilibrium and optimal allocations are equivalent. When it was introduced, it seemed to many—myself included—to be much too narrow a framework to be useful in thinking about cyclical issues." Robert E. Lucas, 1994, p. 184, "Money and Macroeconomics," in: *General Equilibrium 40th Anniversary Conference*, CORE DP # 9482, pp. 184–187.

In its simplest version, the real business cycles theory is an extension of the neoclassical growth model in which random productivity shocks are juxtaposed with the basic deterministic model. The ambition is to explain macroeconomic fluctuations by means of random shocks of real nature uniquely. This is in contrast with the Lucas models of the 70s, in which the engine of fluctuations was attributable to the existence of information delays with which agents discover the nature of a shock (real or monetary), and/or the quality of signals (noninformative equilibria, or "noisy" rational expectations, according to the terminology of Grossman). Nevertheless, the research methodology is the same here. Macroeconomic fluctuations are going to be generated by optimal responses of the economy with respect to metasystematic shocks: agents implement action plans that are state-contingent, i.e. they decide to consume, work and invest according to the history of shocks as well as the present shocks they observe. On a technical standpoint, the solution of the model is only possible if agents have rational expectations.

3.4.4.1 Basic model

We consider an economy with complete markets, and without any form of frictions. The resulting equilibrium allocations are Pareto-optimal. Fot this reason, these equilibrium allocations can be analyzed by "centralizing" the economy. On a strictly technical standpoint, we may write a single constraint. So the planner's problem is,

$$V(k_0, s_0) = \max_{\{c,\theta\}} E\left[\sum_{t=0}^{\infty} \beta^t u(c_t)\right], \tag{3.16}$$

subject to some capital accumulation constraint. We develop a capital accumulation constraint with capital depreciation. To explain this issue, consider the following definition:

$$\text{Invest}_t \equiv I_t = K_{t+1} - (1-\delta)\, K_t. \tag{3.17}$$

The previous identity follows by the following considerations. At time $t-1$, the firm chooses a capital level K_t. At time t, a portion δK_t is lost due to (capital) depreciation. Hence at time t, the firm is left with $(1-\delta)\, K_t$. Therefore, the capital chosen by the firm at time t, K_{t+1}, equals the capital the firm disposes already of, $(1-\delta)\, K_t$, plus new investments, which is exactly eq. (3.17).

Next, assume that $K_t = k_t$ (i.e. population normalized to one), and consider the equilibrium condition,

$$\tilde{y}\,(k_t, \epsilon_t) = c_t + \text{Invest}_t,$$

where $\tilde{y}(k_t, s_t)$ is the production function, $\mathcal{F}_t$-measurable, and s is the "metasystemic" source of randomness. By replacing eq. (3.17) into the previous equilibrium condition,

$$k_{t+1} = \tilde{y}\,(k_t, \epsilon_t) - c_t + (1-\delta)\, k_t. \tag{3.18}$$

So the program is to mazimize the utility in eq. (3.16) under the capital accumulation constraint in eq. (3.18).

We assume that $\tilde{y}\,(k_t, s_t) \equiv s_t y\,(k_t)$, where y is as in section 3.1, and $(s_t)_{t=0}^{\infty}$ is solution to

$$s_{t+1} = s_t^{\rho} \epsilon_{t+1}, \quad \rho \in (0,1), \tag{3.19}$$

where $(\epsilon_t)_{t=0}^{\infty}$ is a i.i.d. sequence with support s.t. $s_t \geq 0$.

In this economy, every asset is priced as in the Lucas model of the previous section. Therefore, the gross return on savings $s_{.}y'(k_{.})$ satisfies:

$$u'\,(c_t) = \beta E_t \left\{u'\,(c_{t+1})\left[s_{t+1} y'\,(k_{t+1}) + 1 - \delta\right]\right\}.$$

A rational expectation equilibrium is thus a statistical sequence $\{(c_t(\omega), k_t(\omega))_t\}_{t=0}^{\infty}$ (i.e., ω by ω) satisfying the constraint of program (3.20), (3.21) and

$$\begin{cases} u'\,(c\,(s,k)) &= \beta \int u'\,(c\,(s^+, k^+))\,[s^+ y'(k^+) + 1 - \delta]\, dP\,(s^+ |\, s) \\ k^+ &= sy\,(k) - c\,(s,k) + (1-\delta)\, k \end{cases} \tag{3.20}$$

where k_0 and s_0 are given.

Implicit in (3.22) is the assumption of rational expectations. Consumption is a function $c_t = c(s_t, k_t)$ that represents an "optimal response" of the economy with respect to exogenous shocks. As is clear, rational expectations and "optimal responses" go together with the *dynamic programming principle*.

We now show the existence of a sort of a saddlepoint path, which implies determinacy of the stochastic equilibrium.[6] We study the behavior of $(c, k, s)_t$ in a neighborhood of $\epsilon \equiv E(\epsilon_t)$. Let (c, k, s) be the values of consumption, capital and shocks arising at ϵ. These (c, k, s) are obtained by replacing them into the constraint of (3.20), (3.21) and (3.22), leaving,

$$\begin{cases} \beta &= \dfrac{1}{sy'(k) + 1 - \delta} \\ c &= sy(k) - \delta k \\ s &= \epsilon^{\frac{1}{1-\rho}} \end{cases}$$

A first-order Taylor approximation of the l.h.s. of (3.22) and a first-order Taylor approximation of the r.h.s. of (3.22) with respect to (k, c, s) yields:

$$\begin{aligned} & u''(c)(c_t - c) \\ & = \beta E_t \left[u''(c) \left(sy'(k) + 1 - \delta \right) (c_{t+1} - c) + u'(c) sy''(k) (k_{t+1} - k) + u'(c) y'(k) (s_{t+1} - s) \right]. \end{aligned} \tag{3.21}$$

By repeating the same procedure as regards the capital propagation equation and the productivity shocks propagation equation we get:

$$\begin{aligned} \hat{k}_{t+1} &= \beta^{-1} \hat{k}_t + \frac{sy(k)}{k} \hat{s}_t - \frac{c}{k} \hat{c}_t \\ \hat{s}_{t+1} &= \rho \hat{s}_t + \hat{\epsilon}_{t+1} \end{aligned} \tag{3.22}$$

where we have defined $\hat{x}_t \equiv \frac{x_t - x}{x}$.

The system (3.23)-(3.24) can be written as,

$$\hat{z}_{t+1} = \Phi \hat{z}_t + R u_{t+1}, \tag{3.23}$$

where $\hat{z}_t = (\hat{k}_t, \hat{c}_t, \hat{s}_t)^T$, $u_t = (u_{c,t}, u_{s,t})^T$, $u_{c,t} = \hat{c}_t - E_{t-1}(\hat{c}_t)$, $u_{s,t} = \hat{s}_t - E_{t-1}(\hat{s}_t) = \hat{\epsilon}_t$,

$$\Phi = \begin{pmatrix} \beta^{-1} & -\frac{c}{k} & s\frac{y(k)}{k} \\ -\frac{u'(c)}{cu''(c)} sky''(k) & 1 + \frac{\beta u'(c)}{u''(c)} sy''(k) & -\frac{\beta u'(c)}{cu''(c)} s(sy(k)y''(k) + \rho y'(k)) \\ 0 & 0 & \rho \end{pmatrix},$$

and

$$R = \begin{pmatrix} 0 & 0 \\ 1 & 0 \\ 0 & 1 \end{pmatrix}.$$

[6] Strictly speaking, the determination of a stochastic equilibrium is the situation in which one shows that there is one and only one stationary measure (definition: $p(+) = \int \pi(+/-) dp(-)$, where π is the transition measure) generating $(c_t, k_t)_{t=1}^{\infty}$. It is also known that the standard linearization practice has not a sound theoretical foundation. (See Brock and Mirman (1972) for the theoretical study of the general case.) The general nonlinear case can be numerically dealt with the tools surveyed by Judd (1998).

Let us consider the characteristic equation:

$$0 = \det(\Phi - \lambda I) = (\rho - \lambda)\left[\lambda^2 - \left(\beta^{-1} + 1 + \beta\frac{u'(c)}{u''(c)}sy''(k)\right)\lambda + \beta^{-1}\right].$$

A solution is $\lambda_1 = \rho$. By repeating the same procedure as in the deterministic case (see appendix 1), one finds that $\lambda_2 \in (0,1)$ and $\lambda_3 > 1$.[7] By using the same approach as in the deterministic case, we "diagonalize" the system by rewriting $\Phi = P\Lambda P^{-1}$, where Λ is a diagonal matrix that has the eigenvalues of Φ on the diagonal, and P is a matrix of the eigenvectors associated to the roots of Φ. The system can thus be rewritten as:

$$\hat{y}_{t+1} = \Lambda\hat{y}_t + w_{t+1},$$

where $\hat{y}_t \equiv P^{-1}\hat{z}_t$ and $w_t \equiv P^{-1}Ru_t$. The third equation is:

$$\hat{y}_{3,t+1} = \lambda_3\hat{y}_{3t} + w_{3,t+1},$$

and $\hat{y}_3$ exploses unless

$$\hat{y}_{3t} = 0 \quad \forall t,$$

which would also require that $w_{3t} = 0\ \forall t$. Another way to see that is the following one. Since $\hat{y}_{3t} = \lambda_3^{-1}E_t(\hat{y}_{3,t+1})$, then:

$$\hat{y}_{3t} = \lambda_3^{-T}E_t(\hat{y}_{3,t+T}) \qquad \forall T.$$

Because $\lambda_3 > 1$, this relationship is satisfied only for

$$\hat{y}_{3t} = 0\ \forall t, \tag{3.24}$$

which also requires that:

$$w_{3t} = 0\ \forall t. \tag{3.25}$$

Let us analyze condition (3.26). We have $\hat{y}_t = P^{-1}\hat{z}_t \equiv \Pi\hat{z}_t$, whence:

$$0 = \hat{y}_{3t} = \pi_{31}\hat{k}_t + \pi_{32}\hat{c}_t + \pi_{33}\hat{s}_t. \tag{3.26}$$

This means that the three state variables are mutually linked by a proportionality relationship. More precisely, we are redescovering in dimension 3 the saddlepoint of the economy:

$$\mathcal{S} = \left\{x \in \mathcal{O} \subset \mathbb{R}^3 /\ \pi_{3.}x = 0\right\},\ \pi_{3.} = (\pi_{31}, \pi_{32}, \pi_{33}).$$

In addition, condition (3.27) implies a linear relationship between the two errors which we found to be after some simple calculations:

$$u_{ct} = -\frac{\pi_{33}}{\pi_{32}}u_{st}\ \forall t \qquad (\text{"no sunspots"}).$$

[7] There is a slight difference in the formulation of the model of this section and the deterministic model of section 4.? because the state variables are expressed in terms of growth rates here. However, one can always reformulate this model in terms of the other one by pre- and post- multiplying Φ by appropriate normalizing matrices. As an example, if G i the 3×3 matrix that has $\frac{1}{k}$, $\frac{1}{c}$ and $\frac{1}{s}$ on its diagonal, system (4.25) can be written as: $E(z_{t+1} - z) = G^{-1}\Phi G \cdot (z_t - z)$, where $z_t = (k_t, c_t, s_t)$. In this case, one would have: $G^{-1}\Phi G = \begin{pmatrix} \beta^{-1} & -1 & y \\ -\frac{u'}{u''}sy'' & 1 + \frac{\beta u'}{u''}sy'' & -\frac{\beta u'}{u''}(sy''y + \rho y') \\ 0 & 0 & \rho \end{pmatrix}$, where $\det(G^{-1}\Phi G - \lambda I) = (\rho - \lambda)(\lambda^2 - (\beta^{-1} + 1 + \beta\frac{u'(c)}{u''(c)}sy''(k))\lambda + \beta^{-1})$, and the conclusions are exactly the same. In this case, we clearly see that the previous system collapses to the deterministic one when $\epsilon_t = 1, \forall t$ and $s_0 = 1$.

Which is the implication of relation (3.28)? It is very similar to the one described in the deterministic case. Here the convergent subspace is the plan $\mathcal{S}$, and $\dim(\mathcal{S}) = 2 =$ number of roots with modulus less than one. Therefore, with 2 predetermined variables (i.e. $\hat{k}_0$ and $\hat{s}_0$), there exists one and only one "jump value" of $\hat{c}_0$ in $\mathcal{S}$ which ensures stability, and this is $\hat{c}_0 = -\frac{\pi_{31}\hat{k}_0+\pi_{33}\hat{s}_0}{\pi_{32}}$.

As is clear, there is a deep analogy between the deterministic and stochastic cases. This is not fortuitous, since the methodology used to derive relation (3.28) is exactly the same as the one used to treat the deterministic case. Let us elaborate further on this analogy, and compute the solution of the model. The solution $\hat{y}$ takes the following form,

$$\hat{y}_t^{(i)} = \lambda_i^t \hat{y}_0^{(i)} + \zeta_t^{(i)}, \quad \zeta_t^{(i)} \equiv \sum_{j=0}^{t-1} \lambda_i^j w_{i,t-j},$$

and the solution for $\hat{z}$ has the form,

$$\hat{z}_t = P\hat{y}_t = (v_1\ v_2\ v_3)\hat{y}_t = \sum_{i=1}^{3} v_i \hat{y}_t^{(i)} = \sum_{i=1}^{3} v_i \hat{y}_0^{(i)} \lambda_i^t + \sum_{i=1}^{3} v_i \zeta_t^{(i)}.$$

To pin down the components of $\hat{y}_0$, observe that $\hat{z}_0 = P\hat{y}_0 \Rightarrow \hat{y}_0 = P^{-1}\hat{z}_0 \equiv \Pi\hat{z}_0$. The stability condition then imposes that the state variables $\in \mathcal{S}$, or $\hat{y}_0^{(3)} = 0$, which is exactly what was established before.

How to implement the solution in practice ? First, notice that:

$$\hat{z}_t = v_1\lambda_1^t\hat{y}_0^{(1)} + v_2\lambda_2^t\hat{y}_0^{(2)} + v_3\lambda_3^t\hat{y}_0^{(3)} + v_1\zeta_t^{(1)} + v_2\zeta_t^{(2)} + v_3\zeta_t^{(3)}.$$

Second note that the terms $v_3\lambda_3^t\hat{y}_0^{(3)}$ and $v_3\zeta_t^{(3)}$ are killed because $\hat{y}_0^{(3)} = 0$, as argued before. Furthermore, $\zeta_t^{(i)} = \sum_{j=0}^{t-1} \lambda_i^j w_{i,t-j}$, and in particular

$$\zeta_t^{(3)} = \sum_{j=0}^{t-1} \lambda_3^j w_{3,t-j}.$$

But $w_{3,t} = 0\ \forall t$ as argued before (see eq. (3.??)), and then $\zeta_t^{(3)} = 0$, $\forall t$.

Therefore, the solution to implement in practice is:

$$\hat{z}_t = v_1\lambda_1^t\hat{y}_0^{(1)} + v_2\lambda_2^t\hat{y}_0^{(2)} + v_1\zeta_t^{(1)} + v_2\zeta_t^{(2)}.$$

3.4.4.2 A digression: indeterminacy and sunspots

The neoclassic deterministic growth model exhibits determinacy of the equilibrium. This is also the case of the basic version of the real business cycles model. Determinacy is due to the fact that the number of predetermined variables is equal to the dimension of the convergent subspace.

If we were only able to build up models in which we managed to increase the dimension of the converging subspace, we would thus be introducing *indeterminacy* phenomena (see appendix 1). In addition, it turns out that this circumstance would be coupled with the the existence of *sunspots* that we would be able to identify as "expectation shocks". Such an approach has been initiated in a series of articles by Farmer.[8] Let us mention that on an econometric standpoint, the basic Real Business Cycle (RBC, henceforth) model is rejected with high probabilities; this has unambigously been shown by Watson (1993)[9] by means of very fine spectral methods. Now it seems that the performance of models generating sunspots is higher than the one of the basic RBC models; much is left to be done to achieve definitive results in this direction, however.[10] The main reasons explaining failure of the basic RBC models are that they offer little space to propagation mechanisms. Essentially, the properties of macroeconomic fluctuations is inherited by the ones of the productivity shocks, which in turn are imposed by the modeler. In other terms, the system dynamics is driven mainly by an impulse mechanism. Empirical evidence suggests that an impulse-propagation mechanism should be more appropriate to model actual macroeconomic fluctuations: this is, after all, the main message contained in the pioneer work of Frisch in the 1930s.

It is impossible to introduce inderminacy and sunspots phenomena in the economies that we considered up to here. As originally shown by Cass, a Pareto-optimal economy can not display sunspots. On the contrary, any market imperfection has the potential to become a source of sunspots; an important example of this kind of link is the one between incomplete markets and sunspots. Therefore, the only way to introduce the possibility of sunspots in a growth model is to consider mild deviations from optimality (such as imperfect competition and/or externality effects). To better grasp the nature of such a phenomenon, it is useful to analyze why such a phenomenon can not take place in the case of the models in the previous sections.

Let us begin with the deterministic model of section 3.2. We wish to examine whether it is possible to generate "stochastic outcomes" without introduicing any shock on the fundamentals of the economy - as in the RBC models, in which the shocks on the fundamentals are represented by the productivity shocks. In other terms, we wish to examine whether there exist conditions under which the deterministic neoclassic model can generate a system dynamics driven uniquely by "shocks" that are totally extraneous to the economy's fundamentals. If this is effectivley the case, we must imagine that the optimal trajectory of consumption, for instance, is generated by a controlled stochastic process, even in the absence of shocks on the fundamentals. In this case, the program 3.6 is recast in an expectation format, and all results of section 3.2 remain the same, with the importan exception that eq. (3.7) is written under an expectation format:

$$E_t \begin{pmatrix} \hat{k}_{t+1} \\ \hat{c}_{t+1} \end{pmatrix} = A \begin{pmatrix} \hat{k}_t \\ \hat{c}_t \end{pmatrix}.$$

Then we introduce the expectation error process $u_{c,t} \equiv \hat{c}_t - E_{t-1}(\hat{c}_t)$, and plug it into the previous system to obtain:

$$\begin{pmatrix} \hat{k}_{t+1} \\ \hat{c}_{t+1} \end{pmatrix} = A \begin{pmatrix} \hat{k}_t \\ \hat{c}_t \end{pmatrix} + \begin{pmatrix} 0 \\ u_{c,t+1} \end{pmatrix}.$$

Such a thought-experiment did not change A and as in section 3.2, we still have $\lambda_1 \in (0,1)$ and $\lambda_2 > 1$. This implies that when we decompose A as $P\Lambda P^{-1}$ to write

$$\hat{y}_{t+1} = \Lambda \hat{y}_t + P^{-1}(0 \quad u_{c,t+1})^T,$$

[8] See Farmer,R. (1993): *The Macroeconomics of Self-Fulfilling Prophecies*, The MIT Press, for an introduction to sunspots in the kind of models considered in this section.

[9] Watson, M. (1993): "Measures of Fit for Calibrated Models," *Journ. Pol. Econ.*, 101, pp.1011-1041.

[10] An important problem pointed out by Kamihigashi (1996) is the observational equivalence of the two kinds of models: Kamihigashi, T. (1996): "Real Business Cycles and Sunspot Fluctuations are Observationally Equivalent," *J. Mon. Econ.*, 37, 105-117.

we must also have that for $\hat{y}_{2t} = \lambda_2^{-T} E_t(\hat{y}_{2,t+T})$ $\forall T$ to hold, the necessary condition is $\hat{y}_{2t} = 0$ $\forall t$, which implies that the second element of $P^{-1}(0 \quad u_{c,t+1})^T$ is zero:

$$0 = \pi_{22} u_{c,t} \ \forall t \Leftrightarrow 0 = u_{c,t} \ \forall t.$$

There is no role for expectation errors here: there can not be sunspots in the neoclassic growth model. This is due to the fact that $\lambda_2 > 1$. The existence of a saddlepoint equilibrium is due to the classical restrictions imposed on functions u and y. Now we wish to study how to modify these functions to change the typology of solutions for the eigenvalues of A.

3.5 Production based asset pricing

We assume complete markets.

3.5.1 Firms

The capital accumulation of the firm does always satisfy the identity in eq. (3.17),

$$K_{t+1} = (1-\delta) K_t + I_t, \tag{3.27}$$

but we now make the assumption that capital adjustment is costly. Precisely, the real profit (or dividend) of the firm as of at time t is given by,

$$D_t \equiv \tilde{y}_t - w_t N_t - I_t - \phi\left(\frac{I_t}{K_t}\right) K_t,$$

where w_t is the real wage, N_t is the employment level, and $\tilde{y}_t$ is the firm's production - that we take to be stochastic here. The function ϕ is the adjustment-cost function. It satisfies $\phi \geq 0$, $\phi' \geq 0$, $\phi'' \geq 0$. This formulation says that capital adjustment is very costly when the adjustment is done fastly. Naturally, ϕ is zero in the absence of adjustment costs. We assume that $\tilde{y}_t = \tilde{y}(K_t, N_t)$.

Next we evaluate the *value* of this profit from the perspective of time zero. We make use of Arrow-Debreu state prices introduced in the previous chapter. At time t, and in state s, the profit $D_t^{(s)}$ (say) is worth,

$$\phi_{0,t}^{(s)} D_t^{(s)} = m_{0,t}^{(s)} D_t^{(s)} P_{0,t}^{(s)},$$

with the same notation as in chapter 2.

3.5.1.1 The value of the firm

We assume that $\tilde{y}_t = \tilde{y}(K_t, N_t)$. The the *cum-dividend* value of the firm is,

$$V^c(K_0, N_0) = D_0 + E\left[\sum_{t=1}^{\infty} m_{0,t} D_t\right]. \tag{3.28}$$

We assume that the firm maximizes its value in eq. (3.28) subject to the capital accumulation in eq. (3.27). That is,

$$V^c(K_0, N_0) = \max_{\{K,N,I\}} \left[D(K_0, N_0, I_0) + E\left(\sum_{t=1}^{\infty} m_{0,t} D(K_t, N_t, I_t)\right)\right],$$

where

$$D(K_t, N_t, I_t) = \tilde{y}(K_t, N_t) - w_t N_t - I_t - \phi\left(\frac{I_t}{K_t}\right) K_t.$$

The value of the firm can recursively be found through the Bellman's equation,

$$V^c(K_t, N_t) = \max_{N_t, I_t} \left[D(K_t, N_t, I_t) + E\left(m_{t+1} V^c(K_{t+1}, N_{t+1})\right) \right], \tag{3.29}$$

where the expectation is taken with respect to the information set as of time t.

The first order conditions for N_t are,

$$\tilde{y}_N(K_t, N_t) = w_t.$$

The first order conditions for I_t lead to,

$$0 = D_I(K_t, N_t, I_t) + E\left[m_{t+1} V_K^c(K_{t+1}, N_{t+1})\right]. \tag{3.30}$$

Optimal investment I is thus a function $I(K_t, N_t)$. The value of the firm thus satisfies,

$$V^c(K_t, N_t) = D(K_t, N_t, I(K_t, N_t)) + E\left[m_{t+1} V^c(K_{t+1}, N_{t+1})\right].$$

We need to find the envelop condition. We have,

$$\begin{aligned} V_K^c(K_t, N_t) &= D_K(K_t, N_t, I(K_t, N_t)) + D_I(K_t, N_t, I(K_t, N_t)) I_K(K_t, N_t) \\ &\quad + E\left\{m_{t+1} V_K^c(K_{t+1}, N_{t+1}) \left[(1-\delta) + I_K(K_t, N_t)\right]\right\} \\ &= D_K(K_t, N_t, I(K_t, N_t)) + (1-\delta) E\left[m_{t+1} V_K^c(K_{t+1}, N_{t+1})\right] \\ &= D_K(K_t, N_t, I(K_t, N_t)) - (1-\delta) D_I(K_t, N_t, I(K_t, N_t)). \end{aligned} \tag{3.31}$$

where the second and third lines follow by the optimality condition (3.30), and

$$\begin{aligned} D_K(K_t, N_t, I(K_t, N_t)) &\equiv \tilde{y}_K(K_t, N_t) - \bar{\phi}_K(K_t, I(K_t, N_t)) \\ -D_I(K_t, N_t, I(K_t, N_t)) &\equiv 1 + \phi'\left(\frac{I_t}{K_t}\right) \end{aligned}$$

and $\bar{\phi}_K(K_t, I(K_t, N_t)) = \phi\left(\frac{I_t}{K_t}\right) - \phi'\left(\frac{I_t}{K_t}\right)\frac{I_t}{K_t}$.

3.5.1.2 q theory

We now introduce the notion of Tobin's qs. We have the following definition,

$$\text{Tobin's marginal } q_t = E\left[m_{t+1} V_K^c(K_{t+1}, N_{t+1})\right]. \tag{3.32}$$

We claim that q is the shadow price of installed capital. To show this, consider the Lagrangean,

$$\mathcal{L}(K_t, N_t) = D(K_t, N_t, I_t) + E\left(m_{t+1} V^c(K_{t+1}, N_{t+1})\right) - q_t\left(K_{t+1} - (1-\delta) K_t - I_t\right).$$

The first order condition for I_t is,

$$0 = D_I(K_t, N_t, I_t) + q_t.$$

So by comparing with the optimality condition (3.30) we get back to the definition in (3.32). Clearly, we also have,

$$V_K^c(K_t, N_t) = D_K(K_t, N_t, I(K_t, N_t)) + (1-\delta) q_t.$$

Therefore, by eq. (3.32) and the previous equality,

$$q_t = E\left[m_{t+1}\left(\tilde{y}_K(K_{t+1}, N_{t+1}) - \bar{\phi}_K(K_{t+1}, I(K_{t+1}, N_{t+1})) + (1-\delta) q_{t+1}\right)\right]. \tag{3.33}$$

So to sum-up,

$$\begin{aligned} q_t &= 1 + \phi'\left(\frac{I_t}{K_t}\right) \\ q_t &= E\left[m_{t+1}\left(\tilde{y}_K(K_{t+1}, N_{t+1}) - \phi\left(\frac{I_{t+1}}{K_{t+1}}\right) + \phi'\left(\frac{I_{t+1}}{K_{t+1}}\right)\frac{I_{t+1}}{K_{t+1}} + (1-\delta) q_{t+1}\right)\right] \end{aligned}$$

For example, suppose that there are no adjustement costs, i.e. $\phi(x) = 0$. Then $q_t = 1$. By plugging $q_t = 1$ in eq. (3.33), we obtain the usual condition,

$$1 = \underset{t-1}{E}\left\{m_t\left[\tilde{y}_K(K_t, N_t) + (1-\delta)\right]\right\}.$$

But empirically, the return $\tilde{y}_K(K_t, N_t) + (1-\delta)$ is a couple of orders of magnitude less than actual stock-market returns. So we need adjustment costs - or other stories.

The previous computations make sense. We have,

$$V^c(K_t, N_t) - D(K_t, N_t, I(K_t, N_t)) = E\left[m_{t+1} V^c(K_{t+1}, N_{t+1})\right] =^{def} V(K_t, N_t).$$

Now by definition, the LHS is the ex-dividend value of the firm, and so must be the RHS.

Finally, we also have,

$$q_t = E\left[\sum_{s=1}^{\infty}(1-\delta)^{s-1} m_{0,t+s}\left(\tilde{y}_K(K_{t+s}, N_{t+s}) - \bar{\phi}_K(K_{t+s}, I_{t+s})\right)\right].$$

Hence, Tobin's marginal q is the firm's marginal value of a further unit of capital invested in the firm, and it is worth the discounted sum of all future marginal net productivity.

We now show an important result - originally obtained by Hayahsi (1982) in continuous time.

Theorem. *Marginal q and average q coincide. That is, we have,*

$$V(K_t, N_t) = q_t K_{t+1}$$

Proof. By the optimality condition (3.30), and eq. (3.31),

$$-D_I(K_t, N_t, I_t) = E\left[m_{t+1}\left(D_K(K_{t+1}, N_{t+1}, I_{t+1}) - (1-\delta) D_I(K_{t+1}, N_{t+1}, I_{t+1})\right)\right]. \tag{3.34}$$

And by various homogeneity properties assumed so far, and $w_t = \tilde{y}_N(K_t, N_t)$,

$$D(K_t, N_t, I_t) = D_K(K_t, N_t, I_t) K_t + D_I(K_t, N_t, I_t) I_t.$$

Hence,

$$V_0 = E\left[\sum_{t=1}^{\infty} m_{0,t} D_t\right] = E\left[\sum_{t=1}^{\infty} m_{0,t}\left(D_{K,t} - (1-\delta) D_{I,t}\right) K_t\right] + E\left[\sum_{t=1}^{\infty} m_{0,t} D_{I,t} K_{t+1}\right].$$

where the second line follows by eq. (3.27). By eq. (3.34), and the law of iterated expectations,

$$E\left[\sum_{t=1}^{\infty} m_{0,t}\left(D_{K,t} - (1-\delta) D_{I,t}\right) K_t\right] = -D_{I,0} K_1 - E\left[\sum_{t=1}^{\infty} m_{0,t} K_{t+1} D_{I,t}\right].$$

Hence, $V_0 = -D_{I,0}K_1$, and the proof follows by eq. (3.30) and the definition of q. ‖

Classical references on these issues are Abel (1990), Christiano and Fisher (1995), Cochrane (1991), and Hayashi (1982).[11]

3.5.2 Consumers

In a deterministic world, eq. (3.8) holds,

$$V_0 = c_0 + \sum_{t=1}^{\infty} \frac{c_t - w_t N_t}{\prod_{i=1}^{t} R_i}.$$

We now show that in the uncertainty case, the relevant budget constraint is,

$$V_0 = c_0 + E\left[\sum_{t=1}^{\infty} m_{0,t}\left(c_t - w_t N_t\right)\right]. \tag{3.35}$$

Indeed,

$$c_t + q_t \theta_{t+1} = \left(q_t + D_t\right)\theta_t + w_t N_t.$$

We have,

$$\begin{aligned}
E\left[\sum_{t=1}^{\infty} m_{0,t}\left(c_t - w_t N_t\right)\right] &= E\left[\sum_{t=1}^{\infty} m_{0,t}\left(q_t + D_t\right)\theta_t\right] - E\left[\sum_{t=1}^{\infty} m_{0,t} q_t \theta_{t+1}\right] \\
&= E\{\sum_{t=1}^{\infty} \underset{t-1}{E}[\underbrace{\frac{m_{0,t}}{m_{t-1,t}}}_{=m_{0,t-1}} m_{t-1,t}\left(q_t + D_t\right)\theta_t]\} - E\left[\sum_{t=2}^{\infty} m_{0,t-1} q_{t-1} \theta_t\right] \\
&= E\left[\sum_{t=1}^{\infty} m_{0,t-1} q_{t-1} \theta_t\right] - E\left[\sum_{t=2}^{\infty} m_{0,t-1} q_{t-1} \theta_t\right] \\
&= q_0 \theta_1 = V_0 - c_0.
\end{aligned}$$

[11]References: Abel, A. (1990). "Consumption and Investment," Chapter 14 in Friedman, B. and F. Hahn (eds.): *Handbook of Monetary Economics*, North-Holland, 725-778. Christiano, L. J. and J. Fisher (1998). "Stock Market and Investment Good Prices: Implications for Macroeconomics," unpublished manuscript. Cochrane (1991). "Production-Based Asset Pricing and the Link Between Stock Returns and Economic Fluctuations," *Journal of Finance* 46, 207-234. Hayashi, F. (1982). "Tobin's Marginal q and Average q: A Neoclassical Interpretation," *Econometrica* 50, 213-224.

where the third line follows by the property of the discount factor $m_t \equiv m_{t-1,t}$. So, many consumers maximize

$$U = E\left[\sum_{t=0}^{\infty} \beta^t u(c_t)\right]$$

under the budget constraint in eq. (3.35). Optimality conditions are fairly easy to derive. We have,

$$\begin{array}{lll} m_{t+1} & = \beta \dfrac{u_1(c_{t+1}, N_{t+1})}{u_1(c_t, N_t)} & (\textit{inter}\text{temporal optimality}) \\ w_t & = -\dfrac{u_2(c_t, N_t)}{u_1(c_t, N_t)} & (\textit{intra}\text{temporal optimality}) \end{array}$$

3.5.3 Equilibrium

For all t,

$$\tilde{y}_t = c_t + I_t, \quad \text{and } \theta_t = 1.$$

3.6 Money, asset prices, and overlapping generations models

3.6.1 Introductory examples

The population is constant and equal to two. The young agent solves the following problem:

$$\begin{array}{l} \max_{\{c\}} \left(u(c_{1t}) + \beta u(c_{2,t+1})\right) \\ \text{s.t.} \left\{ \begin{array}{lll} s_t + c_{1t} & = & w_{1t} \\ c_{2,t+1} & = & s_t R_{t+1} + w_{2,t+1} \end{array} \right. \end{array} \tag{3.36}$$

The previous constraints can be embedded in a single, lifetime budget constraint:

$$c_{2,t+1} = -R_{t+1} c_{1,t} + R_{t+1} w_{1t} + w_{2,t+1}. \tag{3.37}$$

The agent born at time $t-1$ faces the constraints: $s_{t-1} + c_{1,t-1} = w_{1,t-1}$ and $c_{2t} = s_{t-1} R_t + w_{2t}$; by combining her second period constraint with the first period constraint of the agent born at time t one obtains:

$$s_{t-1} R_t + w_t = s_t + c_{1t} + c_{2t}, \quad w_t \equiv w_{1t} + w_{2t} \tag{3.38}$$

The financial market equilibrium,

$$s_t = 0, \quad \forall t, \tag{3.39}$$

implies that the goods market is also in equilibrium: $w_t = \sum_{i=1}^{2} c_{i,t}\ \forall t$. Therefore, we can analyze the economic dynamics by focussing on the inside money equilibrium.

The first order condition requires that the slope of the indifference curve $(-\beta^{-1} \frac{u'(c_{1,t})}{u'(c_{2,t+1})})$ be equal to the slope of the lifetime budget constraint $(-R_{t+1})$, or:[12]

$$\beta \frac{u'(c_{2,t+1})}{u'(c_{1,t})} = \frac{1}{R_{t+1}}. \tag{3.40}$$

The equilibrium is then a sequence of gross returns satisfying (3.32) and (3.33), or:

$$b_t \equiv \frac{1}{R_{t+1}} = \beta \frac{u'(w_{2,t+1})}{u'(w_{1t})}.$$

[12] This result is qualitatively similar to the results presented in the previous section.

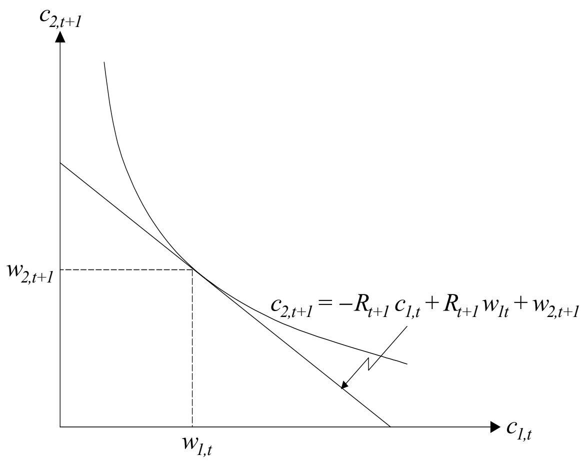


FIGURE 3.4.

b_t can be interpreted as a "shadow price" of a bond issued at t and promising 1 unit of numéraire at $t+1$. One may think of $\{b_t\}_{t=0,1,\cdots}$ as a sequence of prices (fixed by an auctioneer living forever) which does not incentivate agents to lend to or borrow from anyone (given that one can not lend to or borrow from anyone else), which justifies our term "shadow price"; cf. figure 3.3.

The situation is different if heterogeneous members are present within the same generation. In this case, everyone solves the same program as in (3.29):

$$\max_{\{c^j\}} \left(u_j(c_{1t}^{(j)}) + \beta_j u_j(c_{2,t+1}^{(j)}) \right)$$
$$\text{s.t.} \left\{ \begin{array}{lcl} s_t^{(j)} + c_{1t}^{(j)} & = & w_{1t}^{(j)} \\ c_{2,t+1}^{(j)} & = & s_t^{(j)} R_{t+1} + w_{2,t+1}^{(j)} \end{array} \right.$$

The first order condition is still:

$$\beta_j \frac{u_j'(c_{2,t+1}^{(j)})}{u'(c_{1t}^{(j)})} = \frac{1}{R_{t+1}} \equiv b_t, \; j = 1, \cdots, n; \; t = 0, 1, \cdots, \tag{3.41}$$

and an equilibrium is a sequence $\{b_t\}_{t=0,1,\cdots}$ satisfying (3.34) and

$$\sum_{j=1}^{n} s_t^{(j)} = 0, \; t = 0, 1, \cdots.$$

The equilibrium is a sequence of "generation by generation" equilibria. That is, a sequence of temporary equilibria.

Example: $u_j(x) = \log x$ and $\beta_j = \beta \; \forall j$.

The first order condition is:

$$\beta_j \frac{c_{1t}^{(j)}}{c_{2,t+1}^{(j)}} = \frac{1}{R_{t+1}} \equiv b_t,$$

and using the budget constraints:

$$\left\{ \begin{array}{lcll} c_{1t}^{(j)} & = & \frac{1}{1+\beta} \delta_j(R_{t+1}), & \delta_j(x) = w_{1t}^{(j)} + \frac{w_{2,t+1}^{(j)}}{x} \\ c_{2,t+1}^{(j)} & = & \frac{\beta}{1+\beta} R_{t+1} \delta_j(R_{t+1}) & \\ s_t^{(j)} & = & \frac{1}{1+\beta} (\beta w_{1t}^{(j)} - \frac{w_{2,t+1}^{(j)}}{R_{t+1}}) & \end{array} \right.$$

The equilibrium is $0 = \sum_{j=1}^{n} s_t^{(j)}$, whence

$$b_t = \frac{1}{R_{t+1}} = \frac{\beta \sum_{j=1}^{n} w_{1t}^{(j)}}{\sum_{j=1}^{n} w_{2,t+1}^{(j)}}.$$

This finding is similar to the finding of chapter 1. It can be generalized to the case $\beta_j \neq \beta_{j'}$, $j \neq j'$. Such a result clearly reveals the existence of a representative agent with aggregated endowments.

Example. The agent solves the following program:

$$\begin{array}{l} \max_{\{c,\theta\}} \left\{ u(c_{1,t}) + \beta E\left[u(c_{2,t+1}) | \mathcal{F}_t \right] \right\} \\ \text{s.t.} \left\{ \begin{array}{l} q_t \cdot \theta_t + c_{1,t} = w_{1t} \\ c_{2,t+1} = (q_{t+1} + D_{t+1}) \cdot \theta_t + w_{2,t+1} \end{array} \right. \end{array}$$

The notation is the same as the notation in section 3.?. The agent born at time $t-1$ faces the constraints $q_{t-1} \cdot \theta_{t-1} + c_{1,t-1} = w_{1,t-1}$ and $w_{2t} + (q_t + D_t) \cdot \theta_{t-1} = c_{2,t}$. By combining the second period constraint of the agent born at time $t-1$ with the first period constraint of the agent born at time t one obtains:

$$(q_t + D_t) \cdot \theta_{t-1} - q_t \cdot \theta_t + w_t = c_{1,t} + c_{2,t},$$

Finally, by using the financial market clearing condition:

$$\theta_t = 1_{m \times 1} \quad \forall t,$$

one gets:

$$\sum_{i=1}^{m} D_t^{(i)} + \bar{w}_t = c_{1,t} + c_{2,t},$$

The financial market equilibrium implies that all fruits (plus the exogeneous endowments) are totally consumed.

By eliminating c from the program,

$$V(w) = \max_{\theta} \left\{ u\left(w_{1t} - q_t \cdot \theta\right) + \beta E\left[u\left((q_{t+1} + D_{t+1}) \cdot \theta \right) | \mathcal{F}_t \right] \right\}.$$

The first order condition leads to the same result of the model with a representative agent:

$$u'(c_{1,t}) q_t^{(i)} = \beta E\left[u'(c_{2,t+1}) \left(q_{t+1}^{(i)} + D_{t+1}^{(i)} \right) \Big| \mathcal{F}_t \right].$$

Example: $u(c, c_+) = \frac{c^\sigma}{\sigma} + \beta \frac{c_+^\sigma}{\sigma}$. As in the previous chapter,

$$???c_+ = \frac{w_{0j}}{1 + \beta^{1/(1-\sigma)} R^{\sigma/(1-\sigma)}}.???$$

3.6.2 Money

Next, we consider the "monetary" version of the model presented in the introductory section. The agent solves program (3.29), but the constraint has now a different interpretation. Specifically,

$$\begin{cases} \frac{m_t}{p_t} + c_{1,t} = w_{1t} \\ c_{2,t+1} = \frac{m_t}{p_{t+1}} + w_{2,t+1} \end{cases} \tag{3.42}$$

Here real savings are given by:

$$s_t \equiv \frac{m_t}{p_t}.$$

In addition, by letting:

$$R_{t+1} \equiv \frac{p_t}{p_{t+1}},$$

we obtain $\frac{m_t}{p_{t+1}} = R_{t+1}s_t$. *Formally*, the two constraints in (3.30) and (3.35) are the same. The good is perishable here, too, but agents may want to transfer value intertemporally with money. Naturally, we also have here (as in (3.31)) that

$$s_{t-1}R_t = s_t - (w_t - c_{1t} - c_{2t}), \tag{3.43}$$

but the dynamics is different because the equilibrium *does not* necessarily impose that $s_t \equiv \frac{m_t}{p_t} = 0$: money *may* be transferred from a generation to another one, and the equilibrium is:

$$s_t = \frac{\bar{m}_t}{p_t},$$

where $\bar{m}$ denotes money supply. In addition, the second term of the r.h.s. of eq. (3.36) is not necessarily zero due to the existence of monetary transfers. Generally, one has that additional quantity of money in circulation are transferred from a generation to another. In equilibrium,

$$p_t(c_{1,t} + c_{2,t}) = p_t w_t - \Delta\bar{m}_t. \tag{3.44}$$

By replacing (3.37) into (3.36),

$$s_{t-1}R_t = s_t - \frac{\Delta\bar{m}_t}{p_t}. \tag{3.45}$$

Next suppose wlg that the rule of monetary creation has the form:

$$\frac{\Delta\bar{m}_t}{\bar{m}_{t-1}} = \mu_t \in [0, \infty) \ \ \forall t.$$

In equilibrium $\frac{\Delta\bar{m}_t}{p_t} = \frac{\Delta\bar{m}_t}{m_{t-1}}\frac{m_{t-1}}{p_t} = \mu_t R_t s_{t-1}$, and eq. (3.38) becomes:

$$(1 + \mu_t)s_{t-1}R_t = s_t.$$

Next, note that the optimal decisions are such that real savings are of the form:

$$s_t = s(R_{t+1}),$$

and the dynamics of the nominal interest rates becomes, for $s'(R) \neq 0$,

$$(1 + \mu_t)s(R_t)R_t = s(R_{t+1}). \tag{3.46}$$

Here is another way to derive relation (3.39). By definition, $(1+\mu_t)\bar{m}_{t-1} = \bar{m}_t$, and by dividing by p_t, $(1+\mu_t)\frac{\bar{m}_{t-1}}{p_{t-1}}\frac{p_{t-1}}{p_t} = \frac{\bar{m}_t}{p_t}$. At the equilibrium, supply of real money by old ($\bar{m}_{t-1}$) plus new creation of money (i.e. $\mu \cdot \bar{m}_{t-1}$) (i.e. the l.h.s.) is equal to demand of real money by young (i.e. the r.h.s.), and this is exactly relation (3.39).[13]

The previous things can be generalized to the case of population growth. At time t, N_t individuals are born, and N_t obeys the relation $\frac{N_t}{N_{t-1}} = (1+n)$, $n \in \mathbb{R}$. In this case, the relation defining the evolution of money supply is:

$$\frac{\Delta M_t}{M_{t-1}} = \mu_t \in [0, \infty) \ \ \forall t,$$

where

$$M_t \equiv N_t \cdot \bar{m}_t,$$

or,

$$(1+\mu_t)\bar{m}_{t-1} = (1+n)\bar{m}_t.$$

By a reasoning similar to the one made before,

$$\frac{1+\mu_t}{1+n} \cdot s(R_t)R_t = s(R_{t+1}). \tag{3.47}$$

Suppose that the trajectory μ does not depend on R, and that $(\mu_t)_{t=0}^{\infty}$ has a unique stationary solution denoted as μ.

We have two stationary equilibria:

- $R = \frac{1+n}{1+\mu}$. This point corresponds to the "golden rule" when $\mu = 0$ (cf. section 3.9). In the general case, the solution for the price is $p_t = \left(\frac{1+\mu}{1+n}\right)^t p_0$, and one has that (1) $\frac{\bar{m}_t}{p_t} = \frac{M_t}{N_t \cdot p_t} = \frac{M_0}{N_0 \cdot p_0}$, and (2) $\frac{\bar{m}_t}{p_{t+1}} = \frac{M_0}{N_0 \cdot p_0}\frac{1+n}{1+\mu}$. Therefore, the agents' budget constraints are bounded and in addition, the real value of transferrable money is strictly positive. This is the stationary equilibrium in which agents "trust" money.

- $s(R_a) = 0$. This point corresponds to the autarchic state. Generally, one has that $R_a < R$, which means that prices increase more rapidly than the per-capita money stock. Of course this is the stationary equilibrium in which agents do not trust money. Analytically, in this state one has that $R_a < R \Leftrightarrow \frac{p_{t+1}}{p_t} > \frac{1+\mu}{1+n} = \frac{M_{t+1}/M_t}{N_{t+1}/N_t} = \frac{\bar{m}_{t+1}}{\bar{m}_t} \Leftrightarrow \frac{\bar{m}_{t+1}}{p_{t+1}} < \frac{\bar{m}_t}{p_t}$, whence $(\frac{\bar{m}}{p})_t \underset{t\to\infty}{\to} 0+$. As regards $\frac{\bar{m}_t}{p_{t+1}}$, one has $\frac{\bar{m}_t}{p_{t+1}} = \frac{\bar{m}_t}{p_t}R_a < \frac{\bar{m}_t}{p_t}R = \frac{\bar{m}_t}{p_t}\frac{1+n}{1+\mu}$, and since $(\frac{\bar{m}}{p})_t \underset{t\to\infty}{\to} 0+$, then $\frac{\bar{m}_t}{p_{t+1}} \underset{t\to\infty}{\to} 0+$.

If $s(\cdot)$ is differentiable and $s'(\cdot) \neq 0$, the dynamics of $(R_t)_{t=0}^{\infty}$ can be studied through the slope,

$$\frac{dR_{t+1}}{dR_t} = \frac{s'(R_t)R_t + s(R_t)}{s'(R_{t+1})}\frac{1+\mu_t}{1+n}. \tag{3.48}$$

There are three cases:

[13] The analysis here is based on the assumption that money transfers are made to youngs: youngs not only receive money by old ($\bar{m}_{t-1}$), but also new money by the central bank ($\mu_t\bar{m}_{t-1}$). One can consider an alternative model in which transfers are made directly to old. In this case, the budget constraint becomes:

$$\begin{cases} s_t + c_{1,t} = w_{1t} \\ c_{2,t+1} = s_t R_{t+1}\left(1+\mu_{t+1}\right) + w_{2,t+1} \end{cases}$$

and at the aggregate level, $(1+\mu_t)\, s_{t-1}R_t = s_t - (w_t - c_{1t} - c_{2t})$.

- $s'(R) > 0$. Gross substituability: the revenue effect is dominated by the substitution effect.
- $s'(R) = 0$. Revenue and substitution effects compensate each other.
- $s'(R) < 0$. Complementarity: the revenue effect dominates the substitution effect.

An example of gross substituability was provided during the presentation of the introductory examples of the present section (log utility functions). The second case can be obtained with the same examples after imposing that agents have no endowments in the second period. The equilibrium is seriously compromised in this case, however. Another example is obtained with Cobb-Douglas utility functions: $u(c_{1t}, c_{2,t+1}) = c_{1t}^{l_1} \cdot c_{2,t+1}^{l_2}$, which generates a real savings function $s(R_{t+1}) = \frac{\frac{l_2}{l_1} w_{1t} - \frac{w_{2,t+1}}{R_{t+1}}}{1+\frac{l_2}{l_1}}$ the derivative of which is nil when one assumes that $w_{2,t} = 0$ for all t, which also implies $\frac{\bar{m}_t}{p_t} = s_t = \frac{1}{\nu} w_{1t}$, $\nu \equiv \frac{l_1+l_2}{l_2}$ and, by reorganizing,

$$\bar{m}_t \nu = p_t w_{1t},$$

an equation supporting the view of the Quantitative Theory of money. In this case, the sequence of gross returns is $R_{t+1} = \frac{p_t}{p_{t+1}} = \frac{\bar{m}_t}{\bar{m}_{t+1}} \frac{w_{1,t+1}}{w_{1,t}}$, or

$$R_{t+1} = \frac{(1+n) \cdot (1+g_{t+1})}{1+\mu_{t+1}},$$

where g_{t+1} denotes the growth rate of endowments of young between time t and time $t+1$. The inflation factor R_t^{-1} is equal to the monetary creation factor corrected for the the growth rate of the economy.

Another example is $u(c_{1t}, c_{2,t+1}) = \left(l c_{1t}^{(\eta-1)/\eta} + (1-l) c_{2,t+1}^{(\eta-1)/\eta} \right)^{\eta/(\eta-1)}$. Note that $\lim_{\eta \to 1} u(c_{1t}, c_{2,t+1}) = c_{1t}^{l} \cdot c_{2,t+1}^{1-l}$, Cobb-Douglas. We have

$$\begin{cases} c_{1t} & = & \frac{R_{t+1} w_{1t} + w_{2,t+1}}{R_{t+1} + K^{\eta} R_{t+1}^{\eta}}, \ K \equiv \frac{1-l}{l} \\ c_{2,t+1} & = & \frac{R_{t+1} w_{1t} + w_{2,t+1}}{1 + K^{-\eta} R_{t+1}^{1-\eta}} \\ s_t & = & \frac{K^{\eta} R_{t+1}^{\eta} w_{1t} - w_{2,t+1}}{R_{t+1} + K^{\eta} R_{t+1}^{\eta}} \end{cases}$$

To simplify, suppose that $K = 1$, and $0 = w_{2t} = \mu_t = n$, and $w_{1t} = w_{1,t+1}$ $\forall t$. It is easily checked that

$$\text{sign}\,(s'(R)) = \text{sign}\,(\eta - 1)\,.$$

The interest factor dynamics is:

$$R_{t+1} = (R_t^{-\eta} + R^{-1} - 1)^{1/(1-\eta)}. \tag{3.49}$$

The stationary equilibria are the solutions of

$$R^{1-\eta} = R^{-\eta} + R^{-1} - 1,$$

and one immediately verifies the existence of the monetary state $R = 1$.

When $\eta > 1$, the dynamics is quite simple studying: in general, one has $R_a = 0$ and $R = 1$, and R_a is stable and R is unstable. Figure 3.5 illustrates this situation in the case $\eta = 2$.

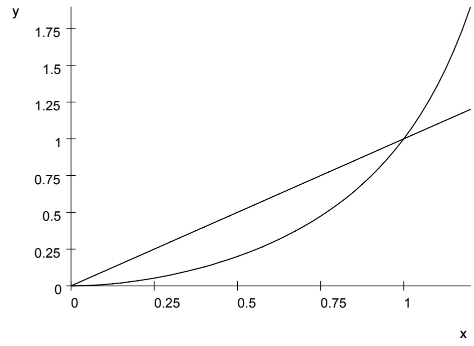


FIGURE 3.5. $\eta = 2$

When $\eta < 1$, the situation is more delicate. In this case, R_a is generically ill-defined, and $R = 1$ is not necessarily stable. One can observe a dynamics converging towards R, or even the emergence of more or less "regular" cycles. In this respect, it is important to examine the slope of the slope of the map in (3.42) in correspondence with $R = 1$:

$$\left.\frac{dR_{t+1}}{dR_t}\right|_{R_{t+1}=R_t=1} = \frac{\eta+1}{\eta-1}.$$

Here are some <u>hints</u> concerning the general case. Figure 3.6 depicts the shape of the map $R_t \mapsto R_{t+1}$ in the case of gross substituability (in fact, the following arguments can also be adapted verbatim to the complementarity case whenever $\forall x$, $\frac{s'(x)x}{s(x)} < -1$: indeed, in this case $\frac{dR_{t+1}}{dR_t} > 0$ since the numerator is negative and the denominator is also negative by assumption. Such a case does not have any significative economic content, however). This is an increasing function since the slope $\frac{dR_{t+1}}{dR_t} = \frac{s'(R_t)R_t+s(R_t)}{s'(R_{t+1})}\frac{1+\mu_t}{1+n} > 0$. In addition, the slope (3.41) computed in correspondence with the monetary state $R = \frac{1+n}{1+\mu}$ is:

$$\left.\frac{dR_{t+1}}{dR_t}\right|_{R_{t+1}=R_t=R} = 1 + \frac{s(R)}{Rs'(R)}.$$

This is always greater than 1 if $s'(.) > 0$, and in this case the monetary state M is unstable and the autarchic state A is stable. In particular, all paths starting from the right of M are unstables. They imply an increasing sequence of R, i.e. a decreasing sequence of p. This can not be an equilibrium because it contradicts the budget constraints (in fact, there would not be a solution to the agents' programs). It is necessary that the economy starts from A and M, but we have not endowed with additional pieces of information: there is a continuum of points $R_1 \in [R_a, R)$ which are candidates for the beginning of the equilibria sequence. Contrary to the models of the previous sections, here we have indeterminacy of the equilibrium, which is parametrized by p_0.

Is the autarchic state the only possible stable configuration? The answer is no. It is sufficient that the map $R_t \mapsto R_{t+1}$ bends backwards and that $\left.\frac{dR_{t+1}}{dR_t}\right|_{R_{t+1}=R_t=R} < -1$ to make M a stable state. A condition for the curve to bend backward is that $\frac{s'(R)R}{s(R)} > -1$, and the condition for $\left.\frac{dR_{t+1}}{dR_t}\right|_{R_{t+1}=R_t=R} < -1$ to hold is that $\frac{s'(R)R}{s(R)} > -\frac{1}{2}$. If $\frac{s'(R)R}{s(R)} > -\frac{1}{2}$, M is attained starting from a

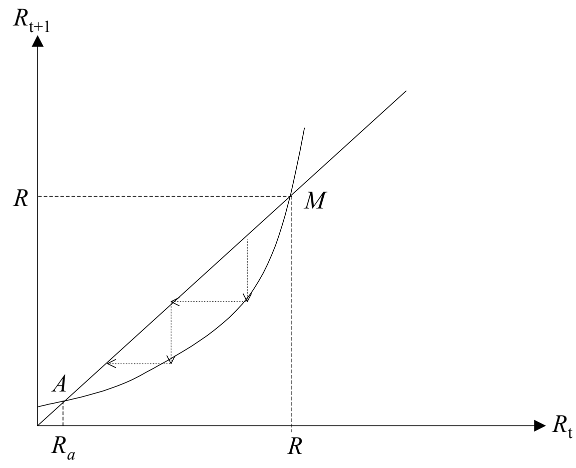


FIGURE 3.6. Gross substituability

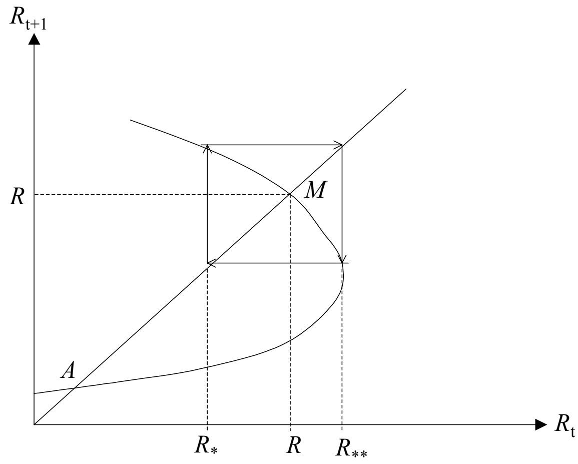


FIGURE 3.7.

sufficiently small neighborhood of M. Figure 3.7 shows the emergence of a cycle of order two,[14] in which $R_{**} = \frac{R^2}{R_*}$.[15] Notice that in this case, the dynamics of the system has been analyzed in a backward-looking manner, not in a forward-looking manner. The reason is that there is an indeterminacy of the forward-looking dynamics, and it is thus necessary to analyze the system dynamics in a backward-looking manner ... In any case, the condition $s'(R) < 0$ is not appealing on an economic standpoint.

[14]There are more complicated situations in which cycles of order 3 may exist, whence the emergence of what is known as "chaotic" trajectories.

[15]Here is the proof. Starting from relation (3.40), we have that for a 2-cycle,

$$\begin{cases} \frac{1+\mu}{1+n} R_* s(R_*) = s(R_{**}) \\ \frac{1+\mu}{1+n} R_{**} s(R_{**}) = s(R_*) \end{cases}$$

By multiplying the two sides of these equations, one recovers the desired result.

3.6.3 Money in a model with real shocks

The Lucas (1972) model[16] is the first to address issues concerning the neutrality of money in a context of overlapping generations with stochastic features. Here we present a very simplified version of the model.[17]

The agent works during the first period and consumes in the second period: this is, of course, nothing but a pleasant simplification. Disutility of work is $-v$. We suppose that $v'(x) > 0$, $v''(x) > 0$, $\forall x \in \mathbb{R}_+$. The second period consumption utility is u, and has the usual properties.

The program is:

$$\max_{\{n,c\}} \{-v(n_t) + \beta E[u(c_{t+1})|\,\mathcal{F}_t]\}$$
$$\text{s.t.} \begin{cases} m = p_t y_t, \ y_t = \epsilon_t n_t \\ p_{t+1} c_{t+1} = m \end{cases}$$

where m denotes money, y denotes the production obtained by means of labour n, and $\{\epsilon_t\}_{t=0,1,\cdots}$ is a sequence of strictly positive shocks affecting labour productivity. The price level is p.

Let us rewrite the program by using the single constraint $c_{t+1} = R_{t+1}\epsilon_t n_t$, $R_{t+1} \equiv \frac{p_t}{p_{t+1}}$,

$$\max_n \{-v(n_t) + \beta E[u(R_{t+1}\epsilon_t n_t)|\,\mathcal{F}_t)]\}.$$

The first order condition is $v'(n_t) = \beta E(u'(R_{t+1}\epsilon_t n_t)R_{t+1}\epsilon_t \,/\, \mathcal{F}_t) \Leftrightarrow$

$$v'(n_t)\frac{1}{p_t\epsilon_t} = \beta E\left[u'(R_{t+1}\epsilon_t n_t)\frac{1}{p_{t+1}}\middle|\,\mathcal{F}_t\right]$$

At the equilibrium, $y_t = c_t \Rightarrow c_t = \epsilon_t \cdot n_t$, and

$$c_{t+1} = \frac{m}{p_{t+1}} = \epsilon_{t+1} \cdot n_{t+1}.$$

By replacing the previous relation into the first order condition, and simplifying,

$$v'(n_t)n_t = \beta E[u'(\epsilon_{t+1} \cdot n_{t+1})(\epsilon_{t+1} \cdot n_{t+1})|\,\mathcal{F}_t]. \tag{3.50}$$

As in the model of section 3.6, the rational expectation assumption consists in regarding all the model's variables as functions of the state varibales. Here, the states of natures are generated by ϵ, and we have:

$$n = n(\epsilon).$$

By plugging $n(\epsilon)$ into (3.43) we get:

$$v'(n(\epsilon))n(\epsilon) = \beta \int_{\text{supp}(\epsilon)} u'\left(\epsilon^+ n(\epsilon^+)\right)\epsilon^+ n(\epsilon^+) dP(\epsilon^+|\,\epsilon). \tag{3.51}$$

A rational expectations equilibrium is a statistical sequence $n(\epsilon)$ satisfying the functional equation (3.44).

Eq. (3.44) considerably simplifies when shocks are i.i.d.: $P(\epsilon^+ \,/\, \epsilon) = P(\epsilon^+)$,

$$v'(n(\epsilon))n(\epsilon) = \beta \int_{\text{supp}(\epsilon)} u'\left(\epsilon^+ n(\epsilon^+)\right)\epsilon^+ n(\epsilon^+) dP(\epsilon^+). \tag{3.52}$$

[16] Lucas, R.E., Jr. (1972): "Expectations and the Neutrality of Money," *J. Econ. Theory*, 4, 103-124.

[17] Parts of this simplified version of the model are taken from Stokey et Lucas (1989, p. 504): Stokey, N.L. and R.E. Lucas (with E.C. Prescott) (1989): *Recursive Methods in Economic Dynamics*, Harvard University Press.

In this case, the r.h.s. of the previous relationship does not depend on ϵ, which implies that the l.h.s. does not depend on ϵ neither. Therefore, the only candidate for the solution for n is a constant $\bar{n}$:[18]

$$n(\epsilon) = \bar{n}, \quad \forall \epsilon.$$

Provided such a $\bar{n}$ exists, this is a result on money neutrality. More precisely, relation (3.45) can be written as:

$$v'(\bar{n}) = \beta \int_{\text{supp}(\epsilon)} u'(\epsilon^+ \bar{n})\epsilon^+ dP(\epsilon^+),$$

and it is always possible to impose reasonable conditions on v and u that ensure existence and unicity of a strictly positive solution for $\bar{n}$, as in the following example.

Example. $v(x) = \frac{1}{2}x^2$ and $u(x) = \log x$. The solution is $\bar{n} = \sqrt{\beta}$, $y(\epsilon) = \epsilon\sqrt{\beta}$ and $p(\epsilon) = \frac{m}{\epsilon\sqrt{\beta}}$.

Exercise. Extend the previous model when the money supply follows the stochastic process: $\frac{\Delta m_t}{m_{t-1}} = \mu_t$, where $\{\mu_t\}_{t=0,1,\cdots}$ is a i.i.d. sequence of shocks.

3.6.4 The Diamond's model

3.7 Optimality

3.7.1 Models with productive capital

The starting point is the relation

$$K_{t+1} = S_t = Y(K_t, N_t) - C_t, \; K_0 \text{ given.}$$

Dividing both sides by N_t one gets:

$$k_{t+1} = \frac{1}{1+n}\left(y(k_t) - c_t\right), \; k_0 \text{ given and: } \inf k_t \geq 0, \; \sup k_t < \infty. \tag{3.53}$$

The stationary state of this economy is:

$$c = y(k) - (1+n)k,$$

and we see that the per-capita consumption attains its maximum at:

$$\bar{k} : y'(\bar{k}) = 1 + n.$$

This is the *golden rule.*

If $y'(\bar{k}) < 1 + n$, it is always possible to increase per-capita consumption at the stationary state; indeed, since $y(k)$ is given, one can decrease k and obtain $dc = -(1+n)dk > 0$, and in the next periods, one has $dc = (y'(k) - (1+n))\, dk > 0$. We wish to provide a formal proof of these facts along the entire capital accumulation path of the economy. Notice that the foregoing

[18] A rigorous proof that $n(\epsilon) = \bar{n}$, $\forall \epsilon$ is as follows. Let's suppose the contrary, i.e. there exists a point ϵ_0 and a neighborhood of ϵ_0 such that either $n(\epsilon_0 + A) > n(\epsilon_0)$or $n(\epsilon_0 + A) < n(\epsilon_0)$, where the constant $A > 0$. Let's consider the first case (the proof of the second case being entirely analogous). Since the r.h.s. of (4.43) is constant for all ϵ, we have, $v'(n(\epsilon_0 + A)) \cdot n(\epsilon_0 + A) = v'(n(\epsilon_0)) \cdot n(\epsilon_0) \leq v'(n(\epsilon_0 + A)) \cdot n(\epsilon_0)$, where the inequality is due to the assumption that $v'' > 0$ always holds. We have thus shown that, $v'(n(\epsilon_0 + A)) \cdot [n(\epsilon_0 + A) - n(\epsilon_0)] \leq 0$. Now $v' > 0$ always holds, so that $n(\epsilon_0 + A) < n(\epsilon_0)$, a contradiction with the assumption $n(\epsilon_0 + A) > n(\epsilon_0)$.

results are valid whatever the structure of the economy is (e.g., a finite number of agents living forever or overlapping generations). As an example, in the overlapping generations case one can interpret relation (3.46) as the one describing the capital accumulation path in the Diamond's model once c_t is interpreted as: $c_t \equiv \frac{C_t}{N_t} = c_{1t} + \frac{c_{2,t+1}}{1+n}$.

The notion of efficiency that we use is the following one:[19] a path $\{(k,c)_t\}_{t=0}^{\infty}$ is consumption-inefficient if there exists another path $\{(\tilde{k},\tilde{c})_t\}_{t=0}^{\infty}$ satisfying (3.46) and such that, for all t, $\tilde{c}_t \geq c_t$, with at least one strict inequality.

We now present a weaker version of thm. 1 p. 161 of Tirole (1988), but easier to show.[20]

Theorem 3.2 ((weak version of the) Cass-Malinvaud theory). (a) *A path* $\{(k,c)_t\}_{t=0}^{\infty}$ *is consumption efficient if* $\frac{y'(k_t)}{1+n} \geq 1\ \forall t$. (b) *A path* $\{(k,c)_t\}_{t=0}^{\infty}$ *is consumption inefficient if* $\frac{y'(k_t)}{1+n} < 1\ \forall t$.

Proof. (a) Let $\tilde{k}_t = k_t + \epsilon_t$, $t = 0, 1, \cdots$, an alternative consumption efficient path. Since k_0 is given, $\epsilon_t = 0$. Furthermore, by relation (3.46) one has that

$$(1+n)\cdot(\tilde{k}_{t+1} - k_{t+1}) = y(\tilde{k}_t) - y(k_t) - (\tilde{c}_t - c_t),$$

and because $\tilde{k}$ has been supposed to be efficient, $\tilde{c}_t \geq c_t$ with at least one strictly equality over the ts. Therefore, by concavity of y,

$$\begin{aligned} 0 &\leq y(\tilde{k}_t) - y(k_t) - (1+n)(\tilde{k}_{t+1} - k_{t+1}) \\ &= y(\tilde{k}_t) - y(k_t) - (1+n)\epsilon_{t+1} \\ &< y(k_t) + y'(k_t)\epsilon_t - y(k_t) - (1+n)\epsilon_{t+1}, \end{aligned}$$

or

$$\epsilon_{t+1} < \frac{y'(k_t)}{1+n}\epsilon_t.$$

Evaluating the previous inequality at $t = 0$ yields $\epsilon_1 < \frac{y'(k_0)}{1+n}\epsilon_0$, and since $\epsilon_0 = 0$, one has that $\epsilon_1 < 0$. Since $\frac{y'(k_t)}{1+n} \geq 1\ \forall t$, $\epsilon_t \underset{t\to\infty}{\to} -\infty$, which contradicts (3.46).

(b) The proof is nearly identical to the one of part (a) with the obvious exception that $\liminf \epsilon_t >> -\infty$ here. Furthermore, note that there are infinitely many such sequences that allow for efficiency improvements. ‖

Are actual economies dynamically efficient? To address this issue, Abel et al. (1989)[21] modified somehow the previous setup to include uncertainty, and conclude that the US economy does satisfy their dynamic efficiency requirements.

The conditions of the previous theorem are somehow restrictive. As an example, let us take the model of section 3.2 and fix, as in section 3.2, $n = 0$ to simplify. As far as $k_0 < k = (y')^{-1}[\beta^{-1}]$, per-capita capital is such that $y'(k_t) > 1\ \forall t$ since the dynamics here is of the saddlepoint type

[19] Tirole, J. (1988): "Efficacité intertemporelle, transferts intergénérationnels et formation du prix des actifs: une introduction," in: *Melanges économiques. Essais en l'honneur de Edmond Malinvaud.* Paris: Editions Economica & Editions EHESS, p. 157-185.

[20] The proof we present here appears in Touzé, V. (1999): *Financement de la sécurité sociale et équilibre entre les générations*, unpublished PhD dissertation Univ. Paris X Nanterre.

[21] Abel, A.B., N.G. Mankiw, L.H. Summers and R.J. Zeckhauser (1989): "Assessing Dynamic Efficiency: Theory and Evidence," *Review Econ. Studies*, 56, 1-20.

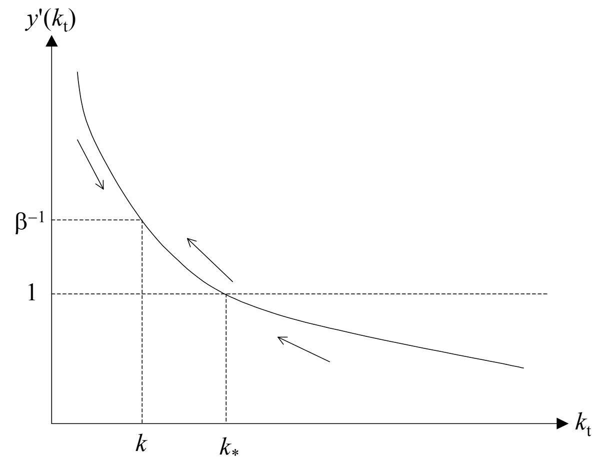


FIGURE 3.8. Non-necessity of the conditions of thm. 3.2 in the model with a representative agent.

and then monotone (see figure 3.8). Therefore, the conditions of the theorem are fulfilled. Such conditions also hold when $k_0 \in [k, k_*]$, again by the monotone dynamics of k_t. Nevertheless, the conditions of the theorem do not hold anymore when $k_0 > k_*$ and yet, the capital accumulation path is still efficient! While it is possible to show this with the tools of the evaluation equilibria of Debreu (1954), here we provide the proof with the same tools used to show thm. 3.2. Indeed, let

$$\tau = \inf\{t : k_t \leq k_*\} = \inf\{t : y'(k_t) \geq 1\}.$$

We see that $\tau < \infty$, and since the dynamics is monotone, $\begin{cases} y'(k_t) < 1, \ t = 0, 1, \cdots, \tau - 1 \\ y'(k_t) \geq 1, \ t = \tau, \tau + 1, \cdots \end{cases}$. By using again the same arguments used to show thm. 3.2, we see that since τ is finite, $-\infty < \epsilon_{\tau+1} < 0$. From τ onwards, an explosive sequence ϵ starts unfolding, and $\epsilon_t \underset{t\to\infty}{\to} -\infty$.

3.7.2 Models with money

The decentralized economy is characterized by the presence of money. Here we are interested in first best optima i.e., optima that a social planner may choose by acting directly on agents consumptions.[22] Let us first analyze the stationary state $R = \frac{1+n}{1+\mu}$ and show that it corresponds to the stationary state in which consumptions and endowments are constants, and agents' utility is maximized when $\mu = 0$. Indeed, here the social planner allocates resources without caring about monetary phenomena, and the only constraint is the following "natural" constraint:

$$w_n \equiv w_1 + \frac{w_2}{1+n} = c_1 + \frac{c_2}{1+n},$$

in which case the utility of the "stationary agent" is:

$$u(c_1, c_2) = u(w_n - \frac{c_2}{1+n}, c_2).$$

[22] In our terminology, a second best optimum is the one in which the social planner makes the thought experiment to let the market "play" first (with money) and then parametrizes such virtual equilibria by μ_t. The resulting indirect utility functions are expressed in terms of such μ_ts and after creating an aggregator of such indirect utility functions, the social planner maximises such an aggregator with respect to μ_t.

The first order condition is $\frac{u_{c2}}{u_{c1}} = \frac{1}{1+n}$. In the market equilibrium, the first order condition is $\frac{u_{c2}}{u_{c1}} = \frac{1}{R}$, which means that in the market equilibrium, the golden rule is attained at R if and only if $\mu = 0$, as claimed in section 3.?

The convergence of the optimal policy of the social planner towards the gloden rule can be verified as it follows. The social planner solves the program:

$$\begin{cases} \max \sum_{t=0}^{\infty} \vartheta^t \cdot u^{(t)}(c_{1t}, c_{2,t+1}) \\ w_{nt} \equiv w_{1t} + \frac{w_{2t}}{1+n} = c_{1t} + \frac{c_{2t}}{1+n} \end{cases}$$

or $\max \sum_{t=0}^{\infty} \vartheta^t \cdot u^{(t)} \left(w_{nt} - \frac{c_{2t}}{1+n}, c_{2,t+1} \right)$. Here the notation $u^{(t)}(.)$ has been used to stress the fact that endowments may change from one generation to another generation, and ϑ is the weighting coefficient of generations that is used by the planner. The first order conditions,

$$\frac{u_{c2}^{(t-1)}}{u_{c1}^{(t)}} = \frac{\vartheta}{1+n},$$

lead towards the modified golden rule at the stationary state (modified by ϑ).

3.8 Appendix 1: Finite difference equations and economic applications

Let $z_0 \in \mathbb{R}^d$, and consider the following linear system of finite difference equations:

$$z_{t+1} = A \cdot z_t, \quad t = 0, 1, \cdots, \tag{3A1.1}$$

where A is conformable matrix. Solution is,

$$z_t = v_1 \kappa_1 \lambda_1^t + \cdots + v_d \kappa_d \lambda_d^t,$$

where λ_i and v_i are the eigenvalues and the corresponding eigenvectors of A, and κ_i are constants which will be determined below.

The classical method of proof is based on the so-called diagonalization of system (3A1.1). Let us consider the system of characteristic equations for A, $(A - \lambda_i I)\, v_i = \mathbf{0}_{d\times 1}$, λ_i scalar and v_i a column vector $d \times 1$, $i = 1, \cdots, n$, or in matrix form, $AP = P\Lambda$, where $P = (v_1, \cdots, v_d)$ and Λ is a diagonal matrix with λ_i on the diagonal. By post-multiplying by P^{-1} one gets the decomposition[23]

$$A = P\Lambda P^{-1}. \tag{3A1.2}$$

By replacing (3A1.2) into (3A1.1), $P^{-1} z_{t+1} = \Lambda \cdot P^{-1} z_t$, or

$$y_{t+1} = \Lambda \cdot y_t, \quad y_t \equiv P^{-1} z_t.$$

The solution for y is:

$$y_{it} = \kappa_i \lambda_i^t,$$

and the solution for z is: $z_t = P y_t = (v_1, \cdots, v_d) y_t = \sum_{i=1}^d v_i y_{it} = \sum_{i=1}^d v_i \kappa_i \lambda_i^t$.

To determine the vector of the constants $\kappa = (\kappa_1, \cdots, \kappa_d)^T$, we first evaluate the solution at $t = 0$,

$$z_0 = (v_1, \cdots, v_d)\kappa = P\kappa,$$

whence

$$\hat{\kappa} \equiv \kappa(P) = P^{-1} z_0,$$

where the columns of P are vectors $\in$ the space of the eigenvectors. Naturally, there is an infinity of such Ps, but the previous formula shows how $\kappa(P)$ must "adjust" to guarantee the stability of the solution with respect to changes of P.

3A.1 Example. $d = 2$. Let us suppose that $\lambda_1 \in (0, 1)$, $\lambda_2 > 1$. The system is unstable in correspondence with any initial condition but a set of zero measure. This set gives rise to the so-called saddlepoint path. Let us compute its coordinates. The strategy consists in finding the set of initial conditions for which $\kappa_2 = 0$. Let us evaluate the solution at $t = 0$,

$$\begin{pmatrix} x_0 \\ y_0 \end{pmatrix} = z_0 = P\kappa = (v_1, v_2) \begin{pmatrix} \kappa_1 \\ \kappa_2 \end{pmatrix} = \begin{pmatrix} v_{11}\kappa_1 + v_{12}\kappa_2 \\ v_{21}\kappa_1 + v_{22}\kappa_2 \end{pmatrix},$$

where we have set $z = (x, y)^T$. By replacing the second equation into the first one and solving for κ_2,

$$\kappa_2 = \frac{v_{11} y_0 - v_{21} x_0}{v_{11} v_{22} - v_{12} v_{21}}.$$

This is zero when

$$y_0 = \frac{v_{21}}{v_{11}} x_0.$$

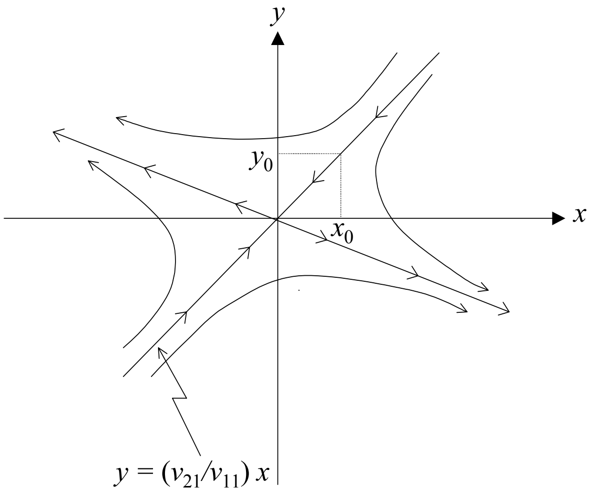


FIGURE 3.9.

Here the saddlepoint is a line with slope equal to the ratio of the components of the eigenvector associated with the root with modulus less than one. The situation is represented in figure 3.9, where the "divergent" line has as equation $y_0 = \frac{v_{22}}{v_{12}} x_0$, and corresponds to the case $\kappa_1 = 0$.

The economic content of the saddlepoint is the following one: if x is a *predetermined* variable, y must "jump" to $y_0 = \frac{v_{21}}{v_{11}} x_0$ to make the system display a non-explosive behavior. Notice that there is a major conceptual difficulty when the system includes two predetermined variables, since in this case there are generically no stable solutions. Such a possibility is unusual in economics, however.

4A.2 Example. The previous example can be generated by the neoclassic growth model. In section 3.2.3, we showed that in a small neighborhood of the stationary values k, c, the dynamics of $(\hat{k}_t, \hat{c}_t)_t$ (deviations of capital and consumption from their respective stationary values k, c) is:

$$\begin{pmatrix} \hat{k}_{t+1} \\ \hat{c}_{t+1} \end{pmatrix} = A \begin{pmatrix} \hat{k}_t \\ \hat{c}_t \end{pmatrix}$$

where

$$A \equiv \begin{pmatrix} y'(k) & -1 \\ -\frac{u'(c)}{u''(c)} y''(k) & 1 + \beta \frac{u'(c)}{u''(c)} y''(k) \end{pmatrix}, \quad \beta \in (0,1).$$

By using the relationship $\beta y'(k) = 1$, and the conditions imposed on u and y, we have

- $\det(A) = y'(k) = \beta^{-1} > 1$;
- $\operatorname{tr}(A) = \beta^{-1} + 1 + \beta \frac{u'(c)}{u''(c)} y''(k) > 1 + \det(A)$.

The two eigenvalues are solutions of a quadratic equation, and are: $\lambda_{1/2} = \frac{\operatorname{tr}(A) \mp \sqrt{\operatorname{tr}(A)^2 - 4\det(A)}}{2}$. Now,

$$\begin{aligned} a &\equiv \operatorname{tr}(A)^2 - 4\det(A) \\ &= \left[\beta^{-1} + 1 + \beta \frac{u'(c)}{u''(c)} y''(k)\right]^2 - 4\beta^{-1} \\ &> \left(\beta^{-1} + 1\right)^2 - 4\beta^{-1} \\ &= \left(1 - \beta^{-1}\right)^2 \\ &> 0. \end{aligned}$$

[23] The previous decomposition is known as the *spectral decomposition* if $P^\top = P^{-1}$. When it is not possible to diagonalize A, one may make reference to the canonical transformation of Jordan.

It follows that $\lambda_2 = \frac{\mathrm{tr}(A)}{2} + \frac{\sqrt{a}}{2} > \frac{1+\det(A)}{2} + \frac{\sqrt{a}}{2} > 1 + \frac{\sqrt{a}}{2} > 1$.

To show that $\lambda_1 \in (0,1)$, notice first that since $\det(A) > 0$, $2\lambda_1 = \mathrm{tr}(A) - \sqrt{\mathrm{tr}(A)^2 - 4\det(A)} > 0$. It remains to be shown that $\lambda_1 < 1$. But $\lambda_1 < 1 \Leftrightarrow \mathrm{tr}(A) - \sqrt{\mathrm{tr}(A)^2 - 4\det(A)} < 2$, or $(\mathrm{tr}(A) - 2)^2 < \mathrm{tr}(A)^2 - 4\det(A)$, which can be confirmed to be always true by very simple calculations.

Next, we wish to generalize the previous examples to the case $d > 2$. The counterpart of the saddlepoint seen before is called the *convergent*, or *stable subspace*: it is the locus of points for which the solution does not explode. (In the case of nonlinear systems, such a convergent subspace is termed *convergent*, or *stable manifold*. In this appendix we only study linear systems.)

Let $\Pi \equiv P^{-1}$, and rewrite the system determining the solution for κ:

$$\hat{\kappa} = \Pi z_0.$$

We suppose that the elements of z and matrix A have been reordered in such a way that $\exists s : |\lambda_i| < 1$, for $i = 1, \cdots, s$ and $|\lambda_i| > 1$ for $i = s+1, \cdots, d$. Then we partition Π in such a way that:

$$\hat{\kappa} = \begin{pmatrix} \underset{s\times d}{\Pi_{\overline{s}}} \\ \underset{(d-s)\times d}{\Pi_u} \end{pmatrix} z_0.$$

As in example (3A.1), the objective is to make the system "stay prisoner" of the convergent space, which requires that

$$\hat{\kappa}_{s+1} = \cdots = \hat{\kappa}_d = 0,$$

or, by exploiting the previous system,

$$\begin{pmatrix} \hat{\kappa}_{s+1} \\ \vdots \\ \hat{\kappa}_d \end{pmatrix} = \underset{(d-s)\times d}{\Pi_u} z_0 = \mathbf{0}_{(d-s)\times 1}.$$

Let $d \equiv k + k^*$ (k free and k^* predetermined), and partition Π_u and z_0 in such a way to distinguish the predetermined from the free variables:

$$\mathbf{0}_{(d-s)\times 1} = \underset{(d-s)\times d}{\Pi_u} z_0 = \begin{pmatrix} \underset{(d-s)\times k}{\Pi_u^{(1)}} & \underset{(d-s)\times k^*}{\Pi_u^{(2)}} \end{pmatrix} \begin{pmatrix} \underset{k\times 1}{z_0^{\mathrm{free}}} \\ \underset{k^*\times 1}{z_0^{\mathrm{pre}}} \end{pmatrix} = \underset{(d-s)\times k}{\Pi_u^{(1)}} \underset{k\times 1}{z_0^{\mathrm{free}}} + \underset{(d-s)\times k^*}{\Pi_u^{(2)}} \underset{k^*\times 1}{z_0^{\mathrm{pre}}},$$

or,

$$\underset{(d-s)\times k}{\Pi_u^{(1)}} \underset{k\times 1}{z_0^{\mathrm{free}}} = - \underset{(d-s)\times k^*}{\Pi_u^{(2)}} \underset{k^*\times 1}{z_0^{\mathrm{pre}}}.$$

The previous system has $d - s$ equations and k unknowns (the components of z_0^{free}): this is so because z_0^{pre} is known (it is the k^*-dimensional vector of the predetermined variables) and $\Pi_u^{(1)}, \Pi_u^{(2)}$ are primitive data of the economy (they depend on A). We assume that $\underset{(d-s)\times k}{\Pi_u^{(1)}}$ has full rank.

Therefore, there are three cases: 1) $s = k^*$; 2) $s < k^*$; and 3) $s > k^*$. Before analyzing these case, let us mention a word on terminology. We shall refer to s as the dimension of the convergent subspace ($\mathcal{S}$). The reason is the following one. Consider the solution:

$$z_t = v_1 \hat{\kappa}_1 \lambda_1^t + \cdots + v_s \hat{\kappa}_s \lambda_s^t + v_{s+1} \hat{\kappa}_{s+1} \lambda_{s+1}^t + \cdots + v_d \hat{\kappa}_d \lambda_d^t.$$

In order to be in $\mathcal{S}$, it must be the case that

$$\hat{\kappa}_{s+1} = \cdots = \hat{\kappa}_d = 0,$$

in which case the solution reduces to:

$$\underset{d\times 1}{z_t} = \underset{d\times 1}{v_1}\hat{\kappa}_1\lambda_1^t + \cdots + \underset{d\times 1}{v_s}\hat{\kappa}_s\lambda_s^t = (v_1\hat{\kappa}_1, \cdots, v_s\hat{\kappa}_s)\cdot\left(\lambda_1^t, \cdots, \lambda_s^t\right)^\top,$$

i.e.,

$$\underset{d\times 1}{z_t} = \underset{d\times s}{\hat{V}}\cdot\underset{s\times 1}{\hat{\lambda}_t},$$

where $\hat{V} \equiv (v_1\hat{\kappa}_1, \cdots, v_s\hat{\kappa}_s)$ and $\hat{\lambda}_t \equiv \left(\lambda_1^t, \cdots, \lambda_s^t\right)^{\mathrm{T}}$. Now for each t, introduce the vector subspace:

$$\left\langle \hat{V} \right\rangle_t \equiv \{z_t \in \mathbb{R}^d : \underset{d\times 1}{z_t} = \underset{d\times s}{\hat{V}}\cdot\underset{s\times 1}{\hat{\lambda}_t}, \quad \hat{\lambda}_t \in \mathbb{R}^s\}.$$

Clearly, for each t, $\dim\left\langle \hat{V} \right\rangle_t = \operatorname{rank}\left(\hat{V}\right) = s$, and we are done.

Let us analyze now the three above mentioned cases.

- $d-s=k$, or $s=k^*$. The dimension of the divergent subspace is equal to the number of the free variables or, the dimension of the convergent subspace is equal to the number of predetermined variables. In this case, the system is determined. The previous conditions are easy to interpret. The predetermined variables identify one and only one point in the convergent space, which allows us to compute the only possible jump in correspondence of which the free variables can jump to make the system remain in the convergent space: $z_0^{\mathrm{free}} = -\Pi_u^{(1)-1}\Pi_u^{(2)}z_0^{\mathrm{pre}}$. This is exactly the case of the previous examples, in which $d=2$, $k=1$, and the predetermined variable was x: there x_0 identified one and only one point in the saddlepoint path, and starting from such a point, there was one and only one y_0 guaranteeing that the system does not explode.

- $d-s>k$, or $s<k^*$. There are generically no solutions in the convergent space. This case was already reminded at the end of example 4A.1.

- $d-s<k$, or $s>k^*$. There exists an infinite number of solutions in the convergent space, and such a phenomenon is typically referred to as *indeterminacy*. In the previous example, $s=1$, and this case may emerge only in the absence of predetermined variables. This is also the case in which *sunspots* may arise.

3.9 Appendix 2: Neoclassic growth model - continuous time

3.9.1 Convergence results

Consider chopping time in the population growth law as,

$$N_{hk} - N_{h(k-1)} = \bar{n} \cdot h \cdot N_{h(k-1)}, \quad k = 1, \cdots, \ell,$$

where $\bar{n}$ is an *instantaneous rate*, and $\ell = \frac{t}{h}$ is the number of subperiods in which we have chopped a given time period t. The solution is $N_{h\ell} = (1 + \bar{n}h)^{\ell} N_0$, or

$$N_t = (1 + \bar{n} \cdot h)^{t/h} N_0.$$

By taking limits:

$$N(t) = \lim_{h \downarrow 0} (1 + \bar{n} \cdot h)^{t/h} N(0) = e^{\bar{n}t} N(0).$$

On the other hand, an exact discretization yields:

$$N(t - \Delta) = e^{\bar{n}(t-\Delta)} N(0).$$

$\Leftrightarrow$

$$\frac{N(t)}{N(t-\Delta)} = e^{\bar{n}\Delta} \equiv 1 + n_{\Delta}.$$

$\Leftrightarrow$

$$\bar{n} = \frac{1}{\Delta} \log (1 + n_{\Delta}).$$

E.g., $\Delta = 1$, $n_{\Delta} = n_1 \equiv n : \bar{n} = \log(1 + n)$.

Now let's try to do the same thing for the capital accumulation law:

$$K_{h(k+1)} = \left(1 - \bar{\delta} \cdot h\right) K_{hk} + I_{h(k+1)} \cdot h, \quad k = 0, \cdots, \ell - 1,$$

where $\bar{\delta}$ is an instantaneous rate.

By iterating on as in the population growth case we get:

$$K_{h\ell} = \left(1 - \bar{\delta} \cdot h\right)^{\ell} K_0 + \sum_{j=1}^{\ell} \left(1 - \bar{\delta} \cdot h\right)^{\ell - j} I_{hj} \cdot h, \quad \ell = \frac{t}{h},$$

or

$$K_t = \left(1 - \bar{\delta} \cdot h\right)^{t/h} K_0 + \sum_{j=1}^{t/h} \left(1 - \bar{\delta} \cdot h\right)^{t/h - j} I_{hj} \cdot h,$$

As $h \downarrow 0$ we get:

$$K(t) = e^{-\bar{\delta}t} K_0 + e^{-\bar{\delta}t} \int_0^t e^{\bar{\delta}u} I(u) du,$$

or in differential form:

$$\dot{K}(t) = -\bar{\delta} K(t) + I(t),$$

and starting from the IS equation:

$$Y(t) = C(t) + I(t),$$

we obtain the capital accumulation law:

$$\dot{K}(t) = Y(t) - C(t) - \bar{\delta} K(t).$$

Discretization issues

An exact discretization gives:

$$K(t+1) = e^{-\bar{\delta}}K(t) + e^{-\bar{\delta}(t+1)} \int_t^{t+1} e^{\bar{\delta}u} I(u)du.$$

By identifying with the standard capital accumulation law in the discrete time setting:

$$K_{t+1} = (1-\delta) K_t + I_t,$$

we get:

$$\bar{\delta} = \log \frac{1}{(1-\delta)}.$$

It follows that

$$\delta \in (0,1) \Rightarrow \bar{\delta} > 0 \text{ and } \delta = 0 \Rightarrow \bar{\delta} = 0.$$

Hence, while δ can take on only values on $[0,1)$, $\bar{\delta}$ can take on values on the entire real line.

An important restriction arises in the continuous time model when we note that:

$$\lim_{\delta \to 1-} \log \frac{1}{(1-\delta)} = \infty,$$

It is impossible to think about a "maximal rate of capital depreciation" in a continuous time model because this would imply an infinite depreciation rate!

Finally, substitute δ into the exact discretization (?.?):

$$K(t+1) = (1-\delta) K(t) + e^{-\bar{\delta}(t+1)} \int_t^{t+1} e^{\bar{\delta}u} I(u)du$$

so that we have to interpret investments in $t+1$ as $e^{-\bar{\delta}(t+1)} \int_t^{t+1} e^{\bar{\delta}u} I(u)du$.

Per capita dynamics

Consider dividing the capital accumulation equation by $N(t)$:

$$\frac{\dot{K}(t)}{N(t)} = \frac{F(K(t), N(t))}{N(t)} - c(t) - \bar{\delta}k(t) = y(k(t)) - c(t) - \bar{\delta}k(t).$$

By plugging the relationship $\dot{k} = d(K(t)/N(t)) = \dot{K}(t) \big/ N(t) - \bar{n}k(t)$ into the previous equation we get:

$$\dot{k} = y(k(t)) - c(t) - \left(\bar{\delta} + \bar{n}\right) \cdot k(t)$$

This is the constraint we use in the problem of the following subsection.

3.9.2 The model itself

We consider directly the social planner problem. The program is:

$$\begin{cases} \max_c \int_0^\infty e^{-\rho t} u(c(t))dt \\ \text{s.t. } \dot{k}(t) = y(k(t)) - c(t) - \left(\bar{\delta} + \bar{n}\right) \cdot k(t) \end{cases} \tag{3A2.1}$$

where all variables are expressed in per-capita terms. We suppose that there is no capital depreciation (in the discrete time model, we supposed a total capital depreciation). More general results can be obtained with just a change in notation.

The Hamiltonian is,

$$H(t) = u(c(t)) + \lambda(t)\left[y(k(t)) - c(t) - \left(\bar{\delta} + \bar{n}\right) \cdot k(t)\right],$$

where λ is a co-state variable.

As explained in appendix 4 of the present chapter, the first order conditions for this problem are:

$$\begin{cases} 0 & = & \dfrac{\partial H}{\partial c}(t) & \Leftrightarrow & \lambda(t) & = & u'(c(t)) \\ \dfrac{\partial H}{\partial \lambda}(t) & = & \dot{k}(t) \\ \dfrac{\partial H}{\partial k}(t) & = & -\dot{\lambda}(t) + \rho\lambda(t) & \Leftrightarrow & \dot{\lambda}(t) & = & \left[\rho + \bar{\delta} + \bar{n} - y'(k(t))\right]\lambda(t) \end{cases} \tag{3A2.2}$$

By differentiating the first equation in (4A2.2) we get:

$$\dot{\lambda}(t) = \left(\frac{u''(c(t))}{u'(c(t))}\dot{c}(t)\right)\lambda(t).$$

By identifying with the help of the last equation in (4A2.2),

$$\dot{c}(t) = \frac{u'(c(t))}{u''(c(t))}\left[\rho + \bar{\delta} + \bar{n} - y'(k(t))\right]. \tag{3A2.3}$$

The equilibrium is the solution of the system consisting of the constraint of (4A2.1), and (4A2.3). As in section 3.2.3, here we analyze the equilibrium dynamics of the system in a small neighborhood of the stationary state.[24] Denote the stationary state as the solution (c, k) of the constraint of program (4A2.1), and (4A2.3) when $\dot{c}(t) = \dot{k}(t) = 0$,

$$\begin{cases} c & = & y(k) - \left(\bar{\delta} + \bar{n}\right)k \\ \rho + \bar{\delta} + \bar{n} & = & y'(k) \end{cases}$$

Warning! these are instantaneous figures, so that don't worry if they are not such that $y'(k) \geq 1 + n$**!**. A first-order Taylor approximation of both sides of the constraint of program (4A2.1) and (4A2.3) near (c, k) yields:

$$\begin{cases} \dot{c}(t) & = & -\dfrac{u'(c)}{u''(c)}y''(k)\left(k(t) - k\right) \\ \dot{k}(t) & = & \rho \cdot (k(t) - k) - (c(t) - c) \end{cases}$$

where we used the equality $\rho + \bar{\delta} + \bar{n} = y'(k)$. By setting $x(t) \equiv c(t) - c$ and $y(t) \equiv k(t) - k$ the previous system can be rewritten as:

$$\dot{z}(t) = A \cdot z(t), \tag{3A2.4}$$

where $z \equiv (x, y)^T$, and

$$A \equiv \begin{pmatrix} 0 & -\dfrac{u'(c)}{u''(c)}y''(k) \\ -1 & \rho \end{pmatrix}.$$

Warning! There must be some mistake somewhere. Let us diagonalize system (4A2.4) by setting $A = P\Lambda P^{-1}$, where P and Λ have the same meaning as in the previous appendix. We have:

$$\dot{\nu}(t) = \Lambda \cdot \nu(t),$$

[24] In addition to the theoretical results that are available in the literature, the general case can also be treated numerically with the tools surveyed by Judd (1998).

where $\nu \equiv P^{-1}z$.

The eigenvalues are solutions of the following quadratic equation:

$$0 = \lambda^2 - \rho\lambda - \frac{u'(c)}{u''(c)}y''(k). \tag{3A2.5}$$

We see that $\lambda_1 < 0 < \lambda_2$, and $\lambda_1 \equiv \frac{\rho}{2} - \frac{1}{2}\sqrt{\rho^2 + 4\frac{u'(c)}{u''(c)}y''(k)}$. The solution for $\nu(t)$ is:

$$\nu_i(t) = \kappa_i e^{\lambda_i t},\ i = 1, 2,$$

whence

$$z(t) = P \cdot \nu(t) = v_1\kappa_1 e^{\lambda_1 t} + v_2\kappa_2 e^{\lambda_2 t},$$

where the v_is are 2×1 vectors. We have,

$$\begin{cases} x(t) &= v_{11}\kappa_1 e^{\lambda_1 t} + v_{12}\kappa_2 e^{\lambda_2 t} \\ y(t) &= v_{21}\kappa_1 e^{\lambda_1 t} + v_{22}\kappa_2 e^{\lambda_2 t} \end{cases}$$

Let us evaluate this solution in $t = 0$,

$$\begin{pmatrix} x(0) \\ y(0) \end{pmatrix} = P\kappa = \begin{pmatrix} v_1 & v_2 \end{pmatrix} \begin{pmatrix} \kappa_1 \\ \kappa_2 \end{pmatrix}.$$

By repeating verbatim the reasoning of the previous appendix,

$$\kappa_2 = 0 \Leftrightarrow \frac{y(0)}{x(0)} = \frac{v_{21}}{v_{11}}.$$

As in the discrete time model, the saddlepoint path is located along a line that has as a slope the ratio of the components of the eigenvector associated with the negative root. We can explicitely compute such ratio. By definition, $A \cdot v_1 = \lambda_1 v_1 \Leftrightarrow$

$$\begin{cases} -\frac{u'(c)}{u''(c)}y''(k) &= \lambda_1 v_{11} \\ -v_{11} + \rho v_{21} &= \lambda_1 v_{21} \end{cases}$$

i.e., $\frac{v_{21}}{v_{11}} = -\frac{\lambda_1}{\frac{u'(c)}{u''(c)}y''(k)}$ and simultaneously, $\frac{v_{21}}{v_{11}} = \frac{1}{\rho - \lambda_1}$, which can be verified with the help of (3A2.5).

4
Continuous time models

4.1 Lambdas and betas in continuous time

4.1.1 The pricing equation

Let q_t be the price of a long lived asset as of time t, and let D_t the dividend paid by the asset at time t. In the previous chapters, we learned that in the absence of arbitrage opportunities,

$$q_t = E_t \left[m_{t+1} \left(q_{t+1} + D_{t+1} \right) \right], \tag{4.1}$$

where E_t is the conditional expectation given the information set at time t, and m_{t+1} is the usual stochastic discount factor.

Next, let us introduce the pricing kernel, or state-price process,

$$\xi_{t+1} = m_{t+1} \xi_t, \quad \xi_0 = 1.$$

In terms of the pricing kernel ξ, Eq. (4.1) is,

$$0 = E_t \left(\xi_{t+1} q_{t+1} - q_t \xi_t \right) + E_t \left(\xi_{t+1} D_{t+1} \right).$$

For small trading periods h, this is,

$$0 = E_t \left(\xi_{t+h} q_{t+h} - q_t \xi_t \right) + E_t \left(\xi_{t+h} D_{t+h} h \right).$$

As $h \downarrow 0$,

$$0 = E_t \left[d \left(\xi(t) q(t) \right) \right] + \xi(t) D(t) dt. \tag{4.2}$$

Eq. (4.2) can now be integrated to yield,

$$\xi(t) q(t) = E_t \left[\int_t^T \xi(u) D(u) du \right] + E_t \left[\xi(T) q(T) \right].$$

Finally, let us assume that $\lim_{T \to \infty} E_t \left[\xi(T) q(T) \right] = 0$. Then, provided it exists, the price q_t of an infinitely lived asset price satisfies,

$$\xi(t) q(t) = E_t \left[\int_t^\infty \xi(u) D(u) du \right]. \tag{4.3}$$

4.1.2 Expected returns

Let us elaborate on Eq. (4.2). We have,

$$d(\xi q) = qd\xi + \xi dq + d\xi dq = \xi q\left(\frac{d\xi}{\xi} + \frac{dq}{q} + \frac{d\xi}{\xi}\frac{dq}{q}\right).$$

By replacing this expansion into Eq. (4.2) we obtain,

$$E_t\left(\frac{dq}{q}\right) + \frac{D}{q}dt = -E_t\left(\frac{d\xi}{\xi}\right) - E_t\left(\frac{d\xi}{\xi}\frac{dq}{q}\right). \tag{4.4}$$

This evaluation equation holds for any asset and, hence, for the assets that do not distribute dividends and are locally riskless, i.e. $D = 0$ and $E_t\left(\frac{d\xi}{\xi}\frac{dq_0}{q_0}\right) = 0$, where $q_0(t)$ is the price of these locally riskless assets, supposed to satisfy $\frac{dq_0(t)}{q_0(t)} = r(t)\,dt$, for some short term rate process r_t. By Eq. (4.4), then,

$$E_t\left(\frac{d\xi}{\xi}\right) = -r(t)\,dt.$$

By replacing this into eq. (4.4) leaves the following representation for the expected returns $E_t\left(\frac{dq}{q}\right) + \frac{D}{q}dt$,

$$E_t\left(\frac{dq}{q}\right) + \frac{D}{q}dt = rdt - E_t\left(\frac{d\xi}{\xi}\frac{dq}{q}\right). \tag{4.5}$$

In a diffusion setting, Eq. (4.5) gives rise to a partial differential equation. Moreover, in a diffusion setting,

$$\frac{d\xi}{\xi} = -rdt - \lambda \cdot dW,$$

where W is a vector Brownian motion, and λ is the vector of unit risk-premia. Naturally, the price of the asset, q, is driven by the same Brownian motions driving r and ξ. We have,

$$E_t\left(\frac{d\xi}{\xi}\frac{dq}{q}\right) = -\text{Vol}\left(\frac{dq}{q}\right)\cdot \lambda dt,$$

which leaves,

$$E_t\left(\frac{dq}{q}\right) + \frac{D}{q}dt = rdt + \underbrace{\text{Vol}\left(\frac{dq}{q}\right)}_{\text{“betas”}} \cdot \underbrace{\lambda}_{\text{“lambdas”}} dt.$$

4.1.3 Expected returns and risk-adjusted discount rates

The difference between expected returns and risk-adjusted discount rates is subtle. If dividends and asset prices are driven by only one factor, expected returns and risk-adjusted discount rates are the same. Otherwise, we have to make a distinction. To illustrate the issue, let us make a simplification. Suppose that the price of the asset q takes the following form,

$$q(y, D) = p(y)\,D, \tag{4.6}$$

where y is a vector of state variables that are suggested by economic theory. In other words, we assume that the price-dividend ratio p is independent of the dividends D. Indeed, this "scale-invariant" property of asset prices arises in many model economies, as we shall discuss in detail in the second part of these Lectures. By Eq. (4.6),

$$\frac{dq}{q} = \frac{dp}{p} + \frac{dD}{D}.$$

By replacing the previous expansion into Eq. (4.5) leaves,

$$E_t\left(\frac{dq}{q}\right) + \frac{D}{q}dt = \text{Disc } dt - E_t\left(\frac{dp}{p}\frac{d\xi}{\xi}\right). \tag{4.7}$$

where we define the "risk adjusted discount rate", Disc, to be:

$$\text{Disc} = r - E_t\left(\frac{dD}{D}\frac{d\xi}{\xi}\right) \Big/ dt.$$

If the price-dividend ratio, p, is constant, the "risk adjusted discount rate" Disc has the usual interpretation. It equals the safe interest rate r, plus the premium $-E_t\left(\frac{dD}{D}\frac{d\xi}{\xi}\right)$ that arises to compensate the agents for the fluctuations of the uncertain flow of future dividends. This premium equals,

$$E_t\left(\frac{dD}{D}\frac{d\xi}{\xi}\right) = -\underbrace{\text{Vol}\left(\frac{dD}{D}\right)}_{\text{"cash-flow betas"}} \cdot \underbrace{\lambda}_{\text{"lambdas"}} dt.$$

In this case, expected returns and risk-adjusted discount rates are the same thing, as in the simple Lucas economy with one factor examined in Section 2.

However, if the price-dividend ratio is not constant, the last term in Eq. (4.7) introduces a wedge between expected returns and risk-adjusted discount rates. As we shall see, the risk-adjusted discount rates play an important role in explaining returns *volatility*, i.e. the beta related to the fluctuations of the price-dividend ratio. Intuitively, this is because risk-adjusted discount rates affect prices through rational evaluation and, hence, price-dividend ratios and price-dividend ratios volatility. To illustrate these properties, note that Eq. (5.6) can be rewritten as,

$$p(y(t)) = \mathbb{E}_t\left[\int_t^\infty \frac{D_*(\tau)}{D(t)} \cdot e^{-\int_t^\tau \text{Disc}(y(u))du} \middle| y(t)\right], \tag{4.8}$$

where the expectation is taken under the risk-neutral probability, but the expected dividend growth $\frac{D_*(\tau)}{D(t)}$ is *not* risk-adjusted (that is $\mathbb{E}(\frac{D_*(\tau)}{D(t)}) = e^{g_0(\tau-t)}$). Eq. (4.8) reveals that risk-adjusted discount rates play an important role in shaping the price function p and, hence, the volatility of the price-dividend ratio p. These points are developed in detail in Chapter 6.

4.2 An introduction to arbitrage and equilibrium in continuous time models

4.2.1 A "reduced-form" economy

Let us consider the Lucas (1978) model with one tree and a perishable good taken as the numéraire. The Appendix shows that in continuous time, the wealth process of a representative agent at time τ, $V(\tau)$, is solution to,

$$dV(\tau) = \left(\frac{dq(\tau)}{q(\tau)} + \frac{D(\tau)}{q(\tau)}d\tau - rd\tau\right)\pi(\tau) + rVd\tau - c(\tau)d\tau, \tag{4.9}$$

where $D(\tau)$ is the dividend process, $\pi \equiv q\theta^{(1)}$, and $\theta^{(1)}$ is the number of trees in the portfolio of the representative agent.

We assume that the dividend process, $D(\tau)$, is solution to the following stochastic differential equation,

$$\frac{dD}{D} = \mu_D d\tau + \sigma_D dW,$$

for two positive constants μ_D and σ_D. Under rational expectation, the price function q is such that $q = q(D)$. By Itô's lemma,

$$\frac{dq}{q} = \mu_q d\tau + \sigma_q dW,$$

where

$$\mu_q = \frac{\mu_D Dq'(D) + \frac{1}{2}\sigma_D^2 D^2 q''(D)}{q(D)}; \quad \sigma_q = \frac{\sigma_D Dq'(D)}{q(D)}.$$

Then, by Eq. (4.9), the value of wealth satisfies,

$$dV = \left[\pi\left(\mu_q + \frac{D}{q} - r\right) + rV - c\right]d\tau + \pi\sigma_q dW.$$

Below, we shall show that in the absence of arbitrage, there must be some process λ, the "unit risk-premium", such that,

$$\mu_q + \frac{D}{q} - r = \lambda\sigma_q. \tag{4.10}$$

Let us assume that the short-term rate, r, and the risk-premium, λ, are both constant. Below, we shall show that such an assumption is compatible with a general equilibrium economy. By the definition of μ_q and σ_q, Eq. (4.10) can be written as,

$$0 = \frac{1}{2}\sigma_D^2 D^2 q''(D) + (\mu_D - \lambda\sigma_D) Dq'(D) - rq(D) + D. \tag{4.11}$$

Eq. (4.11) is a second order differential equation. Its solution, provided it exists, is the the rational price of the asset. To solve Eq. (4.11), we initially assume that the solution, q_{F} say, tales the following simple form,

$$q_{\mathrm{F}}(D) = K \cdot D, \tag{4.12}$$

where K is a constant to be determined. Next, we verify that this is indeed *one* solution to Eq. (4.11). Indeed, if Eq. (9.30) holds then, by plugging this guess and its derivatives into Eq. (4.11) leaves, $K = (r - \mu_D + \lambda\sigma_D)^{-1}$ and, hence,

$$q_{\mathrm{F}}(D) = \frac{1}{r + \lambda\sigma_D - \mu_D} D. \tag{4.13}$$

This is a Gordon-type formula. It merely states that prices are risk-adjusted expectations of future expected dividends, where the risk-adjusted discount rate is given by $r + \lambda\sigma_D$. Hence, in a comparative statics sense, stock prices are inversely related to the risk-premium, a quite intuitive conclusion.

Eq. (4.13) can be thought to be the Feynman-Kac representation to the Eq. 4.11), viz

$$q_{\mathrm{F}}(D(t)) = \mathbb{E}_t \left[\int_t^{\infty} e^{-r(\tau - t)} D(\tau)\, d\tau \right], \tag{4.14}$$

where $\mathbb{E}_t[\cdot]$ is the conditional expectation taken under the risk neutral probability Q (say), the dividend process follows,

$$\frac{dD}{D} = (\mu_D - \lambda\sigma_D)\, d\tau + \sigma_D d\tilde{W},$$

and $\tilde{W}(\tau) = W(\tau) + \lambda(\tau - t)$ is a another standard Brownian motion defined under Q. Formally, the true probability, P, and the risk-neutral probability, Q, are tied up by the Radon-Nikodym derivative,

$$\eta = \frac{dQ}{dP} = e^{-\lambda(W(\tau) - W(t)) - \frac{1}{2}\lambda^2(\tau - t)}.$$

4.2.2 Preferences and equilibrium

The previous results do not let us see, precisely, how preferences affect asset prices. In Eq. (4.13), the asset price is related to the interest rate, r, and the risk-premium, λ. In equilibrium, the agents preferences affect the interest rate and the risk-premium. However, such an impact can have a non-linear pattern. For example, when the risk-aversion is low, a small change of risk-aversion can make the interest rate and the risk-premium change in the same direction. If the risk-aversion is high, the effects may be different, as the interest rate reflects a variety of factors, including precautionary motives.

To illustrate these points, let us rewrite, first, Eq. (4.9) under the risk-neutral probability Q. We have,

$$dV = (rV - c)\, d\tau + \pi\sigma_q d\tilde{W}. \tag{4.15}$$

We assume that the following transversality condition holds,

$$\lim_{\tau \to \infty} \mathbb{E}_t \left[e^{-r(\tau - t)} V(\tau) \right] = 0. \tag{4.16}$$

By integrating Eq. (4.15), and using the previous transversality condition,

$$V(t) = \mathbb{E}_t \left[\int_t^{\infty} e^{-r(\tau - t)} c(\tau)\, d\tau \right]. \tag{4.17}$$

By comparing Eq. (4.14) with Eq. (4.17) reveals that the equilibrium in the real markets, $D = c$, also implies that $q = V$. Next, rewrite (4.17) as,

$$V(t) = \mathbb{E}_t \left[\int_t^{\infty} e^{-r(\tau - t)} c(t) d\tau \right] = E_t \left[\int_t^{\infty} m_t(\tau) c(t) d\tau \right],$$

where

$$m_t(\tau) \equiv \frac{\xi(\tau)}{\xi(t)} = e^{-\left(r + \frac{1}{2}\lambda^2\right)(\tau - t) - \lambda(W(\tau) - W(t))}.$$

We assume that a representative agent solves the following intertemporal optimization problem,

$$\max_c E_t \left[\int_t^{\infty} e^{-\rho(\tau - t)} u(c(\tau))\, d\tau \right] \quad \text{s.t. } V(t) = E_t \left[\int_t^{\infty} m_t(\tau) c(\tau) d\tau \right] \tag{P1}$$

for some instantaneous utility function $u(c)$ and some subjective discount rate ρ.

To solve the program [P1], we form the Lagrangean

$$L = E_t \left[\int_t^\infty e^{-\rho(\tau-t)} u(c(\tau)) d\tau \right] + \ell \cdot \left[V(t) - E_t \left(\int_t^\infty m_t(\tau) c(\tau) d\tau \right) \right],$$

where ℓ is a Lagrange multiplier. The first order conditions are,

$$e^{-\rho(\tau-t)} u'(c(\tau)) = \ell \cdot m_t(\tau).$$

Moreover, by the equilibrium condition, $c = D$, and the definition of $m_t(\tau)$,

$$u'(D(\tau)) = \ell \cdot e^{-\left(r+\frac{1}{2}\lambda^2-\rho\right)(\tau-t)-\lambda(W(\tau)-W(t))}. \tag{4.18}$$

That is, by Itô's lemma,

$$\frac{du'(D)}{u'(D)} = \left[\frac{u''(D)D}{u'(D)} \mu_D + \frac{1}{2}\sigma_D^2 D^2 \frac{u'''(D)}{u'(D)} \right] d\tau + \frac{u''(D)D}{u'(D)} \sigma_D dW. \tag{4.19}$$

Next, let us define the right hand side of Eq. (7.24) as $U(\tau) \equiv \ell \cdot e^{-\left(r+\frac{1}{2}\lambda^2-\rho\right)(\tau-t)-\lambda(W(\tau)-W(t))}$. By Itô's lemma, again,

$$\frac{dU}{U} = (\rho - r)\, d\tau - \lambda dW. \tag{4.20}$$

By Eq. (7.24), drift and volatility components of Eq. (4.19) and Eq. (4.20) have to be the same. This is possible if

$$r = \rho - \frac{u''(D)\,D}{u'(D)} \mu_D - \frac{1}{2}\sigma_D^2 D^2 \frac{u'''(D)}{u'(D)}; \quad \text{and} \quad \lambda = -\frac{u''(D)\,D}{u'(D)} \sigma_D.$$

Let us assume that λ is constant. After integrating the second of these relations two times, we obtain that besides some irrelevant integration constant,

$$u(D) = \frac{D^{1-\eta} - 1}{1-\eta}, \quad \eta \equiv \frac{\lambda}{\sigma_D},$$

where η is the CRRA. Hence, under CRRA preferences we have that,

$$r = \rho + \eta\mu_D - \frac{\eta(\eta+1)}{2}\sigma_D^2; \quad \lambda = \eta\sigma_D.$$

Finally, by replacing these expressions for the short-term rate and the risk-premium into Eq. (4.13) leaves,

$$q(D) = \frac{1}{\rho - (1-\eta)\left(\mu_D - \frac{1}{2}\eta\sigma_D^2\right)} D,$$

provided the following conditions holds true:

$$\rho > (1-\eta)\left(\mu_D - \frac{1}{2}\eta\sigma_D^2\right). \tag{4.21}$$

We are only left to check that the transversality condition (4.16) holds at the equilibrium $q = V$. We have that under the previous inequality,

$$\begin{aligned}
\lim_{\tau\to\infty} \mathbb{E}_t\left[e^{-r(\tau-t)}V(\tau)\right] &= \lim_{\tau\to\infty} \mathbb{E}_t\left[e^{-r(\tau-t)}q(\tau)\right] \\
&= \lim_{\tau\to\infty} E_t\left[m_t(\tau)q(\tau)\right] \\
&= \lim_{\tau\to\infty} E_t\left[e^{-\left(r+\frac{1}{2}\lambda^2\right)(\tau-t)-\lambda(W(\tau)-W(t))}q(\tau)\right] \\
&= q\left(t\right)\lim_{\tau\to\infty} E_t\left[e^{\left(\mu_D-\frac{1}{2}\sigma_D^2-r-\frac{1}{2}\lambda^2\right)(\tau-t)+(\sigma_D-\lambda)(W(\tau)-W(t))}\right] \\
&= q\left(t\right)\lim_{\tau\to\infty} e^{-(r-\mu_D+\sigma_D\lambda)(\tau-t)} \\
&= q\left(t\right)\lim_{\tau\to\infty} e^{-\left[\rho-(1-\eta)\left(\mu_D-\frac{1}{2}\eta\sigma_D^2\right)\right](\tau-t)} \\
&= 0. \qquad (4.22)
\end{aligned}$$

4.2.3 Bubbles

The transversality condition in Eq. (4.16) is often referred to as a *no-bubble condition.* To illustrate the reasons underlying this definition, note that Eq. (4.11) admits an infinite number of solutions. Each of these solutions takes the following form,

$$q(D) = KD + AD^{\delta}, \quad K, A, \delta \text{ constants.} \qquad (4.23)$$

Indeed, by plugging Eq. (4.23) into Eq. (4.11) reveals that Eq. (4.23) holds if and only if the following conditions holds true:

$$0 = K\left(r + \lambda\sigma_D - \mu_D\right) - 1; \quad \text{and} \quad 0 = \delta\left(\mu_D - \lambda\sigma_D\right) + \frac{1}{2}\delta\left(\delta - 1\right)\sigma_D^2 - r. \qquad (4.24)$$

The first condition implies that K equals the price-dividend ratio in Eq. (4.13), i.e. $K = q_{\mathrm{F}}\left(D\right)/D$. The second condition leads to a quadratic equation in δ, with the two solutions,

$$\delta_1 < 0 \qquad \text{and} \qquad \delta_2 > 0.$$

Therefore, the asset price function takes the following form:

$$q(D) = q_{\mathrm{F}}\left(D\right) + A_1 D^{\delta_1} + A_2 D^{\delta_2}.$$

It satisfies:

$$\lim_{D\to 0} q\left(D\right) = \mp\infty \quad \text{if} \quad A_1 \lesseqgtr 0 \quad ; \quad \lim_{D\to 0} q\left(D\right) = 0 \quad \text{if} \quad A_1 = 0.$$

To rule out an explosive behavior of the price as the dividend level, D, gets small, we must set $A_1 = 0$, which leaves,

$$q\left(D\right) = q_{\mathrm{F}}\left(D\right) + \mathcal{B}\left(D\right); \qquad \mathcal{B}\left(D\right) \equiv A_2 D^{\delta_2}. \qquad (4.25)$$

The component, $q_{\mathrm{F}}(D)$, is the *fundamental value of the asset*, as by Eq. (4.14), it is the risk-adjusted present value of the expected dividends. The second component, $\mathcal{B}(D)$, is simply the difference between the market value of the asset, $q(D)$, and the fundamental value, $q_{\mathrm{F}}(D)$. Hence, it is a bubble.

We seek conditions under which Eq. (4.25) satisfies the transversality condition in Eq. (4.16). We have,

$$\lim_{\tau\to\infty}\mathbb{E}_t\left[e^{-r(\tau-t)}q(\tau)\right]=\lim_{\tau\to\infty}\mathbb{E}_t\left[e^{-r(\tau-t)}q_{\mathrm{F}}(D(\tau))\right]+\lim_{\tau\to\infty}\mathbb{E}_t\left[e^{-r(\tau-t)}\mathcal{B}(D(\tau))\right].$$

By Eq. (4.22), the fundamental value of the asset satisfies the transversality condition, under the condition given in Eq. (4.21). As regards the bubble, we have,

$$\begin{aligned}\lim_{\tau\to\infty}\mathbb{E}_t\left[e^{-r(\tau-t)}\mathcal{B}(D(\tau))\right]&=A_2\cdot\lim_{\tau\to\infty}\mathbb{E}_t\left[e^{-r(\tau-t)}D(\tau)^{\delta_2}\right]\\&=A_2\cdot D(t)^{\delta_2}\cdot\lim_{\tau\to\infty}\mathbb{E}_t\left[e^{(\delta_2(\mu_D-\lambda\sigma_D)+\frac{1}{2}\delta_2(\delta_2-1)\sigma_D^2-r)(\tau-t)}\right]\\&=A_2\cdot D(t)^{\delta_2},\end{aligned}\tag{4.26}$$

where the last line holds as δ_2 satisfies the second condition in Eq. (4.24). Therefore, the bubble can not satisfy the transversality condition, except in the trivial case in which $A_2=0$. In other words, in this economy, the transversality condition in Eq. (4.16) holds if and only if there are no bubbles.

4.2.4 Reflecting barriers and absence of arbitrage

Next, suppose that insofar as the dividend level, $D(\tau)$, wanders above a certain level $\underline{D}>0$, everything goes as in the previous section but that, as soon as the dividends level hits a "barrier" $\underline{D}$, it is "reflected" back with probability one. In this case, we say that the dividend follows a process with *reflecting barriers*. How does the price behave in the presence of such a barrier? First, if the dividend is above the barrier, $D>\underline{D}$, the price is still as in Eq. (4.23),

$$q(D)=\frac{1}{r-\mu_D+\lambda\sigma_D}D+A_1D^{\delta_1}+A_2D^{\delta_2}.$$

First, and as in the previous section, we need to set $A_2=0$ to satisfy the transversality condition in Eq. (4.16) (see Eq. (4.26)). However, in the new context of this section, we do not need to set $A_1=0$. Rather, this constant is needed to pin down the behavior of the price function $q(D)$ in the neighborhood of the barrier $\underline{D}$.

We claim that the following *smooth pasting* condition must hold in the neighborhood of $\underline{D}$,

$$q'(\underline{D})=0.\tag{4.27}$$

This condition is in fact a no-arbitrage condition. Indeed, after hitting the barrier $\underline{D}$, the dividend is reflected back for the part exceeding $\underline{D}$. Since the reflection takes place with probability one, the asset is locally riskless at the barrier $\underline{D}$. However, the dynamics of the asset price is,

$$\frac{dq}{q}=\mu_q d\tau+\underbrace{\frac{\sigma_D Dq'}{q}}_{\sigma_q}dW.$$

Therefore, the local risklessness of the asset at $\underline{D}$ is ensured if $q'(\underline{D}) = 0$. [**Warning: We need to add some local time component here.**] Furthermore, rewrite Eq. (4.10) as,

$$\mu_q + \frac{D}{q} - r = \lambda\sigma_q = \lambda\frac{\sigma_D D q'(D)}{q(D)}.$$

If $D = \underline{D}$ then, by Eq. (4.27), $q'(\underline{D}) = 0$. Therefore,

$$\mu_q + \frac{\underline{D}}{q} = r.$$

This relations tells us that holding the asset during the reflection guarantees a total return equal to the short-term rate. This is because during the reflection, the asset is locally riskless and, hence, arbitrage opportunities are ruled out when holding the asset will make us earn no more than the safe interest rate, r. Indeed, by previous relation into the wealth equation (4.9), and using the condition that $\sigma_q = 0$, we obtain that

$$dV = \left[\pi\left(\mu_q + \frac{\underline{D}}{q} - r\right) + rV - c\right]d\tau + \pi\sigma_q dW = (rV - c)\,d\tau.$$

This example illustrates how the relation in Eq. (4.10) works to preclude arbitrage opportunities.

Finally, we solve the model. We have, $K \equiv q_{\text{F}}(D)/D$, and

$$0 = q'(\underline{D}) = K + \delta_1 A_1 \underline{D}^{\delta_1 - 1}; \quad \underline{Q} \equiv q(\underline{D}) = K\underline{D} + A_1\underline{D}^{\delta_1},$$

where the second condition is the *value matching condition*, which needs to be imposed to ensure continuity of the pricing function w.r.t. D and, hence absence of arbitrage. The previous system can be solved to yield[1]

$$\underline{Q} = \frac{1-\delta_1}{-\delta_1}K\underline{D} \quad \text{and} \quad A_1 = \frac{K}{-\delta_1}\underline{D}^{1-\delta_1}.$$

Note, the price is an increasing and convex function of the fundamentals, D.

4.3 Martingales and arbitrage in a general diffusion model

4.3.1 The information framework

The time horizon of the economy is $T < \infty$, and the primitives include a probability space $(\Omega, \mathcal{F}, P)$. Let

$$W = \left\{W(\tau) = (W_1(\tau), \cdots, W_d(\tau))^\top\right\}_{\tau\in[t,T]}$$

be a standard Brownian motion in $\mathbb{R}^d$. Define $\mathbb{F} = \{\mathcal{F}(t)\}_{t\in[0,T]}$ as the P-augmentation of the natural filtration

$$\mathcal{F}^W(\tau) = \sigma\left(W(s),\ s \le \tau\right),\ \tau \in [t, T],$$

[1] In this model, we take the barrier $\underline{D}$ as given. In other context, we might be interested in "controlling" the dividend D in such a way that as soon as the price, q, hits a level $\underline{Q}$, the dividend level D is activate to induce the price q to increase. The solution for $\underline{Q}$ reveals that this situation is possible when $\underline{D} = \frac{-\delta_1}{1-\delta_1}K^{-1}\underline{Q}$, where $\underline{Q}$ is an exogeneously given constant.

generated by W, with $\mathcal{F} = \mathcal{F}(T)$.

We still consider a Lucas' type economy, but work with a finite horizon (generalizations are easy) and suppose that there are m primitive assets ("trees") and an accumulation factor. These assets, and other "inside money" assets (i.e., assets in zero net supply) to be introduced later, are exchanged without frictions. The primitive assets give rights to fruits (or dividends) $D_i = \{D_i(\tau)\}_{\tau\in[t,T]}$, $i = 1, \cdots, m$, which are positive $\mathcal{F}(\tau)$-adapted bounded processes. The fruits constitute the numéraire. We consider an economy lasting from t to T. Agents know T and the market structure is as if some refugees end up in a isle and get one tree each as endowment. At the end, agents sell back the trees to the owners of the isle, go to another isle, and everything starts back from the beginning.

Let:

$$q_+ = \left\{ q_+(\tau) = (q_0(\tau), \cdots, q_m(\tau))^\top \right\}_{\tau\in[t,T]}$$

be the positive $\mathcal{F}(\tau)$-adapted bounded asset price process. Associate an identically zero dividend process to the accumulation factor, the price of which satisfies:

$$q_0(\tau) = \exp\left(\int_t^\tau r(u)du \right), \ \tau \in [t, T],$$

where $\{r(\tau)\}_{\tau\in[t,T]}$ is $\mathcal{F}(\tau)$-adapted process satisfying $E\left(\int_t^T r(\tau)du\right) < \infty$. We assume that the dynamics of the last components of q_+, i.e. $q \equiv (q_1, \cdots, q_m)^\top$, is governed by:

$$dq_i(\tau) = q_i(\tau)\hat{a}_i(\tau)dt + q_i(\tau)\sigma_i(\tau)dW(\tau), \ i = 1, \cdots, m, \tag{4.28}$$

where $\hat{a}_i(\tau)$ and $\sigma_i(\tau)$ are processes satisfying the same properties as r, with $\sigma_i(\tau) \in \mathbb{R}^d$. We assume that $\operatorname{rank}(\sigma(\tau; \omega)) = m \le d$ a.s., where $\sigma(\tau) \equiv (\sigma_1(\tau), \cdots, \sigma_m(\tau))^\top$.

We assume that D_i is solution to

$$dD_i(\tau) = D_i(\tau)a_{D_i}(\tau)d\tau + D_i(\tau)\sigma_{D_i}(\tau)dW(\tau),$$

where $a_{D_i}(\tau)$ and $\sigma_{D_i}(\tau)$ are $\mathcal{F}(\tau)$-adapted, with $\sigma_{D_i} \in \mathbb{R}^d$.

A *strategy* is a predictable process in $\mathbb{R}^{m+1}$, denoted as:

$$\theta = \left\{ \theta(\tau) = (\theta_0(\tau), \cdots, \theta_m(\tau))^\top \right\}_{\tau\in[t,T]},$$

and satisfying $E\left[\int_t^T \|\theta(\tau)\|^2 d\tau\right] < \infty$. The (net of dividends) *value* of a strategy is:

$$V \equiv q_+ \cdot \theta,$$

where q_+ is a row vector. By generalizing the reasoning in section 4.3.1, we say that a strategy is self-financing if its value is the solution of:

$$dV = \left[\pi^\top (a - \mathbf{1}_m r) + Vr - c\right] dt + \pi^\top \sigma dW,$$

where $\mathbf{1}_m$ is m-dimensional vector of ones, $\pi \equiv (\pi_1, \cdots, \pi_m)^\top, \pi_i \equiv \theta_i q_i$, $i = 1, \cdots, m$, $a \equiv (\hat{a}_1 + \frac{D_1}{q_1}, \cdots, \hat{a}_m + \frac{D_m}{q_m})^\top$. This is exactly the multidimensional counterpart of the budget constraint of section ???. We require V to be strictly positive.

The solution of the previous equation is, for all $\tau \in [t, T]$,

$$\left(q_0^{-1}V^{x,\pi,c}\right)(\tau) = x - \int_t^\tau \left(q_0^{-1}c\right)(u)du + \int_t^\tau \left(q_0^{-1}\pi^\top\sigma\right)(u)dW(u) + \int_t^\tau \left(q_0^{-1}\pi^\top(a - \mathbf{1}_m r)\right)(u)du. \tag{4.29}$$

4.3.2 Viability of the model

Let $\bar{g}_i = \frac{q_i}{q_0} + \bar{z}_i$, $i = 1, \cdots, m$, where $d\bar{z}_i = \frac{1}{q_0} dz_i$ and $z_i(\tau) = \int_t^\tau D_i(u)du$. Let us generalize the definition of $\mathcal{P}$ in (4.2) by introducing the set $\mathcal{P}$ defined as:

$$\mathcal{P} \equiv \{Q \approx P : \bar{g}_i \text{ is a } Q\text{-martingale}\}.$$

Aim of this section is to show the equivalent of thm. 2.? in chapter 2: $\mathcal{P}$ is not empty if and only if there are not arbitrage opportunities.[2] This is possible thanks to the assumption of the previous subsection.

Let us give some further information concerning $\mathcal{P}$. We are going to use a well-known result stating that in correspondence with every $\mathcal{F}(t)$-adapted process $\{\lambda(t)\}_{t\in[0,T]}$ satisfying some basic regularity conditions (essentially, the Novikov's condition),

$$W^*(t) = W(t) + \int_t^\tau \lambda(u)du, \ \tau \in [t, T],$$

is a standard Brownian motion under a probability Q which is equivalent to P, with Radon-Nikodym derivative equal to,

$$\eta(T) \equiv \frac{dQ}{dP} = \exp\left(-\int_t^T \lambda^\top(\tau)dW(\tau) - \frac{1}{2}\int_t^T \|\lambda(\tau)\|^2 d\tau\right).$$

This result is known as the *Girsanov's theorem*. The process $\{\eta(\tau)\}_{\tau\in[t,T]}$ is a martingale under P.

Now let us rewrite eq. (4.3) under such a new probability by plugging in it W^*. We have that under Q,

$$dq_i(\tau) = q_i(\tau)\left(\hat{a}_i(\tau) - (\sigma_i\lambda)(\tau)\right)dt + q_i(\tau)\sigma_i(\tau)dW^*(\tau), \ i = 1, \cdots, m.$$

We also have

$$d\bar{g}_i(\tau) = d\left(\frac{q_i}{q_0}\right)(\tau) + d\bar{z}_i(\tau) = \frac{q_i(\tau)}{q_0(\tau)}\left[\left(a_i(\tau) - r(\tau)\right)d\tau + \sigma_i(\tau)dW(\tau)\right].$$

If $\bar{g}_i$ is a Q-martingale, i.e.

$$q_i(\tau) = E_\tau^Q\left[\frac{q_0(\tau)}{q_0(T)}q_i(T) + \int_\tau^T ds \cdot \frac{q_0(\tau)}{q_0(s)}D_i(s)\middle|\mathcal{F}(\tau)\right], \ i = 1, \cdots, m,$$

it is necessary and sufficient that $a_i - \sigma_i\lambda = r$, $i = 1, \cdots, m$, i.e. in matrix form,

$$a(\tau) - \mathbf{1}_m r(\tau) = (\sigma\lambda)(\tau) \text{ a.s.} \tag{4.30}$$

Therefore, in this case the solution for wealth is, by (4.5),

$$\left(q_0^{-1}V^{x,\pi,c}\right)(\tau) = x - \int_t^\tau \left(q_0^{-1}c\right)(u)du + \int_t^\tau \left(q_0^{-1}\pi^\top\sigma\right)(u)dW^*(u), \ \tau \in [t, T], \tag{4.31}$$

[2] In fact, thm. 2.? is formulated in terms of state prices, but section 2.? showed that there is a one-to-one relationship between state prices and "risk-neutral" probabilities. Here, it is more convenient on an analytical standpoint to face the problem in terms of probabilities $\in \mathcal{P}$.

where $x \equiv \mathbf{1}^\top q(t)$ is initial wealth, i.e. the value of the trees at the beginning of the economy.

Definition 4.1 (Arbitrage opportunity). *A portfolio π is an arbitrage opportunity if $V^{x,\pi,0}(t) \leq \left(q_0^{-1}V^{x,\pi,0}\right)(T)$ a.s. and $P\left(\left(q_0^{-1}V^{x,\pi,0}\right)(T) - V^{x,\pi,0}(t) > 0\right) > 0$.*

We are going to show the equivalent of thm. 2.? in chapter 2.

Theorem 4.2. *There are no arbitrage opportunities if and only if $\mathcal{P}$ is not empty.*

As we will see during the course of the proof provided below, the if part follows easily from elaborating (4.7). The only if part is more elaborated, but its basic structure can be understood as follows. By the Girsanov's theorem, the statement AOA $\Rightarrow \exists Q \in \mathcal{Q}$ is equivalent to AOA $\Rightarrow \exists\lambda$ satisfying (4.6). Now if (4.6) doesn't hold, one can implement an AO. Indeed, it would be possible to find a nonzero $\underline{\pi} : \underline{\pi}^\top\sigma = 0$ and $\underline{\pi}^\top(a - \mathbf{1}_m r) \neq 0$; then one can use a portfolio $\underline{\pi}$ when $a - \mathbf{1}_m r > 0$ and $-\underline{\pi}$ when $a - \mathbf{1}_m r < 0$ and obtain an appreciation rate of V greater than r in spite of having zeroed uncertainty with $\underline{\pi}^\top\sigma = 0$! But if (4.6) holds, one can never find such an AO. This is simply so because if (4.6) holds, it also holds that $\forall\pi$, $\pi^\top(a - \mathbf{1}_m r) = \pi^\top\sigma\lambda$ and we are done. Of course models such as BS trivially fulfill such a requirement, because a, r, and σ are just numbers in the geometric Brownian motion case, and automatically identify λ by $\lambda = \frac{a-r}{\sigma}$. More generally, relation (4.6) automatically holds when $m = d$ and σ has full-rank.

Let

$$\langle\sigma^\top\rangle^\perp \equiv \left\{x \in L^2_{t,T,m} : \sigma^\top x = \mathbf{0}_d \quad \text{a.s.}\right\}$$

and

$$\langle\sigma\rangle \equiv \left\{z \in L^2_{t,T,m} : z = \sigma u, \quad \text{a.s.}, \quad \text{for } u \in L^2_{t,T,d}\right\}.$$

Then the previous reasoning can be mathematically formalized by saying that $a - \mathbf{1}_m r$ must be orthogonal to all vectors in $\langle\sigma^\top\rangle^\perp$ and since $\langle\sigma\rangle$ and $\langle\sigma^\top\rangle^\perp$ are orthogonal, $a - \mathbf{1}_m r \in \langle\sigma\rangle$, or $\exists\lambda \in L^2_{t,T,d} : a - \mathbf{1}_m r = \sigma\lambda$.[3]

Proof of Theorem 4.2. As pointed out in the previous discussion, all statements on nonemptiness of $\mathcal{P}$ are equivalent to statements on relation (4.6) being true on the basis of the Girsanov's theorem. Therefore, we will be working with relation (4.6).

If part. When $c \equiv 0$, relation (4.7) becomes:

$$\left(q_0^{-1}V^{x,\pi,0}\right)(\tau) = x + \int_t^\tau \left(q_0^{-1}\pi^\top\sigma\right)(u)dW^*(u), \ \tau \in [t, T],$$

from which we immediately get that:

$$x = E^Q_\tau\left[\left(q_0^{-1}V^{x,\pi,0}\right)(T)\right].$$

[3] To see that $\langle\sigma\rangle$ and $\langle\sigma'\rangle^\perp$ are orthogonal spaces, note that:

$$\begin{aligned}
\left\{x \in L^2_{t,T,m} : x'z = 0, \quad z \in \langle\sigma\rangle\right\} &= \left\{x \in L^2_{t,T,m} : x'\sigma u = 0, \quad u \in L^2_{t,T,d}\right\} \\
&= \left\{x \in L^2_{t,T,m} : x'\sigma = \mathbf{0}_d\right\} \\
&= \left\{x \in L^2_{t,T,m} : \sigma' x = \mathbf{0}_d\right\} \\
&\equiv \langle\sigma'\rangle^\perp.
\end{aligned}$$

Now an AO is $V^{x,\pi,0}(t) \leq \left(q_0^{-1}V^{x,\pi,0}\right)(T)$ a.s. which, combined with the previous equality leaves: $V^{x,\pi,0}(t) = \left(q_0^{-1}V^{x,\pi,0}\right)(T)$ Q-a.s. and so P-a.s. (if a r.v. $\tilde{y} \geq 0$ and $E_t(\tilde{y}) = 0$, this means that $\tilde{y} = 0$ a.s.), and this contradicts $P\left(\left(q_0^{-1}V^{x,\pi,0}\right)(T) - V^{x,\pi,0}(t) > 0\right) > 0$, as required in definition 4.3.

Only if part. We follow Karatzas (1997, thm. 0.2.4 pp. 6-7) and Øksendal (1998, thm. 12.1.8b, pp. 256-257) and let:

$$\begin{aligned} Z(\tau) &= \{\omega : \text{ eq. (4.6) has no solutions}\} \\ &= \{\omega : a(\tau;\omega) - \mathbf{1}_m r(\tau;\omega) \notin \langle\sigma\rangle\} \\ &= \left\{\omega : \exists \underline{\pi}(\tau;\omega) : \underline{\pi}(\tau;\omega)^\top \sigma(\tau;\omega) = 0 \text{ and } \underline{\pi}(\tau;\omega)^\top \left(a(\tau;\omega) - \mathbf{1}_m r(\tau;\omega)\right) \neq 0\right\}, \end{aligned}$$

and consider the following portfolio,

$$\hat{\pi}(\tau;w) = \begin{cases} k \cdot sign\left[\underline{\pi}(\tau;\omega)^\top \left(a(\tau;\omega) - \mathbf{1}_m r(\tau;\omega)\right)\right] \cdot \underline{\pi}(\tau;\omega) & \text{for } \omega \in Z(\tau) \\ 0 & \text{for } \omega \notin Z(\tau) \end{cases}$$

Clearly $\hat{\pi}$ is $(\tau;\omega)$-measurable, and it generates, by virtue of Eq. (4.5),

$$\begin{aligned} &\left(q_0^{-1}V^{x,\hat{\pi},0}\right)(\tau;\omega) \\ =\ & x + \int_t^\tau \left(q_0^{-1}\hat{\pi}^\top\sigma\right)(u;\omega)\mathbb{I}_{Z(u)}(\omega)dW(u) + \int_t^\tau \left(q_0^{-1}\hat{\pi}^\top(a - \mathbf{1}_m r)\right)(u;\omega)\mathtt{1}_{Z(u)}(\omega)du \\ =\ & x + \int_t^\tau \left(q_0^{-1}\hat{\pi}^\top(a - \mathbf{1}_m r)\right)(u;\omega)\mathtt{1}_{Z(u)}(\omega)du \\ \geq\ & x, \quad \text{for } \tau \in (t,T]. \end{aligned}$$

But the market has no arbitrage, which can be possible only if:

$$\mathbb{I}_{Z(u)}(\omega) = 0, \quad \text{for a.a. } (\tau;\omega),$$

id est only if eq. (4.6) has at least one solution for a.a. $(\tau;\omega)$.

‖

4.3.3 Completeness conditions

Let $Y \in L^2(\Omega, \mathcal{F}, P)$. We have the following definition.

Definition 4.3 (Completeness). Markets are *complete* if there is a portfolio process π : $V^{x,\pi,0}(T) = Y$ a.s. in correspondence with any $\mathcal{F}(T)$-measurable random variable $Y \in L^2(\Omega, \mathcal{F}, P)$.

The previous definition represents the natural continuous time counterpart of the definition given in the discrete time case. In continuous time, there is a theorem reproducing some of the results known in the discrete time case: markets are complete if and only if 1) $m = d$ and 2) the price volatility matrix of the available assets (primitives and derivatives) is a.e. nonsingular. Here we are going to show the sufficiency part of this theorem (see, e.g., Karatzas (1997 pp. 8-9) for the converse); in coming up with the arguments of the proof, the reader will realize that such a result is due to the possibility to implement fully spanning dynamic strategies, as in the discrete time case.

As regards the proof, let us suppose that $m = d$ and that $\sigma(t,\omega)$ is nonsingular a.s. Let us consider the Q-martingale:

$$M(\tau) \equiv E^Q \left[q_0(T)^{-1} \cdot Y \middle| \mathcal{F}(\tau) \right], \quad \tau \in [t,T]. \tag{4.32}$$

By virtue of the representation theorem of continuous local martingales as stochastic integrals with respect to Brownian motions (e.g., Karatzas and Shreve (1991) (thm. 4.2 p. 170)), there exists $\varphi \in L^2_{0,T,d}(\Omega, \mathcal{F}, Q)$ such that M can be written as:

$$M(\tau) = M(t) + \int_t^\tau \varphi^\top(u) dW^*(u), \quad \text{a.s. in } \tau \in [t,T].$$

We wish to find out portfolio processes π which generate a discounted wealth process without consumption $\left\{ (q_0^{-1} V^{x,\pi,0})(\tau) \right\}_{\tau \in [t,T]}$ (defined under a probability $\in \mathcal{P}$) such that $q_0^{-1} V^{x,\pi,0} = M$, P (or Q) a.s. By Eq. (4.7),

$$\left(q_0^{-1} V^{x,\pi,0} \right)(\tau) = x + \int_t^\tau \left(q_0^{-1} \pi^\top \sigma \right)(u) dW^*(u), \quad \tau \in [t,T],$$

and by identifying we pick a portfolio $\hat{\pi}^\top = q_0 \varphi^\top \sigma^{-1}$, and $x = M(t)$. In this case, $M(\tau) = \left(q_0^{-1} \cdot V^{M(t),\hat{\pi},0} \right)(\tau)$ a.s., and in particular,

$$M(T) = \left(q_0^{-1} \cdot V^{M(t),\hat{\pi},0} \right)(T) \text{ a.s.}$$

By comparing with Eq. (4.8),

$$V^{M(t),\hat{\pi},0}(T) = Y \quad \text{a.s.} \tag{4.33}$$

This is completeness.

Theorem 4.4. *$\mathcal{P}$ is a singleton if and only if markets are complete.*

Proof. There exists a unique $\lambda : a - \mathbf{1}_m r = \sigma \lambda \Leftrightarrow m = d$. The result follows by the Girsanov's theorem. ‖

When markets are incomplete, there is an infinity of probabilities $\in \mathcal{P}$, and the information "absence of arbitrage opportunities" is not sufficient to "recover" the *true* probability. One could make use of general equilibrium arguments, but in this case we go beyond the edge of knowledge. To see what happens in the Brownian information case (which is of course a very specific case, but it also allows one to deduce very simple conclusions that very often have counterparts in more general settings), let $\mathcal{P}$ be the set of martingale measures that are equivalentes to P on $(\Omega, \mathcal{F})$ for $\bar{g}_i$, $i = 1, \cdots, m$. As shown in the previous subsection, the asset price model is viable if and only if $\mathcal{P}$ is not empty. Let $L^2_{0,T,d}(\Omega, \mathcal{F}, P)$ be the space of all $\mathcal{F}(t)$-adapted processes $x = \{x(t)\}_{t \in [0,T]}$ in $\mathbb{R}^d$ satisfying:

$$0 < \int_0^T \|x(u)\|^2 du < \infty \quad P\text{-}p.s.$$

Define,

$$\langle \sigma \rangle^\perp \equiv \left\{ x \in L^2_{0,T,d}(\Omega, \mathcal{F}, P) : \sigma(t) x(t) = \mathbf{0}_m \quad \text{a.s.} \right\},$$

where $\mathbf{0}_m$ is a vector of zeros in $\mathbb{R}^m$. Let

$$\hat{\lambda}(t) = \left(\sigma^\top(\sigma\sigma^\top)^{-1}(a - \mathbf{1}_m r)\right)(t).$$

Under the usual regularity conditions (essentially: the Novikov's condition), $\widehat{\lambda}$ can be interpreted as the process of unit risk-premia. In fact, *all* processes belonging to the set:

$$\mathcal{Z} = \left\{\lambda : \lambda(t) = \hat{\lambda}(t) + \eta(t),\ \eta \in \langle\sigma\rangle^\perp\right\}$$

are bounded and can thus be interpreted as unit risk-premia processes. More precisely, define the Radon-Nikodym derivative of Q with respect to P on $\mathcal{F}(T)$:

$$\hat{\eta}(T) \equiv \frac{dQ}{dP} = \exp\left(-\int_0^T \hat{\lambda}^\top(t)dW(t) - \frac{1}{2}\int_0^T \left\|\hat{\lambda}(t)\right\|^2 dt\right),$$

and the density process of all $Q \approx P$ on $(\Omega, \mathcal{F})$,

$$\eta(t) = \hat{\eta}(t)\cdot\exp\left(-\int_0^t \eta^\top(u)dW(u) - \frac{1}{2}\int_0^t \|\eta(u)\|^2 du\right),\ t \in [0,T]),$$

a strictly positive P-martingale, we have:

PROPOSITION 4.5. *$Q \in \mathcal{P}$ if and only if it is of the form*: $Q(A) = E\{1_A \eta(T)\}\ \forall A \in \mathcal{F}(T)$.

PROOF. Standard. Adapt, for instance, Proposition 1 p. 271 of He and Pearson (1991) or Lemma 3.4 p. 429 of Shreve (1991) to the primitive asset price process in eq. (4.26)-(4.27). ‖

We have $\dim(\langle\sigma\rangle^\perp) = d - m$, and the previous lemma immediately reveals that markets incompleteness implies the existence of an infinity of probabilities $\in \mathcal{P}$. Such a result was shown in great generality by Harrison and Pliska (1983).[4]

4.4 Equilibrium with a representative agent

4.4.1 The program

An agent maximises expected utility flows $(u(c))$ and an utility function of terminal wealth $(U(v))$ under the constraint of eq. (4.5); here $v \in L^2(\Omega, \mathcal{F}, P)$.[5] The agent faces the following program,

$$J(0, V_0) = \max_{(\pi,c,v)} E\left[U(v) + \int_t^T u(c(\tau))d\tau\right],$$

$$\text{s.t. } q_0(T)^{-1}\cdot v = \left(q_0^{-1}V^{x,\pi,c}\right)(T)$$
$$= x - \int_t^T \left(q_0^{-1}c\right)(u)du + \int_t^T \left(q_0^{-1}\pi^\top\sigma\right)(u)dW(u) + \int_t^T \left(q_0^{-1}\pi^\top(a - \mathbf{1}_m r)\right)(u)du.$$

[4]The so-called Föllmer and Schweizer (1991) measure, or minimal equivalent martingale measure, is defined as: $\widehat{P}^*(A) \equiv E\left\{1_A\widehat{\xi}(T)\right\}\ \forall A \in \mathcal{F}(T)$.

[5]Moreover, we assume that the agent only considers the choice space in which the control functions satisfy the elementary Markov property and belong to $L^2_{0,T,m}(\Omega, \mathcal{F}, P)$ and $L^2_{0,T,1}(\Omega, \mathcal{F}, P)$.

The standard (first) approach to this problem was introduced by Merton in finance. There are also other approaches in which reference is made to Arrow-Debreu state prices - just as we've made in previous chapters. We now describe the latter approach. The first thing to do is to derive a representation of the budget constraint paralleling the representation,

$$0 = c^0 - w^0 + E\left[m \cdot \left(c^1 - w^1\right)\right]$$

obtained for the two-period model and discount factor m (see chapter 2). The logic here is essentially the same. In chapter, we multiplied the budget constraint by the Arrow-Debreu state prices,

$$\phi_s = m_s \cdot P_s, \quad m_s \equiv (1+r)^{-1}\eta_s, \quad \eta_s = \frac{Q_s}{P_s},$$

and "took the sum over all the states of nature". Here we wish to obtain a similar representation in which we make use of Arrow-Debreu state price *densities* of the form:

$$\phi_{t,T}(\omega) \equiv m_{t,T}(\omega) \cdot dP(\omega), \quad m_{t,T}(\omega) = q_0(\omega,T)^{-1}\eta(\omega,T), \quad \eta(\omega,T) = \frac{dQ}{dP}(\omega).$$

Similarly to the finite state space, we multiply the constraint by the previous process and then "take the integral over all states of nature". In doing so, we turn our original problem from one with an infinity of trajectory constraints (wealth) to a one with only one (average) constraint.

1st step Rewrite the constraint as,

$$0 = V^{x,\pi,c}(T) + \int_t^T \frac{q_0(T)}{q_0(u)} c(u)du - q_0(T)x - \int_t^T \frac{q_0(T)}{q_0(u)} \left[\left(\pi^\top (a - \mathbf{1}_m r)\right)(u)du + (\pi^\top \sigma)(u)dW(u)\right].$$

This constraint must hold for every point $\omega \in \Omega$.

2nd step For a fixed ω, multiply the previous constraint by $\phi_{0,T}(\omega) = q_0(\omega,T)^{-1} \cdot dQ(\omega)$,

$$\begin{aligned} 0 &= \left[q_0(T)^{-1}V^{x,\pi,c}(T) + \int_t^T \frac{1}{q_0(u)} c(u)du - x\right] dQ \\ &\quad - \left[\int_t^T q_0(u)^{-1}\left[\left(\pi^\top (a - \mathbf{1}_m r)\right)(u)du + (\pi^\top \sigma)(u)dW(u)\right]\right] dQ. \end{aligned}$$

3rd step Take the integral over all states of nature. By the Girsanov's theorem,

$$0 = \mathbb{E}\left[q_0(T)^{-1}V^{x,\pi,c}(T) + \int_t^T q_0(u)^{-1}c(u)du - x\right].$$

$\|$

Next, let's retrieve back the true probability P. We have,

$$\begin{aligned}
x &= \mathbb{E}\left[q_0(T)^{-1}V^{x,\pi,c}(T)+\int_t^T q_0(u)^{-1}c(u)du\right] \\
&= \eta(t)^{-1}\cdot E\left[\left(q_0^{-1}\eta V^{x,\pi,c}\right)(T)+\int_t^T q_0(u)^{-1}\eta(T)c(u)du\right] \\
&= \eta(t)^{-1}\cdot E\left[\left(q_0^{-1}\eta V^{x,\pi,c}\right)(T)+\int_t^T E\left(q_0(u)^{-1}\eta(T)c(u)\right|\mathcal{F}(u)\right)du\right] \\
&= \eta(t)^{-1}\cdot E\left[\left(q_0^{-1}\eta V^{x,\pi,c}\right)(T)+\int_t^T q_0(u)^{-1}c(u)E\left(\eta(T)\right|\mathcal{F}(u)\right)du\right] \\
&= \eta(t)^{-1}\cdot E\left[\left(q_0^{-1}\eta V^{x,\pi,c}\right)(T)+\int_t^T \left(q_0^{-1}\eta c\right)(u)du\right] \\
&= E\left[m_{t,T}\cdot V^{x,\pi,c}(T)+\int_t^T m_{t,u}\cdot c(u)du\right],
\end{aligned}$$

where we used the fact that c is adapted, the law of iterated expectations, the martingale property of η, and the definition of $m_{0,t}$.

The program is,

$$\begin{aligned}
J(t,x) &= \max_{(c,v)} E\left[e^{-\rho(T-t)}U(v)+\int_t^T u(\tau,c(\tau))d\tau\right], \\
\text{s.t. } x &= E\left[m_{t,T}\cdot v+\int_t^T m_{t,\tau}\cdot c(\tau)d\tau\right]
\end{aligned}$$

That is,

$$\max_{(c,v)} E\left[\int_t^T \left[u\left(\tau,c(\tau)\right)-\psi\cdot m_{t,\tau}\cdot c(\tau)\right]d\tau+U(v)-\psi\cdot m_{t,T}\cdot v+\psi\cdot x\right],$$

where ψ is the constraint's multiplier.

The first order conditions are:

$$\begin{cases} u_c\left(\tau,c(\tau)\right)=\psi\cdot m_{t,\tau}, & \tau\in[t,T) \\ U'(v)=\psi\cdot m_{t,T} \end{cases} \tag{4.34}$$

They imply that, for any two instants $t_1\leq t_2\in[0,T]$,

$$\frac{u_c\left(t_1,c(t_1)\right)}{u_c\left(t_2,c(t_2)\right)}=\frac{m_{t,t_1}}{m_{t,t_2}}. \tag{4.35}$$

Naturally, such a methodology is only valid when $m=d$. Indeed, in this case there is one and only one Arrow-Debreu density process. Furthermore, there are always portfolio strategies π ensuring any desired consumption plan $\left(\hat{c}(\tau)\right)_{\tau\in[t,T]}$ and $\hat{v}\in L^2(\Omega,\mathcal{F},P)$ when markets are complete. The proof of such assertion coincides with a constructive way to compute such portfolio processes. For $\left(\hat{c}(\tau)\right)_{\tau\in[t,T]}\equiv 0$, the proof is just the proof of (4.9). In the general case, it is sufficient to modify slightly the arguments. Define,

$$M(\tau)\equiv E^Q\left[q_0^{-1}(T)\cdot\hat{v}+\int_t^T q_0(u)^{-1}\hat{c}(u)du\right|\mathcal{F}(\tau)\right],\quad \tau\in[t,T].$$

Notice that:

$$M(\tau)=E^{Q}\left[q_0^{-1}(T)\cdot\hat{v}+\int_t^T q_0(u)^{-1}\hat{c}(u)du\right|\mathcal{F}(\tau)\right]=E\left[m_{t,T}\cdot\hat{v}+\int_t^T m_{t,u}\cdot\hat{c}(u)du\right|\mathcal{F}(\tau)\right].$$

By the predictable representation theorem, $\exists\phi$ such that:

$$M(\tau)=M(t)+\int_t^\tau \phi^\top(u)dW(u).$$

Consider the process $\{m_{0,t}V^{x,\pi,c}(\tau)\}_{\tau\in[t,T]}$. By Itô's lemma,

$$m_{0,t}V^{x,\pi,c}(\tau)+\int_t^\tau m_{t,u}\cdot c(u)du=x+\int_t^\tau m_{t,u}\cdot\left(\pi^\top\sigma-V^{x,\pi,c}\lambda\right)(u)dW(u).$$

By identifying,

$$\pi^\top(\tau)=\left[\left(V^{x,\pi,c}\lambda\right)(\tau)+\frac{\phi^\top(\tau)}{m_{t,\tau}}\right]\sigma(\tau)^{-1}, \tag{4.36}$$

where $V^{x,\pi,c}(\tau)$ can be computed from the constraint of the program (4.10) written in $\tau\in[t,T]$:

$$V^{x,\pi,c}(\tau)=E\left[m_{\tau,T}\cdot v+\int_\tau^T m_{\tau,u}\cdot c(u)du\right|\mathcal{F}(\tau)\right], \tag{4.37}$$

once that the optimal trajectory of c has been computed. As in the two-periods model of chapter 2, the kernel process $\{m_{\tau,t}\}_{\tau\in[t,T]}$ is to be determined as an equilibrium outcome. Furthermore, in the continuous time model of this chapter there also exist strong results on the existence of a representative agent (Huang (1987)).

Example 4.6. $U(v)=\log v$ and $u(x)=\log x$. By exploiting the first order conditions (4.11) one has $\frac{1}{\hat{c}(\tau)}=\psi\cdot m_{t,\tau}$, $\frac{1}{\hat{v}}=\psi\cdot m_{\tau,T}$. By plugging these conditions into the constraint of the program (4.10) one obtains the solution for the Lagrange multiplier: $\psi=\frac{T+1}{x}$. By replacing this back into the previous first order conditions, one eventually obtains: $\hat{c}(t)=\frac{x}{T+1}\frac{1}{m_{t,\tau}}$, and $\hat{v}=\frac{x}{T+1}\frac{1}{m_{t,T}}$. As regards the portfolio process, one has that:

$$M(\tau)=E\left[m_{\tau,T}\cdot\hat{v}+\int_t^T m_{t,u}\hat{c}(u)du\right|\mathcal{F}(\tau)\right]=x,$$

which shows that $\phi=0$ in the representation (4.13). By replacing $\phi=0$ into (4.13) one obtains:

$$\pi^\top(\tau)=\left(V^{x,\pi,\hat{c}}\lambda\right)(\tau)\cdot\sigma(\tau)^{-1}.$$

We can compute $V^{x,\pi,\hat{c}}$ in (4.14) by using $\hat{c}$:

$$V^{x,\pi,\hat{c}}(\tau)=\frac{x}{T+1}E\left[\frac{m_{\tau,T}}{m_{t,T}}+\int_\tau^T\frac{m_{\tau,u}}{m_{t,u}}du\right|\mathcal{F}(\tau)\right]=\frac{x}{m_{t,\tau}}\frac{T+1-(\tau-t)}{T+1},$$

where we used the property that m forms an *evolving semi-group*: $m_{t,a}\cdot m_{t,b}=m_{t,b}$, $t\leq a\leq b$. The solution is:

$$\pi^\top(\tau)=\frac{x}{m_{t,\tau}}\frac{T+1-t}{T+1}\left(\lambda\sigma^{-1}\right)(\tau)=\frac{x}{m_{t,\tau}}\frac{T+1-(\tau-t)}{T+1}\left(\lambda\sigma^{-1}\right)(\tau)$$

whence, by taking into account the relation: $a-\mathbf{1}_m r=\sigma\lambda$,

$$\pi(\tau)=\frac{x}{m_{t,\tau}}\frac{T+1-(\tau-t)}{T+1}\left[(\sigma\sigma^\top)^{-1}(a-\mathbf{1}_m r)\right](\tau).$$

4.4.2 The older, Merton's approach: dynamic programming

Here we show how to derive optimal consumption and portfolio policies using the Bellman's approach in an infinite horizon setting. Suppose to have to solve the problem

$$\begin{aligned} J(V(t)) &= \max_c E\left[\int_t^\infty e^{-\rho(\tau-t)} u(c(\tau))\, d\tau\right] \\ \text{s.t. } dV &= \left[\pi^\top (a - \mathbf{1}_m) + rV - c\right] d\tau + \pi^\top \sigma dW \end{aligned}$$

via the standard Bellman's approach, viz:

$$0 = \max_c E\left[u(c) + J'(V)\left(\pi^\top (a - \mathbf{1}_m r) + rV - c\right) + \frac{1}{2} J''(V) \pi^\top \sigma\sigma^\top \pi - \rho J(V)\right]. \tag{4.38}$$

First order conditions are:

$$\begin{aligned} u'(c) &= J'(V) \\ \pi &= \left(\frac{-J'(V)}{J''(V)}\right) \left(\sigma\sigma^\top\right)^{-1} (a - \mathbf{1}_m r) \end{aligned} \tag{4.39}$$

By plugging these back into the Bellman's equation (4.37) leaves:

$$0 = u(c) + J'(V)\left[\frac{-J'(V)}{J''(V)} \cdot Sh + rV - c\right] + \frac{1}{2} J''(V) \left[\frac{-J'(V)}{J''(V)}\right]^2 Sh - \rho J(V), \tag{4.40}$$

where:

$$Sh \equiv (a - \mathbf{1}_m r)^\top (\sigma\sigma^\top)^{-1} (a - \mathbf{1}_m r),$$

with $\lim_{T\to\infty} e^{-\rho(T-t)} E\left[J(V(T))\right] = 0$.

Let's take

$$u(x) = \frac{x^{1-\eta} - 1}{1-\eta},$$

and conjecture that:

$$J(x) = A\frac{x^{1-\eta} - B}{1-\eta},$$

where A, B are constants to be determined. Using the first line of (4.38), $c = A^{-1/\eta} V$. By plugging this and using the conjectured analytical form of J, eq. (4.39) becomes:

$$0 = AV^{1-\eta}\left(\frac{\eta}{1-\eta} A^{-1/\eta} + \frac{1}{2}\frac{Sh}{\eta} + r - \frac{\rho}{1-\eta}\right) - \frac{1}{1-\eta}(1 - \rho AB).$$

This must hold for every V. Hence

$$A = \left(\frac{\rho - r(1-\eta)}{\eta} - \frac{(1-\eta)Sh}{2\eta^2}\right)^{-\eta}; \qquad B = \frac{1}{\rho}\left(\frac{\rho - r(1-\eta)}{\eta} - \frac{(1-\eta)Sh}{2\eta^2}\right)^{\eta}$$

Clearly, $\lim_{\eta\to 1} J(V) = \rho^{-1} \log V$.

4.4.3 Equilibrium and Walras's consistency tests

An equilibrium is a consumption plan satisfying the first order conditions (4.11) and a portfolio process having the form (4.13), with

$$c(\tau) = D(\tau) \equiv \sum_{i=1}^{m} D_i(\tau), \text{ for } \tau \in [t, T); \quad \text{and} \quad v = \underline{q}(T) \equiv \sum_{i=1}^{m} q_i(T)$$

and

$$\theta_0(\tau) = 0, \;\; \pi(\tau) = q(\tau), \text{ for } \tau \in [t, T].$$

For reasons developed below, it is useful to derive the dynamics of the dividend, D. This is the solution of:

$$dD(\tau) = a_D(\tau)D(\tau)d\tau + \sigma_D(\tau)D(\tau)dW(\tau),$$

where $a_D D \equiv \sum_{i=1}^{m} a_{D_i} D_i$ and $\sigma_D D \equiv \sum_{i=1}^{m} \sigma_{D_i} D_i$.

Equilibrium allocations and Arrow-Debreu state price (*densities*) in the present infinite dimensional commodity space can now be computed as follows.

The first order conditions say that:

$$d \log u_c(\tau, c(\tau)) = d \log m_{t,\tau}, \tag{4.41}$$

and since $\frac{\partial}{\partial u} E(\frac{m_{\tau,u}}{m_{\tau,u}} - 1)\Big|_{u=\tau} = -r$, we have that the rate of decline of marginal utility is equal to the short-term rate. Such a relationship allows one to derive the term-structure of interest rates.

Now by replacing the differential of $\log m_{t,\tau}$ into relation (4.17),

$$d \log u_c(\tau, c(\tau)) = -\left(r(\tau) + \frac{1}{2} \|\lambda(\tau)\|^2 \right) d\tau - \lambda^\top(\tau) dW(\tau).$$

But in equilibrium,

$$d \log u_c(\tau, D(\tau)) = -\left(r(\tau) + \frac{1}{2} \|\lambda(\tau)\|^2 \right) dt - \lambda^\top(\tau) dW(\tau).$$

By Itô's lemma, $\log u_c(\tau, D(\tau))$ is solution of

$$d \log u_c = \left[\frac{u_{\tau c}}{u_c} + a_D D \frac{u_{cc}}{u_c} + \frac{1}{2} \sigma_D^2 D^2 \left(\frac{u_{ccc}}{u_c} - \left(\frac{u_{cc}}{u_c} \right)^2 \right) \right] dt + \frac{u_{cc}}{u_c} D \sigma_D dW.$$

By identifying,

$$\begin{aligned} r &= -\left[\frac{u_{\tau c}}{u_c} + a_D D \frac{u_{cc}}{u_c} + \frac{1}{2} \sigma_D^2 D^2 \left(\frac{u_{ccc}}{u_c} - \left(\frac{u_{cc}}{u_c} \right)^2 \right) + \frac{1}{2} \|\lambda\|^2 \right] \\ \lambda^\intercal &= -\frac{u_{cc}}{u_c} \cdot D \cdot \sigma_D \end{aligned}$$

The first relation is the drift identification condition for $\log u_c$ (Arrow-Debreu state-price density *drift* identification). The second relation is the diffusion identification condition for $\log u_c$ (Arrow-Debreu state-price density *diffusion* identification).

By replacing the first relation of the previous identifying conditions into the second one we get the equilibrium short-term rate:

$$r(\tau)=-\left[\frac{u_{\tau c}}{u_c}(\tau,D(\tau))+a_D(\tau)\cdot D(\tau)\cdot\frac{u_{cc}}{u_c}(\tau,D(\tau))+\frac{1}{2}\sigma_D(\tau)^2D(\tau)^2\cdot\frac{u_{ccc}}{u_c}(\tau,D(\tau))\right]. \tag{4.42}$$

As an example, if $u(\tau,x)=e^{-(\tau-t)\rho}\frac{x^{1-\eta}}{1-\eta}$ and $m=1$, then:

$$r(\tau)=\rho+\eta\cdot a_D(\tau)-\frac{1}{2}\eta(\eta+1)\sigma_D(\tau)^2. \tag{4.43}$$

Furthermore, λ is well identified, which allows one to compute the Arrow-Debreu state price density (whence allocations by the formulae seen in the previous subsection once the model is extended to the case of heterogeneous agents).

We are left to check the following consistency aspect of the model: in the approach of this section, we neglected all possible portfolio choices of the agent. We have to show that (4.15) $\Rightarrow$ (4.16). In fact, we are going to show that in the absence of arbitrage opportunities, (4.15) $\Leftrightarrow$ (4.16). This is shown in appendix 3.

4.4.4 Continuous-time CAPM

We showed that $\mathcal{P}$ is a singleton in this model. We can then write down without any ambiguity an evaluation formula for the primitive asset prices:

$$\begin{aligned}
q_i(\tau) &= E^Q\left[\frac{q_0(\tau)}{q_0(T)}q_i(T)+\int_\tau^T\frac{q_0(\tau)}{q_0(s)}D_i(s)ds\right|\mathcal{F}(\tau)\bigg] \\
&= \eta(\tau)^{-1}E\left[\eta(T)\frac{q_0(\tau)}{q_0(T)}q_i(T)+\int_\tau^T\frac{q_0(\tau)}{q_0(s)}D_i(s)\eta(T)ds\right|\mathcal{F}(\tau)\bigg] \\
&= E\left[\eta(\tau)^{-1}\eta(T)\frac{q_0(\tau)}{q_0(T)}q_i(T)+\int_\tau^T\eta(\tau)^{-1}\eta(s)\frac{q_0(\tau)}{q_0(s)}D_i(s)ds\right|\mathcal{F}(\tau)\bigg] \\
&= E\left[\frac{m_{t,T}}{m_{t,\tau}}q_i(T)+\int_\tau^T\frac{m_{t,s}}{m_{t,\tau}}D_i(s)ds\right|\mathcal{F}(\tau)\bigg].
\end{aligned}$$

By using (4.12) we obtain: $q_i(\tau)=E\left[\frac{u'(c(T))}{u'(c(\tau))}q_i(T)+\int_\tau^T ds\frac{u'(c(s))}{u'(c(\tau))}D_i(s)ds\right|\mathcal{F}(\tau)\bigg]$. At the equilibrium:

$$q_i(\tau)=E\left[\frac{u'\left(\underline{q}(T)\right)}{u'\left(D(\tau)\right)}q_i(T)+\int_\tau^T\frac{u'\left(D(s)\right)}{u'\left(D(\tau)\right)}D_i(s)ds\right|\mathcal{F}(\tau)\bigg].$$

4.4.5 Examples

Evaluation of a pure discount bond. Here one has $b(T)=1$ and $D_{bond}(s)=0$, $s\in[t,T]$. Therefore,

$$
\begin{aligned}
b(\tau) &= E\left[\left.\frac{u'\left(\underline{q}(T)\right)}{u'\left(D(\tau)\right)}\right|\mathcal{F}(\tau)\right] \\
&= E\left[\left.\frac{m_{t,T}}{m_{t,\tau}}\right|\mathcal{F}(\tau)\right] \\
&= E\left[\exp\left(-\int_\tau^T r(u)du - \int_\tau^T \lambda^\top(u)dW(u) - \frac{1}{2}\int_\tau^T \|\lambda(u)\|^2 du\right)\middle|\mathcal{F}(\tau)\right],
\end{aligned}
$$

where λ is as in relation (4.6). As an example, if $u(x(\tau)) = e^{-\rho(\tau-t)}\log x(\tau)$, then

$$\lambda(\tau) = \sigma_D(\tau).$$

4.5 Black & Scholes formula and "invisible" parameters

- The Black & Scholes and Merton model for option evaluation.
  - Hedging

In the Black & Scholes model,

$$-\frac{u''(c)}{u'(c)}c = \frac{\lambda}{\sigma},$$

where λ and σ are constants. Apart from some irrelevant constants, we have then:

$$u(c) = \frac{c^{1-\eta}-1}{1-\eta},\ \eta \equiv \frac{a-r}{\sigma^2} = \frac{\lambda}{\sigma}.$$

The Black & Scholes model is "supported" by an equilibrium model with a representative agent with CRRA preferences (who is risk neutral if and only if $a = r$), i.e.: there exist CRRA utility functions that support any risk-premium level that one has in mind. Such results are known since Rubenstein (1976) (cf. also Bick (1987) for further results). However, the methods used for the proof here are based on the more modern martingale approach.

4.6 Jumps

Brownian motions are well-suited to model price behavior of liquid assets, but there is a fair amount of interest in modeling discontinuous changes of asset prices, expecially as regards the modeling of fixed income instruments, where discontinuities can be generated by sudden changes in liquidity market conditions for instance. In this section, we describe one class of stochastic processes which is very popular in addressing the previous characteristics of asset prices movements, known as Poisson process.

4.6.1 Construction

Let (t, T) be a given interval, and consider events in that interval which display the following properties:

1. The random number of events arrivals in disjoint time intervals of (t, T) are independent.

2. Given two arbitrary disjoint but equal time intervals in (t, T), the probability of a given random number of events arrivals is the same in each interval.

3. The probability that at least two events occur *simultaneously* in any time interval is zero.

Next, let $P_k(\tau - t)$ be the probability that k events arrive in the time interval $\tau - t$. We make use of the previous three properties to determine the functional form of $P_k(\tau - t)$. First, $P_k(\tau - t)$ must satisfy:

$$P_0(\tau + d\tau - t) = P_0(\tau - t)\, P_0(d\tau)\,, \tag{4.44}$$

where $d\tau$ is a small perturbation of $\tau - t$, and we impose

$$\begin{cases} P_0(0) & = & 1 \\ P_k(0) & = & 0,\ k \geq 1 \end{cases} \tag{4.45}$$

Eq. (6.67) and the first relation of (6.68) are simultaneously satisfied by $P_0(x) = e^{-v \cdot x}$, with $v > 0$ in order to ensure that $P_0 \in [0, 1]$. Furthermore,

$$\begin{cases} P_1(\tau + d\tau - t) & = & P_0(\tau - t)\, P_1(d\tau) + P_1(\tau - t)\, P_0(d\tau) \\ & \vdots & \\ P_k(\tau + d\tau - t) & = & P_{k-1}(\tau - t)\, P_1(d\tau) + P_k(\tau - t)\, P_0(d\tau) \\ & \vdots & \end{cases} \tag{4.46}$$

Rearrange the first equation of (6.69) to have:

$$\frac{P_1(\tau + d\tau - t) - P_1(\tau - t)}{d\tau} = -\frac{1 - P_0(d\tau)}{d\tau} P_1(\tau - t) + \frac{P_1(d\tau)}{d\tau} P_0(\tau - t)$$

Taking the limits of the previous expression for $d\tau \to 0$ gives

$$P_1'(\tau - t) = \left(-\lim_{d\tau \to 0} \frac{1 - P_0(d\tau)}{d\tau}\right) P_1(\tau - t) + \left(\lim_{d\tau \to 0} \frac{P_1(d\tau)}{d\tau}\right) P_0(\tau - t)\,. \tag{4.47}$$

Now, we have that:

$$\frac{1 - P_0(d\tau)}{d\tau} = \frac{1 - e^{-v \cdot d\tau}}{d\tau}, \tag{4.48}$$

and for small $d\tau$, $P_1(d\tau) = 1 - P_0(d\tau)$, whence

$$\frac{P_1(d\tau)}{d\tau} = \frac{1 - P_0(d\tau)}{d\tau} = \frac{1 - e^{-v \cdot d\tau}}{d\tau}. \tag{4.49}$$

Taking the limit of (6.71)-(6.72) for $d\tau \to 0$ gives $\lim_{d\tau \to 0} \frac{1 - P_0(d\tau)}{d\tau} = \lim_{d\tau \to 0} \frac{P_1(d\tau)}{d\tau} = v$, and substituting back into (6.70) yields the result that:

$$P_1'(\tau - t) = -vP_1(\tau - t) + vP_0(\tau - t)\,. \tag{4.50}$$

Alternatively, eq. (6.73) can be arrived at by noting that for small $d\tau$, a Taylor approximation of $P_0(d\tau)$ is:

$$P_0(d\tau) = 1 - vd\tau + O\left(d\tau^2\right) \simeq 1 - vd\tau,$$

whence

$$\begin{cases} P_0(d\tau) &= 1 - v d\tau \\ P_1(d\tau) &= v d\tau \end{cases} \tag{4.51}$$

and plugging the previous equations directly into the first equation in (6.69), rearranging terms and taking $d\tau \to 0$ gives exactly (6.73).

Repeating the same reasoning used to arrive at (6.73), one shows that

$$P_k'(\tau - t) = -vP_k(\tau - t) + vP_{k-1}(\tau - t),$$

the solution of which is:

$$P_k(\tau - t) = \frac{v^k (\tau - t)^k}{k!} e^{-v(\tau - t)}. \tag{4.52}$$

4.6.2 Interpretation

The common interpretation of the process described in the previous subsection is the one of *rare events*, and eqs. (6.74) reveal that within this interpretation,

$$E\,(\text{event arrival in } d\tau) = v d\tau,$$

so that v can be thought of as an *intensity* of events arrivals.

Now we turn to provide a precise mathematical justification to the rare events interpretation. Consider the binomial distribution:

$$P_{n,k} = \binom{n}{k} p^k q^{n-k} = \frac{n!}{k!\,(n-k)!} p^k q^{n-k}, \quad p > 0,\ q > 0,\ p + q = 1,$$

which is the probability of k “arrivals” on n trials. We want to consider probability p as being a function of n with the special feature that

$$\lim_{n \to \infty} p(n) = 0.$$

This is a condition that is consistent with the rare event interpretation. One possible choice for such a p is

$$p(n) = \frac{a}{n}, \ \ a > 0.$$

We have,

$$\begin{aligned} P_{n,k} &= \frac{n!}{k!\,(n-k)!} p(n)^k (1 - p(n))^{n-k} \\ &= \frac{n!}{k!\,(n-k)!} \left(\frac{a}{n}\right)^k \left(1 - \frac{a}{n}\right)^{n-k} \\ &= \frac{n!}{k!\,(n-k)!} \left(\frac{a}{n}\right)^k \left(1 - \frac{a}{n}\right)^{n} \left(1 - \frac{a}{n}\right)^{-k} \\ &= \frac{n!}{n^k\,(n-k)!} \frac{a^k}{k!} \left(1 - \frac{a}{n}\right)^{n} \left(1 - \frac{a}{n}\right)^{-k} \\ &= \underbrace{\frac{n}{n} \cdot \frac{n-1}{n} \cdots \frac{n-k+1}{n}}_{k \text{ times}} \frac{a^k}{k!} \left(1 - \frac{a}{n}\right)^{n} \left(1 - \frac{a}{n}\right)^{-k}. \end{aligned}$$

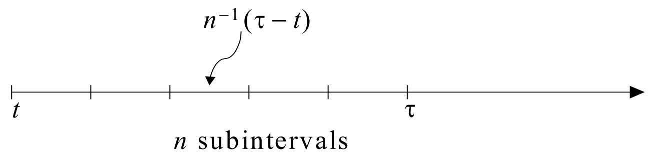


FIGURE 4.1. Heuristic construction of a Poisson process from the binomial distribution.

Therefore,

$$\lim_{n\to\infty} P_{n,k} \equiv P_k = \frac{a^k}{k!}e^{-a}.$$

Next, we split the $(\tau - t)$-interval into n subintervals of length $\frac{\tau-t}{n}$, and then making the probability of one arrival in each subinterval be proportional to each subinterval length (see figure 4.1), viz.,

$$p(n) = v\frac{\tau - t}{n} \equiv \frac{a}{n}, \qquad a \equiv v(\tau - t).$$

The process described in the previous section is thus obtained by taking $n \to \infty$, which is continuous time because in this case each subinterval in figure 6.8 shrinks to $d\tau$. In this case, the interpretation is that the probability that there is one arrival in $d\tau$ is $vd\tau$, and this is also the expected number of events in $d\tau$ because

$$\begin{aligned} E\left(\#\text{ arrivals in } d\tau\right) &= \Pr\left(\text{one arrival in } d\tau\right) \times \text{one arrival} \\ &\quad + \Pr\left(\text{zero arrivals in } d\tau\right) \times \text{zero arrivals} \\ &= \Pr\left(\text{one arrival in } d\tau\right) \times 1 + \Pr\left(\text{zero arrivals in } d\tau\right) \times 0 \\ &= vd\tau. \end{aligned}$$

Such a heuristic construction is also one that works very well to simulate Poisson processes: You just simulate a Uniform random variable $U(0,1)$, and the continuous-time process is replaced by the approximation Y, where:

$$Y = \begin{cases} 0 & \text{if} \quad 0 \leq U < 1 - vh \\ 1 & \text{if} \quad 1 - vh \leq U < 1 \end{cases}$$

where h is the discretization interval.

4.6.3 Properties and related distributions

First we verify that P_k is a probability. We have

$$\sum_{k=0}^{\infty} P_k = e^{-a}\sum_{k=0}^{\infty}\frac{a^k}{k!} = 1,$$

since $\sum_{k=0}^{\infty} a^k / k!$ is the McLaurin expansion of function e^a.

Second, we compute the mean,

$$\text{mean} = \sum_{k=0}^{\infty} k \cdot P_k = e^{-a}\sum_{k=0}^{\infty} k \cdot \frac{a^k}{k!} = a,$$

where the last equality is due to the fact that $\sum_{k=0}^{\infty}\left(ka^k\right)/k! = ae^a$.[6]

A related distribution is the so-called *exponential* (or *Erlang*) distribution. We already know that the probability of zero arrivals in $\tau - t$ is given by $P_0(\tau - t) = e^{-v(\tau-t)}$ [see eq. (6.75)]. It follows that:

$$G(\tau - t) \equiv 1 - P_0(\tau - t) = 1 - e^{-v(\tau-t)}$$

is the probability of at least one arrival in $\tau - t$. Also, this can be interpreted as the probability that the first arrival occurred before τ starting from t.

The density function g of this distribution is

$$g(\tau - t) = \frac{\partial}{\partial\tau}G(\tau - t) = ve^{-v(\tau-t)}. \tag{4.53}$$

Clearly, $\int_0^\infty v \cdot e^{-v \cdot x}dx = 1$, which confirms that g is indeed a density function. From this, we may easily compute,

$$\begin{aligned}\text{mean} &= \int_0^\infty xve^{-vx}dx = v^{-1} \\ \text{variance} &= \int_0^\infty \left(x - v^{-1}\right)^2 ve^{-vx}dx = v^{-2}\end{aligned}$$

The expected time of the first arrival occurred before τ starting from t equals v^{-1}. More generally, v^{-1} can be interpreted as the average time from an arrival to another.[7]

A more general distribution in modeling such issues is the Gamma distribution with density:

$$g_\gamma(\tau - t) = ve^{-v(\tau-t)}\frac{[v(\tau - t)]^{\gamma-1}}{(\gamma - 1)!},$$

which collapses to density g in (6.76) when $\gamma = 1$.

4.6.4 Asset pricing implications

A natural application of the previous schemes is to model asset prices as driven by Brownian motions plus jumps processes. To model jumps, we just interpret an "arrival" as the event that a certain variable experiences a jump of size $\mathcal{S}$, where $\mathcal{S}$ is another random variable with a fixed probability measure p on $\mathbb{R}^d$. A simple (unidimensional, i.e. $d = 1$) model of an asset price is then

$$dq(\tau) = b(q(\tau))d\tau + \sqrt{2\sigma(q(\tau))}dW(\tau) + \ell(q(\tau)) \cdot \mathcal{S} \cdot dZ(\tau), \tag{4.54}$$

where b, σ, ℓ are given functions (with $\sigma > 0$), W is a standard Brownian motion, and Z is a Poisson process with intensity function, or hazard rate, equal to v, i.e.

1. $\Pr(Z(t)) = 0$.
2. $\forall t \leq \tau_0 < \tau_1 < \cdots < \tau_N < \infty$, $Z(\tau_0)$ and $Z(\tau_k) - Z(\tau_{k-1})$ are independent for each $k = 1, \cdots, N$.

---

[6] Indeed, $e^a = \frac{\partial e^a}{\partial a} = \frac{\partial}{\partial a}\sum_{k=0}^{\infty}\frac{a^k}{k!} = \sum_{k=0}^{\infty}k\frac{a^{k-1}}{k!} = a^{-1}\sum_{k=0}^{\infty}k\frac{a^k}{k!}$.

[7] Suppose arrivals are generated by Poisson processes, consider the random variable "time interval elapsing from one arrival to next one", and let τ' be the instant at which the last arrival occurred. The probability that the time $\tau - \tau'$ which will elapse from the last arrival to the next one is less than Δ is then the same as the probability that during the time interval $\overline{\tau'\tau} = \tau - \tau'$ there is at least one arrival.

3. $\forall\tau > t$, $Z(\tau) - Z(t)$ is a random variable with Poisson distribution and expected value $v(\tau - t)$, i.e.
$$\Pr\left(Z(\tau) - Z(t) = k\right) = \frac{v^k\left(\tau - t\right)^k}{k!}e^{-v\cdot(\tau-t)}. \tag{4.55}$$
In the framework considered here, k is interpreted as the number of jumps.[8]

From (6.78), it follows that $\Pr\left(Z(\tau) - Z(t) = 1\right) = v\cdot(\tau-t)\cdot e^{-v\cdot(\tau-t)}$ and by letting $\tau \to t$, we (heuristically) get that:
$$\Pr\left(dZ(\tau) = 1\right) \equiv \Pr\left(\left. Z(\tau) - Z(t)\right|_{\tau\to t} = 1\right) = \left. v\left(\tau - t\right)e^{-v\cdot(\tau-t)}\right|_{\tau\to t} \simeq v\cdot d\tau,$$
because e^{-vx} is of smaller order than x when x is very small.

More generally, the process
$$\left\{Z(\tau) - v\left(\tau - t\right)\right\}_{\tau\geq t}$$
is a martingale.

Armed with these preliminary definitions and interpretation, we now provide a heuristic derivation of Itô's lemma for jump-diffusion processes. Consider any function f which enjoys enough regularity conditions which is a *rational* function of the state in (6.77), i.e.
$$f(\tau) \equiv f\left(q(\tau), \tau\right).$$
Consider the following expansion of f:
$$\begin{aligned} df(\tau) &= \left(\frac{\partial}{\partial\tau} + L\right) f\left(q(\tau),\tau\right)d\tau + f_q\left(q(\tau),\tau\right)\sqrt{2\sigma(q(t))}dW(\tau) \\ &\quad + \left[f\left(q(\tau) + \ell(q(\tau))\cdot\mathcal{S},\tau\right) - f\left(q(\tau),\tau\right)\right]\cdot dZ(\tau). \end{aligned} \tag{4.56}$$
The first two terms in (6.79) are the usual Itô's lemma terms, with $\frac{\partial}{\partial\tau}\cdot + L\cdot$ denoting the usual infinitesimal generator for diffusions. The third term accounts for jumps: if there are no jumps from time τ_- to time τ (where $d\tau = \tau - \tau_-$), then $dZ(\tau) = 0$; but if there is one jump (remember that there can not be more than one jumps because in $d\tau$, only one event can occur), then $dZ(\tau) = 1$, and in this case f, as a "rational" function, also instantaneously jumps to $f\left(q(\tau) + \ell(q(\tau))\cdot\mathcal{S},\tau\right)$, and the jump will then be exactly $f\left(q(\tau) + \ell(q(\tau))\cdot\mathcal{S},\tau\right) - f\left(q(\tau),\tau\right)$, where $\mathcal{S}$ is another random variable with a fixed probability measure. Clearly, by choosing $f(q,\tau) = q$ gives back the original jump-diffusion model (6.77).

From this, we can compute $L^J f$ and the infinitesimal generator for jumps diffusion models:
$$\begin{aligned} E\left\{df\right\} &= \left(\frac{\partial}{\partial\tau} + L\right) f d\tau + E\left\{\left\{f\left(q + \ell\mathcal{S},\tau\right) - f\left(q,\tau\right)\right\}\cdot dZ(\tau)\right\} \\ &= \left(\frac{\partial}{\partial\tau} + L\right) f d\tau + E\left\{\left\{f\left(q + \ell\mathcal{S},\tau\right) - f\left(q,\tau\right)\right\}\cdot v\cdot d\tau\right\}, \end{aligned}$$
or
$$L^J f = Lf + v\cdot\int_{\mathrm{supp}(\mathcal{S})}\left[f\left(q + \ell\mathcal{S},\tau\right) - f\left(q,\tau\right)\right]p\left(d\mathcal{S}\right),$$
so that the infinitesimal generator for jumps diffusion models is just $\left(\frac{\partial}{\partial\tau} + L^J\right)f$.

[8] For simplicity sake, here we are considering the case in which v is a constant. If v is a deterministic function of time, we have that
$$\Pr\left(Z(\tau) - Z(t) = k\right) = \frac{\left(\int_t^\tau v(u)du\right)^k}{k!}\exp\left(-\int_t^\tau v(u)du\right), \quad k = 0, 1, \cdots$$
and there is also the possibility to model v as a function of the state: $v = v(q)$, for example.

4.6.5 An option pricing formula

Merton (1976, JFE), Bates (1988, working paper), Naik and Lee (1990, RFS) are the seminal papers.

4.7 Continuous-time Markov chains

4.8 General equilibrium

4.9 Incomplete markets

4.10 Appendix 1: Convergence issues

We have,

$$c_t + q_t \theta_{t+1}^{(1)} + b_t \theta_{t+1}^{(2)} = (q_t + D_t)\,\theta_t^{(1)} + b_t \theta_t^{(2)} \equiv V_t + D_t \theta_t^{(1)},$$

where $V_t \equiv q_t \theta_t^{(1)} + b_t \theta_t^{(2)}$ is wealth net of dividends. We have,

$$\begin{aligned} V_t - V_{t-1} &= q_t \theta_t^{(1)} + b_t \theta_t^{(2)} - V_{t-1} \\ &= q_t \theta_t^{(1)} + b_t \theta_t^{(2)} - \left(c_{t-1} + q_{t-1} \theta_t^{(1)} + b_{t-1} \theta_t^{(2)} - D_{t-1} \theta_{t-1}^{(1)} \right) \\ &= (q_t - q_{t-1})\,\theta_t^{(1)} + (b_t - b_{t-1})\,\theta_t^{(2)} - c_{t-1} + D_{t-1} \theta_{t-1}^{(1)}, \end{aligned}$$

and more generally,

$$V_t - V_{t-\Delta} = (q_t - q_{t-\Delta})\,\theta_t^{(1)} + (b_t - b_{t-\Delta})\,\theta_t^{(2)} - (c_{t-\Delta} \cdot \Delta) + (D_{t-\Delta} \cdot \Delta)\,\theta_{t-\Delta}^{(1)}.$$

Now let $\Delta \downarrow 0$ and assume that $\theta^{(1)}$ and $\theta^{(2)}$ are approximately constant between t and $t - \Delta$. We have:

$$dV(\tau) = (dq(\tau) + D(\tau)d\tau)\,\theta^{(1)}(\tau) + db(\tau)\theta^{(2)}(\tau) - c(\tau)d\tau.$$

Assume that

$$\frac{db(\tau)}{b(\tau)} = r d\tau.$$

The budget constraint can then be written as:

$$\begin{aligned} dV(\tau) &= (dq(\tau) + D(\tau)d\tau)\,\theta^{(1)}(\tau) + rb(\tau)\theta^{(2)}(\tau)d\tau - c(\tau)d\tau \\ &= (dq(\tau) + D(\tau)d\tau)\,\theta^{(1)}(\tau) + r\left(V - q(\tau)\theta^{(1)}(\tau) \right) d\tau - c(\tau)d\tau \\ &= (dq(\tau) + D(\tau)d\tau - rq(\tau)d\tau)\,\theta^{(1)}(\tau) + rVd\tau - c(\tau)d\tau \\ &= \left(\frac{dq(\tau)}{q(\tau)} + \frac{D(\tau)}{q(\tau)}d\tau - rd\tau \right) \theta^{(1)}(\tau)q(\tau) + rVd\tau - c(\tau)d\tau \\ &= \left(\frac{dq(\tau)}{q(\tau)} + \frac{D(\tau)}{q(\tau)}d\tau - rd\tau \right) \pi(\tau) + rVd\tau - c(\tau)d\tau. \end{aligned}$$

4.11 Appendix 2: Walras consistency tests

(4.15) $\Rightarrow$ *(4.16)*. We can grasp the intituion of the result by better understanding the same phenomenon in the case of the two-period economic studied in chapter 2. There, we showed that in the absence of arbitrage opportunities, $\exists \phi \in \mathbb{R}^d : \phi^\top (c^1 - w^1) = q\theta = -(c_0 - w_0)$. Whence $c_s = w_s, s = 0, \cdots, d \Rightarrow \theta = \mathbf{0}_m$.

In the model of the present chapter, we have that in the absence of arbitrage opportunities, $\exists!$ $Q \in \mathcal{P}$: (4.7) is satisfied, or

$$\begin{aligned} \left(q_0^{-1} V^{x,\pi,c} \right)(\tau) &\equiv \frac{1}{q_0(\tau)} \left(\theta_0(\tau) q_0(\tau) + \pi^\top(\tau) \mathbf{1}_m \right) \\ &= \mathbf{1}_m^\top q(t) + \int_t^\tau \left(q_0^{-1} \pi^\top \sigma \right)(u) dW^*(u) - \int_t^\tau \left(q_0^{-1} c \right)(u) du, \ \tau \in [t, T]. \end{aligned}$$

Let us rewrite the previous relations as:

for $\tau \in [t, T]$, $\quad \dfrac{1}{q_0(\tau)} \left(\theta_0(\tau) q_0(\tau) + \left(\pi^\top(\tau) - q^\top(\tau) \right) \mathbf{1}_m \right) + \dfrac{1}{q_0(\tau)} q^\top(\tau) \mathbf{1}_m$

$$= \mathbf{1}_m^\top q(t) + \int_t^\tau \frac{1}{q_0(u)} \left(\pi^\top(u) - q^\top(u)\right) \sigma(u) dW^*(u) - \int_t^\tau \frac{1}{q_0(u)} c(u) du + \int_t^\tau \frac{1}{q_0(u)} q^\top(u)\sigma(u)dW^*(u)$$

By plugging the solution $\left(\frac{q_i}{q_0}\right)(\tau) = q_i(t) + \int_t^\tau \left(q_0^{-1} q_i\right)(u)\sigma_i(u)dW^*(u) - \int_t^\tau \left(q_0^{-1} D_i\right)(u)du$ in the previous relation,

$$\begin{aligned} & \frac{1}{q_0(\tau)} \left(\theta_0(\tau) q_0(\tau) + \left(\pi^\top(\tau) - q^\top(\tau)\right) \mathbf{1}_m\right) \\ = & \int_t^\tau \frac{1}{q_0(u)} \left(\pi^\top(u) - q^\top(u)\right) \sigma(u) dW^*(u) + \int_t^\tau \frac{1}{q_0(u)} \left(D(u) - c(u)\right) du. \end{aligned} \quad (4.57)$$

Let us rewrite the previous relation in $t = T$,

$$\begin{aligned} & \frac{1}{q_0(T)} \left(\theta_0(T) q_0(T) + \left(\pi^\top(T) - q^\top(T)\right) \mathbf{1}_m\right) \\ = & \int_t^\tau \frac{1}{q_0(u)} \left(\pi^\top(u) - q^\top(u)\right) \sigma(u) dW^*(u) + \int_t^\tau \frac{1}{q_0(u)} \left(D(u) - c(u)\right) du. \end{aligned} \quad (4.58)$$

When (4.15) is verified, we have that $V^{x,\pi,c}(T) = \theta_0(T) q_0(T) + \pi^\top(T)\mathbf{1}_m = v = \underline{q}(T) = q^\top(T)\mathbf{1}_m$, and $D = c$, and the previous equality becomes:

$$0 = x(T) \equiv \int_t^T \frac{1}{q_0(u)} \left(\pi^\top(u) - q^\top(u)\right) \sigma(u) dW^*(u),\ Q\text{-a.s.}$$

a martingale starting at zero, the incremental process of which satisfies:

$$dx(\tau) = \frac{1}{q_0(\tau)} \left(\pi^\top(\tau) - q^\top(\tau)\right) \sigma(\tau) dW^*(\tau) = 0,\ \tau \in [t, T].$$

Since $\ker(\sigma) = \{\emptyset\}$ then, we have that $\pi(\tau) = q(\tau)$ a.s. for $\tau \in [t, T]$ and, hence, $\pi(\tau) = q(\tau)$ a.s. for $\tau \in [t, T]$. It is easily checked that this implies $\theta_0(T) = 0$ P-a.s. and that in fact (with the help of (4.18)), $\theta_0(\tau) = 0$ a.s. $\|$

(4.15) $\Leftarrow$ *(4.16)*. When (4.16) holds, Eq. (4.19) becomes:

$$0 = y(T) \equiv \int_t^T \frac{1}{q_0(u)} \left(D(u) - c(u)\right) du,\ \text{a.s.},$$

a martingale starting at zero. It is sufficient to apply verbatim to $\{y(\tau)\}_{\tau \in [t,T]}$ the same arguments of the proof of the previous part to conclude. $\|$

4.12 Appendix 3: The Green's function

This approach can be useful in applied work (see, for example, Mele and Fornari (2000, chapter 5).

Setup

In section 4.5, we showed that the value of a security as of time τ is:

$$V\left(x(\tau),\tau\right)=E\left[\frac{m_{t,T}}{m_{t,\tau}}V\left(x(T),T\right)+\int_{\tau}^{T}\frac{m_{t,s}}{m_{t,\tau}}h\left(x(s),s\right)ds\right], \tag{5A.2}$$

where $m_{t,\tau}$ is the stochastic discount factor,

$$m_{t,\tau}=q_0(\tau)^{-1}\eta(t,\tau)=q_0(\tau)^{-1}\cdot\left.\frac{dQ}{dP}\right|_{\mathcal{F}(\tau)},\quad q_0(\tau)\equiv e^{\int_t^\tau r(u)du}.$$

Arrow-Debreu state price densities are:

$$\phi_{t,T}=m_{t,T}dP=q_0(T)^{-1}dQ,$$

which now we want to characterize in terms of PDE.

The first thing to note is that by reiterating the same reasoning produced in section 4.5, eq. (5A.2) can be rewritten as:

$$V\left(x(\tau),\tau\right)=\mathbb{E}\left[\frac{q_0(\tau)}{q_0(T)}V\left(x(T),T\right)+\int_{\tau}^{T}\frac{q_0(\tau)}{q_0(s)}h\left(x(s),s\right)ds\right]. \tag{5A.3}$$

Let

$$a\left(t',t''\right)=\frac{q_0(t')}{q_0(t'')}.$$

In terms of a, eq. (5A.3) is:

$$V\left(x(\tau),\tau\right)=\mathbb{E}\left[a\left(\tau,T\right)V\left(x(T),T\right)+\int_{\tau}^{T}a\left(\tau,s\right)h\left(x(s),s\right)ds\right].$$

Next, consider the augmented state vector:

$$y(u)\equiv\left(a\left(\tau,u\right),\ x(u)\right),\quad \tau\leq u\leq T,$$

and let $P\left(y(t'')\right|y(\tau))$ be the density function of the augmented state vector under the risk-neutral probability. We have,

$$\begin{aligned}V\left(x(\tau),\tau\right)&=\mathbb{E}\left[a\left(\tau,T\right)V\left(x(T),T\right)+\int_{\tau}^{T}a\left(\tau,s\right)h\left(x(s),s\right)ds\right]\\&=\int a\left(\tau,T\right)V\left(x(T),T\right)P\left(y(T)\right|y(\tau))\,dy(T)+\int_{\tau}^{T}\int a\left(\tau,s\right)h\left(x(s),s\right)P\left(y(s)\right|y(\tau))\,dy(s)ds.\end{aligned}$$

If $V\left(x(T),T\right)$ and $a\left(\tau,T\right)$ are independent,

$$\begin{aligned}\int a\left(\tau,T\right)V\left(x(T),T\right)P\left(y(T)\right|y(\tau))\,dy(T)&=\int_X\left[\int_A a\left(\tau,T\right)P\left(y(T)\right|y(\tau))\,dy(T)\right]V\left(x(T),T\right)dx(T)\\&\equiv\int_X G(\tau,T)V\left(x(T),T\right)dx(T)\end{aligned}$$

where:

$$G(\tau,T) \equiv \int_A a(\tau,T)\, P\left(y(T)|\, y(\tau)\right) dy(T).$$

If we assume the same thing for h, we eventually get:

$$V\left(x(\tau),\tau\right) = \int_X G(\tau,T) V\left(x(T),T\right) dx(T) + \int_\tau^T \int_X G(\tau,s) h\left(x(s),s\right) dx(s) ds.$$

G is known as the *Green's function*:

$$G\left(t,\ell\right) \equiv G\left(x,t;\xi,\ell\right) = \int_A a\left(t,\ell\right) P\left(y(\ell)|\, y(t)\right) da.$$

It is the value in state $x \in \mathbb{R}^d$ as of time t of a unit of numéraire at $\ell > t$ if future states lie in a neighborhood (in $\mathbb{R}^d$) of ξ. It is thus the Arrow-Debreu state-price density.

For example, a pure discount bond has $V(x,T) = 1\ \forall x$, and $h(x,s) = 1\ \forall x, s$, and

$$V\left(x(\tau),\tau\right) = \int_X G\left(x(\tau),\tau;\xi,T\right) d\xi,$$

with

$$\lim_{\tau\uparrow T} G\left(x(\tau),\tau;\xi,T\right) = \delta\left(x(\tau)-\xi\right),$$

where δ is the Dirac delta.

The PDE connection

We have,

$$V\left(x\left(t\right),t\right) = \int_X G\left(x(t),t;\xi\left(T\right),T\right) V\left(\xi\left(T\right),T\right) d\xi\left(T\right) + \int_t^T \int_X G(x(t),t;\xi(s),s) h\left(\xi(s),s\right) d\xi(s) ds. \tag{5A.4}$$

Consider the scalar case. Eq. (5A.3) and the connection between PDEs and Feynman-Kac tell us that under standard regularity conditions, V is solution to:

$$0 = V_t + \mu V_x + \frac{1}{2}\sigma^2 V_{xx} - rV + h. \tag{5A.5}$$

Now take the appropriate partial derivatives in (4.A4),

$$\begin{aligned}
V_t &= \int_X G_t V d\xi - \int_X \delta(x-\xi) h d\xi + \int_t^T \int_X G_t h d\xi ds = \int_X G_t V d\xi - h + \int_\tau^T \int_X G_t h d\xi ds \\
V_x &= \int_X G_x V d\xi + \int_t^T \int_X G_x h d\xi ds \\
V_{xx} &= \int_X G_{xx} V d\xi + \int_t^T \int_X G_{xx} h d\xi ds
\end{aligned}$$

and replace them into (4.A5) to obtain:

$$\begin{aligned}
0 &= \int_X \left[G_t + \mu G_x + \frac{1}{2}\sigma^2 G_{xx} - rG\right] V\left(\xi\left(T\right),T\right) d\xi(T) \\
&\quad + \int_t^T \int_X \left[G_t + \mu G_x + \frac{1}{2}\sigma^2 G_{xx} - rG\right] h\left(\xi\left(s\right),s\right) d\xi\left(s\right) ds.
\end{aligned}$$

This shows that G is solution to

$$0 = G_t + \mu G_x + \frac{1}{2}\sigma^2 G_{xx} - rG,$$

and

$$\lim_{t \uparrow T} G(x, t; \xi, T) = \delta(x - \xi).$$

The multidimensional case is dealt with a mere change in notation.

4.13 Appendix 4: Models with final consumption only

Sometimes, we may be interested in models with consumption taking place in at the end of the period only. Let $\bar{q} = \left(q^{(0)}, q\right)^{\intercal}$ and $\bar{\theta} = (\theta^{(0)}, \theta)$, where θ and q are both m-dimensional. Define as usual wealth as of time t as $V_t \equiv \bar{q}_t \bar{\theta}_t$. There are no dividends. A *self-financing* strategy $\bar{\theta}$ satisfies,

$$\bar{q}_t^{+} \bar{\theta}_{t+1} = \bar{q}_t \bar{\theta}_t \equiv V_t, \quad t = 1, \cdots, T.$$

Therefore,

$$\begin{aligned} V_t &= \bar{q}_t \bar{\theta}_t + \bar{q}_{t-1} \bar{\theta}_{t-1} - \bar{q}_{t-1} \bar{\theta}_{t-1} \\ &= \bar{q}_t \bar{\theta}_t + \bar{q}_{t-1} \bar{\theta}_{t-1} - \bar{q}_{t-1} \bar{\theta}_t \quad \text{(because } \bar{\theta} \text{ is self-financing)} \\ &= V_{t-1} + \Delta \bar{q}_t \bar{\theta}_t, \quad \Delta \bar{q}_t \equiv \bar{q}_t - \bar{q}_{t-1}, \quad t = 1, \cdots, T. \end{aligned}$$

$\Leftrightarrow$

$$V_t = V_1 + \sum_{n=1}^{t} \Delta \bar{q}_n \bar{\theta}_n.$$

Now suppose that

$$\Delta q_t^{(0)} = r_t q_{t-1}^{(0)}, \quad t = 1, \cdots, T,$$

with $\{r_t\}_{t=1}^{T}$ given and to be defined more precisely below. The term $\Delta \bar{q}_t \theta_t^{+}$ can then be rewritten as:

$$\begin{aligned} \Delta \bar{q}_t \bar{\theta}_t &= \Delta q_t^{(0)} \theta_t^{(0)} + \Delta q_t \theta_t \\ &= r_t q_{t-1}^{(0)} \theta_t^{(0)} + \Delta q_t \theta_t \\ &= r_t q_{t-1}^{(0)} \theta_t^{(0)} + r_t q_{t-1} \theta_t - r_t q_{t-1} \theta_t + \Delta q_t \theta_t \\ &= r_t \bar{q}_{t-1} \bar{\theta}_t - r_t q_{t-1} \theta_t + \Delta q_t \theta_t \\ &= r_t \bar{q}_{t-1} \bar{\theta}_{t-1} - r_t q_{t-1} \theta_t + \Delta q_t \theta_t \quad \text{(because } \bar{\theta} \text{ is self-financing)} \\ &= r_t V_{t-1} - r_t q_{t-1} \theta_t + \Delta q_t \theta_t, \end{aligned}$$

and we obtain

$$V_t = (1 + r_t) V_{t-1} - r_t q_{t-1} \theta_t + \Delta q_t \theta_t, \tag{4.59}$$

or by integrating,

$$V_t = V_1 + \sum_{n=1}^{t} \left(r_n V_{n-1} - r_n q_{n-1} \theta_n + \Delta q_n \theta_n \right). \tag{4.60}$$

Next, considering "small" time intervals. In the limit we obtain:

$$dV(t) = r(t)V(t)dt - r(t)q(t)\theta(t)dt + dq(t)\theta(t). \tag{4.61}$$

Such an equation can also be arrived at by noticing that current wealth is nothing but initial wealth plus gains from trade accumulated up to now:

$$V(t) = V(0) + \int_0^t d\bar{q}(u)\bar{\theta}(u).$$

$\Leftrightarrow$

$$\begin{aligned} dV(t) &= d\bar{q}(t)\bar{\theta}(t)^{+} \\ &= dq_0(t)\theta_0(t) + dq(t)\theta(t) \\ &= r(t)q_0(t)\theta_0(t)dt + dq(t)\theta(t) \\ &= r(t)\left(V(t) - q(t)\theta(t)\right)dt + dq(t)\theta(t) \\ &= r(t)V(t)dt - r(t)q(t)\theta(t)dt + dq(t)\theta(t). \end{aligned}$$

Now consider the sequence of problems of terminal wealth maximization:

$$\text{For } t = 1, \cdots, T, \quad \mathcal{P}_t : \begin{cases} \max_{\theta_t} E\left[u(V(T))\right|\mathcal{F}_{t-1}], \\ \text{s.t. } V_t = (1 + r_t)\,V_{t-1} - r_t q_{t-1}\theta_t + \Delta q_t \theta_t \end{cases}$$

Even if markets are incomplete, agents can solve the sequence of problems $\{\mathcal{P}_t\}_{t=1}^T$ as time unfolds. Using (2.?), each problem can be written as:

$$\max_{\theta_t} E\left[u\left(V_1 + \sum_{t=1}^{T} (r_t V_{t-1} - r_t q_{t-1}\theta_t + \Delta q_t \theta_t) \right) \middle| \mathcal{F}_{t-1} \right].$$

The FOC for $t = 1$ is:

$$E\left[u'(V(T))\,(q_1 - (1 + r_0)\,q_0) \middle| \mathcal{F}_0 \right],$$

whence

$$q_0 = (1 + r_0)^{-1} \frac{E\left[u'(V(T)) \cdot q_1 \right| \mathcal{F}_0]}{E\left[u'(V(T))\right| \mathcal{F}_0]},$$

and in general

$$q_t = (1 + r_t)^{-1} \frac{E\left[u'(V(T)) \cdot q_{t+1} \right| \mathcal{F}_t]}{E\left[u'(V(T))\right| \mathcal{F}_t]}, \quad t = 0, \cdots, T - 1.$$

In deriving such relationships, we were close to the *dynamic programming principle*. More on this below. The previous relationship suggests that we can define a *martingale measure* Q for the discounted price process by defining

$$\left.\frac{dQ}{dP}\right|_{\mathcal{F}_t} = \frac{u'(V(T))}{E\left[u'(V(T))\right| \mathcal{F}_t]}.$$

More on this below.

Connections with the CAPM. It's easy to show that:

$$E\left(\tilde{r}_{t+1}\right) - r_t = cov\left[\frac{u'(V(T))}{E\left[u'(V(T))\right| \mathcal{F}_t]}, \tilde{r}_{t+1} \right],$$

where $\tilde{r}_{t+1} \equiv (q_{t+1} - q_t)/\,q_t$.

4.14 Appendix 5: Further topics on jumps

4.14.1 The Radon-Nikodym derivative

To derive heuristically the Radon-Nikodym derivative, consider the jump times $0 < \tau_1 < \tau_2 < \cdots < \tau_n = \widehat{T}$. The probability of a jump in a neighborhood of τ_i is $v(\tau_i)d\tau$, and to find what happens under the risk-neutral probability or, in general, under any equivalent measure, we just write $v^Q(\tau_i)d\tau$ under measure Q, and set $v^Q = v\lambda^J$. Clearly, the probability of no-jumps between any two adjacent random points τ_{i-1} and τ_i *and* a jump at time τ_{i-1} is, for $i \geq 2$, proportional to

$$v(\tau_{i-1})e^{-\int_{\tau_{i-1}}^{\tau_i} v(u)du} \qquad \text{under the probability } P,$$

and to

$$v^Q(\tau_{i-1})e^{-\int_{\tau_{i-1}}^{\tau_i} v^Q(u)du} = v(\tau_{i-1})\lambda^J(\tau_{i-1})e^{-\int_{\tau_{i-1}}^{\tau_i} v(u)\lambda^J(u)du} \qquad \text{under the probability } Q.$$

As explained in section 6.9.3 (see also formula # 6.76)), *these are in fact densities of time intervals elapsing from one arrival to the next one.*

Next let A be the event of marks at time $\tau_1, \tau_2, \cdots, \tau_n$. The Radon-Nikodym derivative is the likelihood ratio of the two probabilities Q and P of A:

$$\frac{Q(A)}{P(A)} = \frac{e^{-\int_t^{\tau_1} v(u)\lambda^J(u)du} \cdot v(\tau_1)\lambda^J(\tau_1)e^{-\int_{\tau_1}^{\tau_2} v(u)\lambda^J(u)du} \cdot v(\tau_2)\lambda^J(\tau_2)e^{-\int_{\tau_2}^{\tau_3} v(u)\lambda^J(u)du} \cdot \cdots}{e^{-\int_t^{\tau_1} v(u)du} \cdot v(\tau_1)e^{-\int_{\tau_1}^{\tau_2} v(u)du} \cdot v(\tau_2)e^{-\int_{\tau_2}^{\tau_3} v(u)du} \cdot \cdots},$$

where we have used the fact that given that at $\tau_0 = t$, there are no-jumps, the probability of no-jumps from t to τ_1 is $e^{-\int_t^{\tau_1} v(u)du}$ under P and $e^{-\int_t^{\tau_1} v(u)\lambda^J(u)du}$ under Q, respectively. Simple algebra then yields,

$$\begin{aligned}
\frac{Q(A)}{P(A)} &= \lambda^J(\tau_1) \cdot \lambda^J(\tau_2) \cdot e^{-\int_t^{\tau_1} v(u)\left(\lambda^J(u)-1\right)du} \cdot e^{-\int_{\tau_1}^{\tau_2} v(u)\left(\lambda^J(u)-1\right)du} e^{-\int_{\tau_2}^{\tau_3} v(u)\left(\lambda^J(u)-1\right)du} \cdot \cdots \\
&= \prod_{i=1}^n \lambda^J(\tau_i) \cdot e^{-\int_t^{\tau_n} v(u)\left(\lambda^J(u)-1\right)du} \\
&= \exp\left[\log\left(\prod_{i=1}^n \lambda^J(\tau_i) \cdot e^{-\int_t^{\tau_n} v(u)\left(\lambda^J(u)-1\right)du}\right)\right] \\
&= \exp\left[\sum_{i=1}^n \log \lambda^J(\tau_i) - \int_t^{\tau_n} v(u)\left(\lambda^J(u)-1\right)du\right] \\
&= \exp\left[\int_t^{\widehat{T}} \log \lambda^J(u)dZ(u) - \int_t^{\widehat{T}} v(u)\left(\lambda^J(u)-1\right)du\right],
\end{aligned}$$

where the last equality follows from the definition of the Stieltjes integral.

The previous results can be used to say something substantive on an economic standpoint. But before, we need to simplify both presentation and notation. We have:

Definition (Doléans-Dade exponential semimartingale): *Let M be a martingale. The unique solution to the equation*:

$$L(\tau) = 1 + \int_t^{\tau} L(u)dM(u),$$

is called the Doléans-Dade exponential semimartingale *and is denoted as $\mathcal{E}(M)$.*

4.14.2 Arbitrage restrictions

As in the main text, let now q be the price of a primitive asset, solution to:

$$\begin{aligned}\frac{dq}{q} &= bd\tau + \sigma dW + \ell\mathcal{S}dZ \\ &= bd\tau + \sigma dW + \ell\mathcal{S}\left(dZ - vd\tau\right) + \ell\mathcal{S}vd\tau \\ &= \left(b + \ell\mathcal{S}v\right)d\tau + \sigma dW + \ell\mathcal{S}\left(dZ - vd\tau\right).\end{aligned}$$

Next, define

$$d\tilde{Z} = dZ - v^Q d\tau \ \left(v^Q = v\lambda^J\right); \quad d\tilde{W} = dW + \lambda d\tau.$$

Both $\tilde{Z}$ and $\tilde{W}$ are Q-martingales. We have:

$$\frac{dq}{q} = \left(b + \ell\mathcal{S}v^Q - \sigma\lambda\right)d\tau + \sigma d\tilde{W} + \ell\mathcal{S}d\tilde{Z}.$$

The characterization of the equivalent martingale measure for the discounted price is given by the following Radon-Nikodym density of Q w.r.t. P:

$$\frac{dQ}{dP} = \mathcal{E}\left(-\int_t^T \lambda(\tau)dW(\tau) + \int_t^T \left(\lambda^J(\tau) - 1\right)\left(dZ(\tau) - v(\tau)\right)d\tau\right),$$

where $\mathcal{E}(\cdot)$ is the Doléans-Dade exponential semimartingale, and so:

$$b = r + \sigma\lambda - \ell v^Q E_{\mathcal{S}}(\mathcal{S}) = r + \sigma\lambda - \ell v\lambda^J E_{\mathcal{S}}(\mathcal{S}).$$

Clearly, markets are incomplete here.

Exercise. Show that if $\mathcal{S}$ is deterministic, a representative agent with utility function $u(x) = \frac{x^{1-\eta}-1}{1-\eta}$ makes $\lambda^J(\mathcal{S}) = (1+\mathcal{S})^{-\eta}$.

4.14.3 State price density: introduction

We have:

$$L(T) = \exp\left[-\int_t^T v(\tau)\left(\lambda^J(\tau) - 1\right)d\tau + \int_t^T \log\lambda^J(\tau)dZ(\tau)\right].$$

The objective here is to use Itô's lemma for jump processes to express L in differential form. Define the jump process y as:

$$y(\tau) \equiv -\int_t^\tau v(u)\left(\lambda^J(u) - 1\right)du + \int_t^\tau \log\lambda^J(u)dZ(u).$$

In terms of y, L is $L(\tau) = l(y(\tau))$ with $l(y) = e^y$. We have:

$$\begin{aligned}dL(\tau) &= -e^{y(\tau)}v(\tau)\left(\lambda^J(\tau) - 1\right)d\tau + \left(e^{y(\tau)+\text{jump}} - e^{y(\tau)}\right)dZ(\tau) \\ &= -e^{y(\tau)}v(\tau)\left(\lambda^J(\tau) - 1\right)d\tau + e^{y(\tau)}\left(e^{\log\lambda^J(\tau)} - 1\right)dZ(\tau)\end{aligned}$$

or,

$$\frac{dL(\tau)}{L(\tau)} = -v(\tau)\left(\lambda^J(\tau) - 1\right)d\tau + \left(\lambda^J(\tau) - 1\right)dZ(\tau) = \left(\lambda^J(\tau) - 1\right)\left(dZ(\tau) - v(\tau)d\tau\right).$$

The general case (with stochastic distribution) is covered in the following subsection.

4.14.4 State price density: general case

Suppose the primitive is:

$$dx(\tau) = \mu(x(\tau_-))d\tau + \sigma(x(\tau_-))dW(\tau) + dZ(\tau),$$

and that u is the price of a derivative. Introduce the P-martingale,

$$dM(\tau) = dZ(\tau) - v(x(\tau))d\tau.$$

By Itô's lemma for jump-diffusion processes,

$$\begin{aligned} \frac{du(x(\tau),\tau)}{u(x(\tau_-),\tau)} &= \mu^u(x(\tau_-),\tau)d\tau + \sigma^u(x(\tau_-),\tau)dW(\tau) + J^u(\Delta x,\tau)\,dZ(\tau) \\ &= \left(\mu^u(x(\tau_-),\tau) + v(x(\tau_-))J^u(\Delta x,\tau)\right)d\tau + \sigma^u(x(\tau_-),\tau)dW(\tau) + J^u(\Delta x,\tau)\,dM(\tau), \end{aligned}$$

where $\mu^u = \left[\left(\frac{\partial}{\partial t} + L\right)u\right]/u$, $\sigma^u = \left(\frac{\partial u}{\partial x}\cdot\sigma\right)/u$, $\frac{\partial}{\partial t}+L$ is the generator for pure diffusion processes and, finally:

$$J^u(\Delta x,\tau) \equiv \frac{u(x(\tau),\tau) - u(x(\tau_-),\tau)}{u(x(\tau_-),\tau)}.$$

Next generalize the steps made some two subsections ago, and let

$$d\tilde{W} = dW + \lambda d\tau; \quad d\tilde{Z} = dZ - v^Q d\tau.$$

The objective is to find restrictions on both λ and v^Q such that both $\tilde{W}$ and $\tilde{Z}$ are Q-martingales. Below, we show that there is a precise connection between v^Q and J^η, where J^η is the jump component in the differential representation of η:

$$\frac{d\eta(\tau)}{\eta(\tau_-)} = -\lambda(x(\tau_-))dW(\tau) + J^\eta(\Delta x,\tau)\,dM(\tau), \quad \eta(t) = 1.$$

The relationship is

$$v^Q = v\left(1 + J^\eta\right),$$

and a proof of these facts will be provided below. What has to be noted here, is that in this case,

$$\frac{d\eta(\tau)}{\eta(\tau_-)} = -\lambda(x(\tau_-))dW(\tau) + \left(\lambda^J - 1\right)dM(\tau), \quad \eta(t) = 1,$$

which clearly generalizes what stated in the previous subsection.

Finally, we have:

$$\begin{aligned} \frac{du}{u} &= \left(\mu^u + vJ^u\right)d\tau + \sigma^u dW + J^u\left(dZ - vd\tau\right) \\ &= \left(\mu^u + v^Q J^u - \sigma^u\lambda\right)d\tau + \sigma^u d\tilde{W} + J^u d\tilde{Z} \\ &= \left(\mu^u + v\left(1+J^\eta\right)J^u - \sigma^u\lambda\right)d\tau + \sigma^u d\tilde{W} + J^u d\tilde{Z}. \end{aligned}$$

Finally, by the Q-martingale property of the discounted u,

$$\mu^u - r = \sigma^u\lambda - v^Q\cdot E_{\Delta x}\left(J^u\right) = \sigma^u\lambda - v\cdot E_{\Delta x}\left\{\left(1+J^\eta\right)J^u\right\},$$

where $E_{\Delta x}$ is taken with respect to the jump-size distribution, which is the same under Q and P.

Proof that $v^Q = v\left(1+J^\eta\right)$

As usual, the state-price density η has to be a P-martingale in order to be able to price bonds (in addition to all other assets). In addition, η clearly "depends" on W and Z. Therefore, it satisfies:

$$\frac{d\eta(\tau)}{\eta(\tau_-)} = -\lambda(x(\tau_-))dW(\tau) + J_\eta\left(\Delta x, \tau\right) dM(\tau), \quad \eta(t) = 1.$$

We wish to find v^Q in $d\tilde{Z} = dZ - v^Q d\tau$ such that $\tilde{Z}$ is a Q-martingale, viz

$$\tilde{Z}(\tau) = \mathbb{E}[\tilde{Z}(T)],$$

i.e.,

$$\mathbb{E}(\tilde{Z}(t)) = \frac{E\left(\eta(T) \cdot \tilde{Z}(T)\right)}{\eta(t)} = \tilde{Z}(t) \; \Leftrightarrow \; \eta(t)\tilde{Z}(t) = E[\eta(T)\tilde{Z}(T)],$$

i.e.,

$$\eta(t)\tilde{Z}(t) \text{ is a } P\text{-martingale.}$$

By Itô's lemma,

$$\begin{aligned} d(\eta\tilde{Z}) &= d\eta \cdot \tilde{Z} + \eta \cdot d\tilde{Z} + d\eta \cdot d\tilde{Z} \\ &= d\eta \cdot \tilde{Z} + \eta\left(dZ - v^Q d\tau\right) + d\eta \cdot d\tilde{Z} \\ &= d\eta \cdot \tilde{Z} + \eta[\underbrace{dZ - vd\tau}_{dM} + \left(v - v^Q\right) d\tau] + d\eta \cdot d\tilde{Z} \\ &= d\eta \cdot \tilde{Z} + \eta \cdot dM + \eta\left(v - v^Q\right) d\tau + d\eta \cdot d\tilde{Z}. \end{aligned}$$

Because η, M and $\eta\tilde{Z}$ are P-martingales,

$$\forall T, \;\; 0 = E\left\{\int_t^T \left[\eta(\tau) \cdot \left(v(\tau) - v^Q(\tau)\right) d\tau + \int_t^T d\eta(\tau) \cdot d\tilde{Z}(\tau)\right]\right\}.$$

But

$$d\eta \cdot d\tilde{Z} = \eta\left(-\lambda dW + J^\eta dM\right)\left(dZ - v^Q d\tau\right) = \eta\left[-\lambda dW + J^\eta\left(dZ - vd\tau\right)\right]\left(dZ - v^Q d\tau\right),$$

and since $(dZ)^2 = dZ$,

$$E(d\eta \cdot d\tilde{Z}) = \eta \cdot J^\eta v \cdot d\tau,$$

and the previous condition collapses to:

$$\forall T, \;\; 0 = E\left[\int_t^T \eta(\tau) \cdot \left(v(\tau) - v^Q(\tau) + J^\eta(\Delta x)v(\tau)\right) d\tau\right],$$

which implies

$$v^Q(\tau) = v(\tau)\left(1 + J^\eta(\Delta x)\right), \;\; \text{a.s.}$$

∥

Part II

Asset pricing and reality

5

On kernels and puzzles

5.1 A single factor model

This chapter discusses theoretical restrictions that can be used to perform statistical validation of asset pricing models. We reconsider the Lucas' model, and give more structure on the data generating process. We present a simple setting which allows us to obtain closed-form solutions. We then discuss how the model's predictions can be used to test the validity of the model.

5.2 A single factor model

5.2.1 The model

There is a representative agent with CRRA utility, viz $u(x) = x^{1-\eta}/(1-\eta)$. Cum-dividends gross returns $(q_t + D_t)/q_{t-1}$ are generated by:

$$\begin{cases} \log(q_t + D_t) & = & \log q_{t-1} + \mu_q - \frac{1}{2}\sigma_q^2 + \epsilon_{q,t} \\ \log D_t & = & \log D_{t-1} + \mu_D - \frac{1}{2}\sigma_D^2 + \epsilon_{D,t} \end{cases} \tag{5.1}$$

where

$$\begin{bmatrix} \epsilon_{q,t} \\ \epsilon_{D,t} \end{bmatrix} \sim NID\left(\mathbf{0}_2; \begin{bmatrix} \sigma_q^2 & \sigma_{qD} \\ \sigma_{qD} & \sigma_D^2 \end{bmatrix}\right).$$

Given the stochastic price and dividend process in (5.1), we now derive restrictions between the various coefficients μ_q, μ_D, σ_q^2, σ_D^2 and σ_{qD} that are imposed by economic theory. The cum dividend process in (5.1) has been assumed so for analytical purposes only.

By standard consumption-based asset pricing theory (see Part I),

$$q_t = E\left[\beta \frac{u'(D_{t+1})}{u'(D_t)}(q_{t+1} + D_{t+1}) \middle| \mathcal{F}_t\right],$$

with the usual notation. By the preferences assumption,

$$1 = E\left[e^{Z_{t+1}+Q_{t+1}} \middle| \mathcal{F}_t\right], \tag{5.2}$$

where

$$Z_{t+1} = \log\left(\beta\left(\frac{D_{t+1}}{D_t}\right)^{-\eta}\right); \qquad Q_{t+1} = \log\left(\frac{q_{t+1}+D_{t+1}}{q_t}\right).$$

In fact, eq. (5.2) holds for any asset. In particular, it holds for a one-period bond with price $q_t^b \equiv b_t$, $q_{t+1}^b \equiv 1$ and $D_{t+1}^b \equiv 0$. Define, $Q_{t+1}^b \equiv \log\left(b_t^{-1}\right) \equiv \log R_t$. By replacing this into eq. (5.2), one gets $R_t^{-1} = E\left[e^{Z_{t+1}}\middle|\mathcal{F}_t\right]$. We are left with the following system:

$$\begin{cases} \dfrac{1}{R_t} &= E\left[e^{Z_{t+1}}\middle|\mathcal{F}_t\right] \\ 1 &= E\left[e^{Z_{t+1}+Q_{t+1}}\middle|\mathcal{F}_t\right] \end{cases} \tag{5.3}$$

To obtain closed-form solutions, we will need to use the following result:

LEMMA 5.1: *Let Z be conditionally normally distributed. Then, for any $\gamma \in \mathbb{R}$,*

$$\begin{aligned} E\left[e^{-\gamma Z_{t+1}}\middle|\mathcal{F}_t\right] &= e^{-\gamma E(Z_{t+1}|\mathcal{F}_t)+\frac{1}{2}\gamma^2 var(Z_{t+1}|\mathcal{F}_t)} \\ \sqrt{var\left[e^{-\gamma Z_{t+1}}\middle|\mathcal{F}_t\right]} &= e^{-\gamma E(Z_{t+1}|\mathcal{F}_t)+\gamma^2 var(Z_{t+1}|\mathcal{F}_t)}\sqrt{1-e^{-\gamma^2 var(Z_{t+1}|\mathcal{F}_t)}} \end{aligned}$$

By the definition of Z, eq. (5.1), and lemma 5.1,

$$\frac{1}{R_t} = E\left[e^{Z_{t+1}}\middle|\mathcal{F}_t\right] = e^{E[Z_{t+1}|\mathcal{F}_t]+\frac{1}{2}var[Z_{t+1}|\mathcal{F}_t]} = e^{\log\beta-\eta\left(\mu_D-\frac{1}{2}\sigma_D^2\right)+\frac{1}{2}\eta^2\sigma_D^2}.$$

The equilibrium interest rate thus satisfies,

$$\log R_t = -\log\beta + \eta\mu_D - \frac{\eta(\eta+1)}{2}\sigma_D^2, \qquad \text{a constant.} \tag{5.4}$$

The $\eta\mu_D$ term reflects "intertemporal substitution" effects; the last term term reflects "precautionary" motives.

The second equation in (5.3) can be written as,

$$1 = E\left[\exp\left(Z_{t+1}+Q_{t+1}\right)\middle|\mathcal{F}_t\right] = e^{\log\beta-\eta\left(\mu_D-\frac{1}{2}\sigma_D^2\right)+\mu_q-\frac{1}{2}\sigma_q^2}\cdot E\left[e^{\tilde{n}_{t+1}}\middle|\mathcal{F}_t\right],$$

where $\tilde{n}_{t+1} \equiv \epsilon_{q,t+1} - \eta\epsilon_{D,t} \sim N(0, \sigma_q^2+\eta^2\sigma_D^2-2\eta\sigma_{qD})$. The above expectation can be computed through lemma 5.1. The result is,

$$0 = \underbrace{\log\beta - \eta\mu_D + \frac{\eta(\eta+1)}{2}\sigma_D^2}_{-\log R_t} + \mu_q - \eta\sigma_{qD}.$$

By defining $R_t \equiv e^{r_t}$, and rearranging terms,

$$\underbrace{\mu_q - r}_{\text{risk premium}} = \eta\sigma_{qD}.$$

To sum up,

$$\begin{cases} \mu_q &= r + \eta\sigma_{qD} \\ r_t &= -\log\beta + \eta\mu_D - \dfrac{\eta(\eta+1)}{2}\sigma_D^2 \end{cases}$$

Let us compute other interesting objects. The expected gross return on the risky asset is,

$$E\left[\left.\frac{q_{t+1}+D_{t+1}}{q_t}\right|\mathcal{F}_t\right]=e^{\mu_q-\frac{1}{2}\sigma_q^2}\cdot E\left[\left.e^{\epsilon_{q,t+1}}\right|\mathcal{F}_t\right]=e^{\mu_q}=e^{r+\eta\sigma_{qD}}.$$

Therefore, if $\sigma_{qD}>0$, then $E\left[\left.(q_{t+1}+D_{t+1})/\,q_t\right|\mathcal{F}_t\right]>E\left[\left.b_t^{-1}\right|\mathcal{F}_t\right]$, as expected.

Next, we test the internal consistency of the model. The coefficients of the model must satisfy some restrictions. In particular, the asset price volatility must be determined endogeneously. We first conjecture that the following "no-sunspots" condition holds,

$$\epsilon_{q,t}=\epsilon_{D,t}. \tag{5.5}$$

We will demonstrate below that this is indeed the case. Under the previous condition,

$$\mu_q=r+\lambda\sigma_D;\quad \lambda\equiv\eta\sigma_D,$$

and

$$Z_{t+1}=-\left(r+\frac{1}{2}\lambda^2\right)-\lambda u_{D,t+1};\qquad u_{D,t+1}\equiv\frac{\epsilon_{D,t+1}}{\sigma_D}.$$

Under condition (5.5), we have a very instructive way to write the pricing kernel. Precisely, define recursively,

$$m_{t+1}=\frac{\xi_{t+1}}{\xi_t}\equiv\exp\left(Z_{t+1}\right);\qquad \xi_0=1.$$

This is reminiscent of the continuous time representation of Arrow-Debreu state prices (see chapter 4).

Next, let's iterate the asset price equation (5.2),

$$\begin{aligned}q_t &= E\left[\left.\left(\prod_{j=1}^n e^{Z_{t+i}}\right)\cdot q_{t+n}\right|\mathcal{F}_t\right]+\sum_{i=1}^n E\left[\left.\left(\prod_{j=1}^i e^{Z_{t+j}}\right)\cdot D_{t+i}\right|\mathcal{F}_t\right]\\ &= E\left[\left.\frac{\xi_{t+n}}{\xi_t}\cdot q_{t+n}\right|\mathcal{F}_t\right]+\sum_{i=1}^n E\left[\left.\frac{\xi_{t+i}}{\xi_t}\cdot D_{t+i}\right|\mathcal{F}_t\right].\end{aligned}$$

By letting $n\to\infty$ and assuming no-bubbles, we get:

$$q_t=\sum_{i=1}^\infty E\left[\left.\frac{\xi_{t+i}}{\xi_t}\cdot D_{t+i}\right|\mathcal{F}_t\right]. \tag{5.6}$$

The expectation is, by lemma 5.1,

$$E\left[\left.\frac{\xi_{t+i}}{\xi_t}\cdot D_{t+i}\right|\mathcal{F}_t\right]=E\left[\left.e^{\sum_{j=1}^i Z_{t+j}}\cdot D_{t+i}\right|\mathcal{F}_t\right]=D_te^{(\mu_D-r-\sigma_D\lambda)i}.$$

Suppose that the "risk-adjusted" discount rate $r+\sigma_D\lambda$ is higher than the growth rate of the economy, viz.

$$r+\sigma_D\lambda>\mu_D\Leftrightarrow k\equiv e^{\mu_D-r-\sigma_D\lambda}<1.$$

Under this condition, the summation in eq. (5.6) converges, and we obtain:

$$\frac{q_t}{D_t}=\frac{k}{1-k}. \tag{5.7}$$

This is a version of the celebrated Gordon's formula. That is a boring formula because it predicts that price-dividend ratios are constant, a counterfactual feature (see Chapter 6). Chapter 6 reviews some developments addressing this issue.

To find the final restrictions of the model, notice that eq. (5.7) and the second equation in (5.1) imply that

$$\log(q_t + D_t) - \log q_{t-1} = -\log k + \mu_D - \frac{1}{2}\sigma_D^2 + \epsilon_{D,t}.$$

By the first equation in (5.1),

$$\begin{cases} \mu_q - \frac{1}{2}\sigma_q^2 & = & \mu_D - \frac{1}{2}\sigma_D^2 - \log k \\ \epsilon_{q,t} & = & \epsilon_{D,t}, \quad \forall t \end{cases}$$

The second condition confirms condition (5.5). It also reveals that, $\sigma_q^2 = \sigma_{qD} = \sigma_D^2$. By replacing this into the first condition, delivers back $\mu_q = \mu_D - \log k = r + \sigma_D \lambda$.

5.2.2 Extensions

In chapter 3 we showed that in a i.i.d. environment, prices are convex (resp. concave) in the dividend rate whenever $\eta > 1$ (resp. $\eta < 1$). The pricing formula (5.7) reveals that in a dynamic environment, such a property is lost. In this formula, prices are always linear in the dividends' rate. It would be possible to show with the techniques developed in the next chapter that in a dynamic context, convexity properties of the price function would be inherited by properties of the dividend process in the following sense: if the expected dividend growth under the risk-neutral measure is a convex (resp. concave) function of the initial dividend rate, then prices are convex (resp. concave) in the initial dividend rate. In the model analyzed here, the expected dividend growth under the risk-neutral measure is linear in the dividends' rate, and this explains the linear formula (5.7).

5.3 The equity premium puzzle

> "Average excess returns on the US stock market [the equity premium] is too high to be easily explained by standard asset pricing models." Mehra and Prescott

To be consistent with data, the equity premium,

$$\mu_q - r = \lambda \sigma_D, \qquad \lambda = \eta \sigma_D$$

must be "high" enough - as regards US data, approximately an annualized 6%. If the asset we are trying to price is literally a *consumption claim*, then σ_D is consumption volatility, which is very low (approximately 3%). To make $\mu_q - r$ high, one needs very "high" values of η (let's say $\eta \simeq 30$), But assuming $\eta = 30$ doesn't seem to be plausible. This is the equity premium puzzle originally raised by Mehra and Prescott (1985).

Even if we dismiss the idea that $\eta = 30$ is implausible, there is another puzzle, the interest rate puzzle. As we showed in eq. (5.4), very high values of η can make the interest rate very high (see figure 5.1).

In the next section, we show how this failure of the model can be "detected" with a general methodology that can be applied to a variety of related models - more general models.

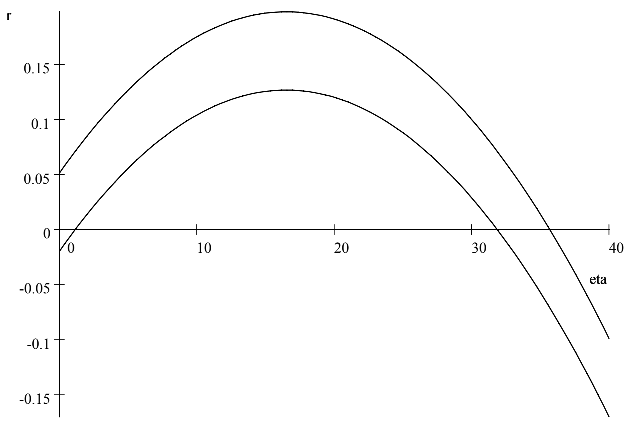


FIGURE 5.1. The risk-free rate puzzle: the two curves depict the graph $\eta \mapsto r(\eta) = -\log\beta + 0.0183\cdot\eta - (0.0328)^2 \cdot \frac{\eta(\eta+1)}{2}$, with $\beta = 0.95$ (top curve) and $\beta = 1.05$ (bottom curve). Even if we accept the idea that risk aversion is as high as $\eta = 30$, we would obtain a resulting equilibrium interest rate as high as 10%. The only way to make low r consistent with high values of η is to make $\beta > 1$.

5.4 The Hansen-Jagannathan cup

Suppose there are n risky assets. The n asset pricing equations for these assets are,

$$1 = E\left[m_{t+1}\left(1+R_{j,t+1}\right)\middle|\,\mathcal{F}_t\right],\quad j=1,\cdots,n.$$

By taking the unconditional expectation of the previous equation, and defining $R_t = (R_{1,t},\cdots,R_{n,t})^\top$,

$$\mathbf{1}_n = E\left[m_t\left(\mathbf{1}_n + R_t\right)\right].$$

Let $\bar{m} \equiv E(m_t)$. We create a family of stochastic discount factors m_t^* parametrized by $\bar{m}$ by projecting m on to the asset returns,

$$Proj\left(m\middle|\,\mathbf{1}_n + R_t\right) \equiv m_t^*(\bar{m}) = \bar{m} + \underset{1\times n}{\left[R_t - E(R_t)\right]^\top}\underset{n\times 1}{\beta_{\bar{m}}},$$

where[1]

$$\underset{n\times 1}{\beta_{\bar{m}}} = \underset{n\times n}{\Sigma^{-1}}\underset{n\times 1}{cov\left(m,\mathbf{1}_n + R_t\right)} = \Sigma^{-1}\left[\mathbf{1}_n - \bar{m}E\left(\mathbf{1}_n + R_t\right)\right],$$

and $\Sigma \equiv E\left[\left(R_t - E(R_t)\right)\left(R_t - E(R_t)\right)^\top\right]$. As shown in the appendix, we also have that,

$$\mathbf{1}_n = E\left[m_t^*(\bar{m})\cdot\left(\mathbf{1}_n + R_t\right)\right].$$

We have,

$$\sqrt{var\left(m_t^*(\bar{m})\right)} = \sqrt{\beta_{\bar{m}}^\top\Sigma\beta_{\bar{m}}} = \sqrt{\left(\mathbf{1}_n - \bar{m}E\left(\mathbf{1}_n + R_t\right)\right)^\top\Sigma^{-1}\left(\mathbf{1}_n - \bar{m}E\left\{\mathbf{1}_n + R_t\right\}\right)}.$$

This is the celebrated *Hansen-Jagannathan "cup"* (Hansen and Jagannathan (1991)). The interest of this object lies in the following theorem.

[1] We have, $cov\left(m,\mathbf{1}_n + R_t\right) = E\left[m\left(\mathbf{1}_n + R\right)\right] - E(m)\,E\left(\mathbf{1}_n + R_t\right) = \mathbf{1}_n - \bar{m}E\left(\mathbf{1}_n + R_t\right)$.

THEOREM 5.1: *Among all stochastic discount factors with fixed expectation $\bar{m}$, $m_t^*(\bar{m})$ is the one with the smallest variance.*

PROOF: Consider another discount factor indexed by $\bar{m}$, i.e. $m_t(\bar{m})$. Naturally, $m_t(\bar{m})$ satisfies $\mathbf{1}_n = E\left[m_t(\bar{m})\left(\mathbf{1}_n + R_t\right)\right]$. And since it also holds that $\mathbf{1}_n = E\left[m_t^*(\bar{m})\left(\mathbf{1}_n + R_t\right)\right]$, we deduce that

$$\begin{aligned}
\mathbf{0}_n &= E\left[\left(m_t(\bar{m}) - m_t^*(\bar{m})\right)\left(\mathbf{1}_n + R_t\right)\right] \\
&= E\left\{\left[m_t(\bar{m}) - m_t^*(\bar{m})\right]\left[\left(\mathbf{1}_n + E(R_t)\right) + \left(R_t - E(R_t)\right)\right]\right\} \\
&= E\left\{\left[m_t(\bar{m}) - m_t^*(\bar{m})\right]\left[R_t - E(R_t)\right]\right\} \\
&= cov\left[m_t(\bar{m}) - m_t^*(\bar{m}), R_t\right]
\end{aligned}$$

where the third line follows from the fact that $E\left[m_t(\bar{m})\right] = E\left[m_t^*(\bar{m})\right] = \bar{m}$, and the fourth line follows because $E\left[\left(m_t(\bar{m}) - m_t^*(\bar{m})\right)\right] = 0$. But $m_t^*(\bar{m})$ is a linear combination of R_t. By the previous equation, it must then be the case that,

$$0 = cov\left[m_t(\bar{m}) - m_t^*(\bar{m}), m_t^*(\bar{m})\right].$$

Hence,

$$\begin{aligned}
var\left[m_t(\bar{m})\right] &= var\left[m_t^*(\bar{m}) + m_t(\bar{m}) - m_t^*(\bar{m})\right] \\
&= var\left[m_t^*(\bar{m})\right] + var\left[m_t(\bar{m}) - m_t^*(\bar{m})\right] + 2 \cdot cov\left[m_t(\bar{m}) - m_t^*(\bar{m}), m_t^*(\bar{m})\right] \\
&= var\left[m_t^*(\bar{m})\right] + var\left[m_t(\bar{m}) - m_t^*(\bar{m})\right] \\
&\geq var\left[m_t^*(\bar{m})\right].
\end{aligned}$$

‖

The previous bound can be improved by using conditioning information as in Gallant, Hansen and Tauchen (1990) and the relatively more recent work by Ferson and Siegel (2002). Moreover, these bounds typically diplay a finite sample bias: they typically overstate the true bounds and thus they reject too often a given model. Finite sample corrections are considered by Ferson and Siegel (2002).

For example, let us consider an application of the Hansen-Jagannathan testing methodology to the model in section 5.1. That model has the following stochastic discount factor,

$$m_{t+1} = \frac{\xi_{t+1}}{\xi_t} = \exp\left(Z_{t+1}\right); \qquad Z_{t+1} = -\left(r + \frac{1}{2}\lambda^2\right) - \lambda u_{D,t+1}; \qquad u_{D,t+1} \equiv \frac{\epsilon_{D,t+1}}{\sigma_D}.$$

First, we have to compute the first two moments of the stochastic discount factor. By lemma 5.1 we have,

$$\bar{m} = E(m_t) = e^{-r} \quad \text{and} \quad \bar{\sigma}_m = \sqrt{var\left(m_t(\bar{m})\right)} = e^{-r+\frac{1}{2}\lambda^2}\sqrt{1 - e^{-\lambda^2}} \tag{5.8}$$

where

$$r = -\log\beta + \eta\mu_D - \frac{\eta\left(\eta + 1\right)}{2}\sigma_D^2 \quad \text{and} \quad \lambda = \eta\sigma_D.$$

For given μ_D and σ_D^2, system (5.8) forms a η-parametrized curve in the space ($\bar{m}$-$\bar{\sigma}_m$). The objective is to see whether there are plausible values of η for which such a η-parametrized curve enters the Hansen-Jagannathan cup. Typically, this is not the case. Rather, one has the situation depicted in Figure 5.2 below.

The general message is that models can be consistent with data with high volatile pricing kernels (for a fixed $\bar{m}$). Dismiss the idea of a representative agent with CRRA utility function. Consider instead models with heterogeneous agents (by generalizing some ideas in Constantinides and Duffie (1996); and/or consider models with more realistic preferences - such as for example the habit preferences considered in Campbell and Cochrane (1999, *J. Pol. Econ.*); and/or combinations of these. These things will be analyzed in depth in the next chapter.

5.5 Simple multidimensional extensions

A natural way to increase the variance of the pricing kernel is to increase the number of factors. We consider two possibilities: one in which returns are normally distributed, and one in which returns are lognormally distributed.

5.5.1 *Exponential affine pricing kernels*

Consider again the simple model in section 5.1. In this section, we shall make a different assumption regarding the returns distributions. But we shall maintain the hypothesis that the pricing kernel satisfies an exponential-Gaussian type structure,

$$m_{t+1} = \exp\left(Z_{t+1}\right); \qquad Z_{t+1} = -\left(r + \frac{1}{2}\lambda^2\right) - \lambda u_{D,t+1}; \qquad u_{D,t+1} \sim \text{NID}\left(0,1\right),$$

where r and λ are some constants. We have,

$$1 = E(m_{t+1} \cdot \tilde{R}_{t+1}) = E\left(m_{t+1}\right) E(\tilde{R}_{t+1}) + cov(m_{t+1}, \tilde{R}_{t+1}), \qquad \tilde{R}_{t+1} \equiv \frac{q_{t+1} + D_{t+1}}{q_t}.$$

By rearranging terms,[2] and using the fact that $E\left(m_{t+1}\right) = R^{-1}$,

$$E(\tilde{R}_{t+1}) - R = -R \cdot cov(m_{t+1}, \tilde{R}_{t+1}). \tag{5.9}$$

The following result is useful:

LEMMA 5.2 (Stein's lemma): *Suppose that two random variables x and y are jointly normal. Then,*

$$cov\left[g\left(x\right), y\right] = E\left[g'\left(x\right)\right] \cdot cov\left(x, y\right),$$

for any function $g : E\left(\left|g'\left(x\right)\right|\right) < \infty$.

We now suppose that $\tilde{R}$ is normally distributed. This assumption is inconsistent with the model in section 5.1. In the model of section 5.1, $\tilde{R}$ is *log*normally distributed in equilibrium

[2] With a portfolio return that is perfectly correlated with m, we have:

$$E_t(\tilde{R}^M_{t+1}) - \frac{1}{E_t\left(m_{t+1}\right)} = -\frac{\sigma_t\left(m_{t+1}\right)}{E_t\left(m_{t+1}\right)} \sigma_t(\tilde{R}^M_{t+1}).$$

In more general setups than the ones considered in this introductory example, both $\frac{\sigma_t(m_{t+1})}{E_t(m_{t+1})}$ and $\sigma_t(\tilde{R}^M_{t+1})$ should be time-varying.

because $\log \tilde{R} = \mu_D - \frac{1}{2}\sigma_D^2 + \epsilon_q$, with ϵ_q normal. But let's explore the asset pricing implications of this tilting assumption. Because $\tilde{R}_{t+1}$ and Z_{t+1} are normal, and $m_{t+1} = m(Z_{t+1}) = \exp(Z_{t+1})$, we may apply Lemma 5.2 and obtain,

$$cov(m_{t+1}, \tilde{R}_{t+1}) = E[m'(Z_{t+1})] \cdot cov(Z_{t+1}, \tilde{R}_{t+1}) = -\lambda R^{-1} \cdot cov(u_{D,t+1}, \tilde{R}_{t+1}).$$

Replacing this into eq. (5.9),

$$E(\tilde{R}_{t+1}) - R = \lambda \cdot cov(u_{D,t+1}, \tilde{R}_{t+1}).$$

We wish to extend the previous observations to more general situations. Clearly, the pricing kernel is some function of K factors $m(\epsilon_{1t}, \cdots, \epsilon_{Kt})$. A particularly convenient analytical assumption is to make m exponential-affine and the factors $(\epsilon_{i,t})_{i=1}^K$ normal, as in the following definition:

Definition 5.1 (EAPK: Exponential Affine Pricing Kernel): *Let,*

$$Z_t \equiv \phi_0 + \sum_{i=1}^K \phi_i \epsilon_{i,t}.$$

A EAPK *is a function*

$$m_t = m(Z_t) = \exp(Z_t).$$

If $(\epsilon_{i,t})_{i=1}^K$ are jointly normal, and each $\epsilon_{i,t}$ has mean zero and variance σ_i^2, $i = 1, \cdots, K$, the EAPK is called a Normal EAPK (NEAPK).

In the previous definition, we assumed that each $\epsilon_{i,t}$ has mean zero. This entails no loss of generality insofar as $\phi_0 \neq 0$.

Now suppose that $\tilde{R}$ is normally distributed. By Lemma 5.2 and the NEAPK structure,

$$cov(m_{t+1}, \tilde{R}_{t+1}) = cov[\exp(Z_{t+1}), \tilde{R}_{t+1}] = R^{-1} cov(Z_{t+1}, \tilde{R}_{t+1}) = R^{-1} \sum_{i=1}^K \phi_i cov(\epsilon_{i,t+1}, \tilde{R}_{t+1}).$$

By replacing this into eq. (5.9) leaves the linear factor representation,

$$E(\tilde{R}_{t+1}) - R = -\sum_{i=1}^K \phi_i \underbrace{cov(\epsilon_{i,t+1}, \tilde{R}_{t+1})}_{\text{“betas”}}. \tag{5.10}$$

We have thus shown the following result:

Proposition 5.1: *Suppose that $\tilde{R}$ is normally distributed. Then, NEAPK $\Rightarrow$ linear factor representation for asset returns.*

The APT representation in eq. (5.10), is close to one result in Cochrane (1996).[3] Cochrane (1996) assumed that m has a linear structure, i.e. $m(Z_t) = Z_t$ where Z_t is as in Definition 5.1.

[3] To recall why eq. (5.10) is indeed a APT equation, suppose that $\tilde{R}$ is a n-(column) vector of returns and that $\tilde{R} = a + bf$, where f is K-(column) vector *with zero mean and unit variance* and a, b are some given vector and matrix with appropriate dimension. Then clearly, $b = cov(\tilde{R}, f)$. A portfolio π delivers $\pi^\top \tilde{R} = \pi^\top a + \pi^\top cov(\tilde{R}, f) f$. Arbitrage opportunity is: $\exists \pi : \pi^\top cov(\tilde{R}, f) = 0$ and $\pi^\top a \neq r$. To rule that out, we may show as in Part I of these *Lectures* that there must exist a K-(column) vector λ s.t. $a = cov(\tilde{R}, f)\lambda + r$. This implies $\tilde{R} = a + bf = r + cov(\tilde{R}, f)\lambda + bf$. That is, $E(\tilde{R}) = r + cov(\tilde{R}, f)\lambda$.

This assumption implies that $cov(m_{t+1}, \tilde{R}_{t+1}) = \sum_{i=1}^{K} \phi_i cov(\epsilon_{i,t+1}, \tilde{R}_{t+1})$. By replacing this into eq. (5.9),

$$E(\tilde{R}_{t+1}) - R = -R \sum_{i=1}^{K} \phi_i cov(\epsilon_{i,t+1}, \tilde{R}_{t+1}), \quad \text{where } R = \frac{1}{E(m)} = \frac{1}{\phi_0}.$$

The advantage to use the NEAPKs is that the pricing kernel is automatically guaranteed to be strictly positive - a condition needed to rule out arbitrage opportunities.

5.5.2 Lognormal returns

Next, we assume that $\tilde{R}$ is *log*normally distributed, and that NEAPK holds. We have,

$$1 = E\left(m_{t+1} \cdot \tilde{R}_{t+1}\right) \iff e^{-\phi_0} = E\left[e^{\sum_{i=1}^{K} \phi_i \epsilon_{i,t+1}} \cdot \tilde{R}_{t+1}\right]. \tag{5.11}$$

Consider first the case $K = 1$ and let $y_t = \log \tilde{R}_t$ be normally distributed. The previous equation can be written as,

$$e^{-\phi_0} = E\left[e^{\phi_1 \epsilon_{t+1} + y_{t+1}}\right] = e^{E(y_{t+1}) + \frac{1}{2}(\phi_1^2 \sigma_\epsilon^2 + \sigma_y^2 + 2\phi_1 \sigma_{\epsilon y})}.$$

This is,

$$E(y_{t+1}) = -\left[\phi_0 + \frac{1}{2}(\phi_1^2 \sigma_\epsilon^2 + \sigma_y^2 + 2\phi_1 \sigma_{\epsilon y})\right].$$

By applying the pricing equation (5.11) to a bond price,

$$e^{-\phi_0} = E\left(e^{\phi_1 \epsilon_{t+1}}\right) e^{\log R_{t+1}} = e^{\log R_{t+1} + \frac{1}{2}\phi_1^2 \sigma_\epsilon^2},$$

and then

$$\log R_{t+1} = -\left(\phi_0 + \frac{1}{2}\phi_1^2 \sigma_\epsilon^2\right).$$

The expected excess return is,

$$E(y_{t+1}) - \log R_{t+1} + \frac{1}{2}\sigma_y^2 = -\phi_1 \sigma_{\epsilon y}.$$

This equation reveals how to derive the simple theory in section 5.1 in an alternate way. Apart from Jensen's inequality effects ($\frac{1}{2}\sigma_y^2$), this is indeed the Lucas model of section 5.1 once $\phi_1 = -\eta$. As is clear, this is a poor model because we are contrived to explain returns with only one "stochastic discount-factor parameter" (i.e. with ϕ_1).

Next consider the general case. Assume as usual that dividends are as in (5.1). To find the price function in terms of the state variable ϵ, we may proceed as in section 5.1. In the absence of bubbles,

$$q_t = \sum_{i=1}^{\infty} E\left[\frac{\xi_{t+i}}{\xi_t} \cdot D_{t+i}\right] = D_t \cdot \sum_{i=1}^{\infty} e^{\left(\mu_D + \phi_0 + \frac{1}{2}\sum_{i=1}^{K} \phi_i \left(\phi_i \sigma_i^2 + 2\sigma_{i,D}\right)\right) \cdot i}, \quad \sigma_{i,D} \equiv cov(\epsilon_i, \epsilon_D).$$

Thus, if

$$\hat{k} \equiv \mu_D + \phi_0 + \frac{1}{2}\sum_{i=1}^{K} \phi_i \left(\phi_i \sigma_i^2 + 2\sigma_{i,D}\right) < 0,$$

then,

$$\frac{q_t}{D_t} = \frac{\hat{k}}{1-\hat{k}}.$$

Even in this multi-factor setting, price-dividend ratios are constant - which is counterfactual. Note that the various parameters can be calibrated so as to make the pricing kernel satisfy the Hansen-Jagannathan theoretical test conditions in section 5.3. But the resulting model always makes the boring prediction that price-dividend ratios are constant. This multifactor model doesn't work even if the variance of the implied pricing kernel is high - and lies inside the Hansen-Jagannathan cup. Living inside the cup doesn't necessarily imply that the resulting model is a good one. We need other theoretical test conditions. The next chapter develops such theoretical test conditions (When are price-dividend ratios procyclical? When is returns volatility countercyclical? Etc.).

5.6 Pricing kernels, Sharpe ratios and the market portfolio

5.6.1 *What does a market portfolio do?*

When is the market portfolio perfectly correlated with the pricing kernel? When this is the case, we say that the market portfolio is *β-CAPM generating*. The answer to the previous question is generically negative. This section aims at clarifying the issue.[4]

Let $r_{i,t+1}^e = \tilde{R}_{i,t+1} - R_{t+1}$ be the excess return on a risky asset. By standard arguments,

$$0 = E_t\left(m_{t+1} r_{i,t+1}^e\right) = E_t\left(m_{t+1}\right) E_t\left(r_{i,t+1}^e\right) + \rho_{i,t} \cdot \sqrt{Var_t\left(m_{t+1}\right)} \cdot \sqrt{Var_t\left(r_{i,t+1}^e\right)},$$

where $\rho_{i,t} = corr_t(m_{t+1}, r_{i,t+1}^e)$. Hence,

$$\text{Sharpe\_Ratio} \equiv \frac{E_t\left(r_{i,t+1}^e\right)}{\sqrt{Var_t\left(r_{i,t+1}^e\right)}} = -\rho_{i,t} \cdot \frac{\sqrt{Var_t\left(m_{t+1}\right)}}{E_t\left(m_{t+1}\right)}.$$

Sharpe ratios computed on Fama-French data take on an average value of 0.45.

We have,

$$\left|\frac{E_t\left(r_{i,t+1}^e\right)}{\sqrt{Var_t\left(r_{i,t+1}^e\right)}}\right| \leq \frac{\sqrt{Var_t\left(m_{t+1}\right)}}{E_t\left(m_{t+1}\right)} = \sqrt{Var_t\left(m_{t+1}\right)} \cdot R_{t+1}.$$

The highest possible Sharpe ratio is bounded. The equality is obtained with returns R_{t+1}^M (say) that are perfectly conditionally negatively correlated with the price kernel, i.e. $\rho_{M,t} = -1$. This portfolio is a *β-CAPM generating portfolio*. Should it be ever possible to create such a portfolio, we could also call it *market portfolio*. The reason is that a feasible and attainable portfolio lying on the kernel volatility bounds is clearly mean-variance efficient.

If $\rho_{M,t} = -1$,

$$\frac{E_t\left(r_{M,t+1}^e\right)}{\sqrt{Var_t\left(r_{M,t+1}^e\right)}} = \text{slope\_of\_the\_capital\_market\_line} = \frac{\sqrt{Var_t\left(m_{t+1}\right)}}{E_t\left(m_{t+1}\right)}.$$

[4] An additional article to read is Cecchetti, Lam, and Mark (1994).

But as we made clear in chapter 2, the Sharpe ratio $E_t(r^e_{M,t+1})\big/\sqrt{Var_t(r^e_{M,t+1})}$ has also the interpretation of unit market risk-premium. Hence,

$$\Pi \equiv \text{unit\_market\_risk\_premium} = \frac{\sqrt{Var_t\left(m_{t+1}\right)}}{E_t\left(m_{t+1}\right)}.$$

For example, the Lucas model in section 5.1 has,

$$\frac{\sqrt{Var_t\left(m_{t+1}\right)}}{E_t\left(m_{t+1}\right)} = \sqrt{e^{\eta^2\sigma_D^2}-1} \approx \eta\sigma_D.$$

In section 5.1, we also obtained that $\left(\mu_q - r\right)\big/\sigma_D = \eta\sigma_D$. As the previous relation reveals, Π is only approximately equal to $\eta\sigma_D$ because the asset in section 5.1 is simply *not* a β-CAPM generating portfolio. For example, suppose that the economy in section 5.1 has only a single risky asset. It would then be very natural to refer this asset to as "market portfolio". Yet this asset wouldn't be β-CAPM generating.

In section 5.1, we found that $E(\tilde{R}) = e^{\mu_q}$, $R = e^{-\log\beta+\eta\left(\mu_D-\frac{1}{2}\sigma_D^2\right)-\frac{1}{2}\eta^2\sigma_D^2}$ and $var(\tilde{R}) = e^{2\mu_q}(e^{\sigma_D^2}-1)$. Therefore, $\mathcal{S} \equiv E(\tilde{R}-R)\Big/\sqrt{var(\tilde{R})}$ is:

$$\mathcal{S} \equiv \text{Sharpe\_Ratio} = \frac{1-e^{-\eta\sigma_D^2}}{\sqrt{e^{\sigma_D^2}-1}}.$$

Indeed, by simple computations,

$$\rho = -\frac{1-e^{-\eta\sigma_D^2}}{\sqrt{e^{\eta^2\sigma_D^2}-1}\sqrt{e^{\sigma_D^2}-1}}.$$

This is not precisely "minus one". Yet in practice $\rho \approx -1$ when σ_D is low. However, consumption claims are not acting as market portfolios - in the sense of chapter 2. If that consumption claim is very highly correlated with the pricing kernel, then it is also a good approximation to the β-CAPM generating portfolio. But as the previous simple example demonstrates, that is only an approximation. To summarize, *the fact that everyone is using an asset* (*or in general a portfolio in a 2-funds separation context*) *doesn't imply that the resulting return is perfectly correlated with the pricing kernel*. In other terms, a market portfolio is not necessarily β-CAPM generating.[5]

We now describe a further complication: a β-CAPM generating portfolio is not necessarily the tangency portfolio. We show the existence of another portfolio producing the same β-pricing relationship as the tangency portfolio. For reasons developed below, such a portfolio is usually referred to as the *maximum correlation portfolio*.

Let $\bar{R} = \frac{1}{E(m)}$. By the CCAPM (see chapter 3),

$$E\left(R^i\right) - \bar{R} = \frac{\beta_{R^i,m}}{\beta_{R_p,m}}\left(E\left(R_p\right) - \bar{R}\right),$$

where R_p is a portfolio return. Next, let

$$R_p = R^m \equiv \frac{m}{E(m^2)}.$$

[5] As is well-known, things are the same in economies with one agent with quadratic utility. This fact can be seen at work in the previous formulae (just take $\eta = -1$). You should also be able to show this claim with more general quadratic utility functions - as in chapter 3.

This is clearly perfectly correlated with the kernel, and by the analysis in chapter 3,

$$E\left(R^i\right) - \bar{R} = \beta_{R^i,R^m}\left[E\left(R^m\right) - \bar{R}\right].$$

This is not yet the β-representation of the CAPM, because we have yet to show that there is a way to construct R^m as a portfolio return. In fact, there is a natural choice: pick $m = m^*$, where m^* is the minimum-variance kernel leading to the Hansen-Jagannathan bounds. Since m^* is linear in all asset retuns, R^{m^*} can be thought of as a return that can be obtained by investing in all assets. Furthermore, in the appendix we show that R^{m^*} satisfies,

$$1 = E\left(m \cdot R^{m^*}\right).$$

Where is this portfolio located? As shown in the appendix, *there is no portfolio yielding the same expected return with lower variance* (i.e., R^{m^*} is mean-variance efficient). In addition, in the appendix we show that,

$$E\left(R^{m^*}\right) - 1 = \frac{r - Sh}{1 + Sh} = r - \frac{1+r}{1+Sh}Sh < r.$$

Mean-variance efficiency of R^{m^*} and the previous inequality imply that this portfolio lies in the *lower branch* of the mean-variance efficient portfolios. And this is so because this portfolio is *positively* correlated with the *true* pricing kernel. Naturally, the fact that this portfolio is β-CAPM generating doesn't necessarily imply that it is also perfectly correlated with the true pricing kernel. As shown in the appendix, R^{m^*} has only the maximum possible correlation with all possible m. Perfect correlation occurs exactly in correspondence of the pricing kernel $m = m^*$ (i.e. when the economy exhibits a pricing kernel exactly equal to m^*).

Proof that R^{m^*} is β-capm generating. The relations $1 = E(m^*R^i)$ and $1 = E(m^*R^{m^*})$ imply

$$\begin{aligned} E(R^i) - R &= -R \cdot cov\left(m^*, R^i\right) \\ E(R^{m^*}) - R &= -R \cdot cov\left(m^*, R^{m^*}\right) \end{aligned}$$

and,

$$\frac{E(R^i) - R}{E(R^{m^*}) - R} = \frac{cov\left(m^*, R^i\right)}{cov\left(m^*, R^{m^*}\right)}.$$

By construction, R^{m^*} is perfectly correlated with m^*. Precisely, $R^{m^*} = m^*/\,E(m^{*2}) \equiv \gamma^{-1}m^*$, $\gamma \equiv E(m^{*2})$. Therefore,

$$\frac{cov\left(m^*, R^i\right)}{cov\left(m^*, R^{m^*}\right)} = \frac{cov\left(\gamma R^{m^*}, R^i\right)}{cov\left(\gamma R^{m^*}, R^{m^*}\right)} = \frac{\gamma \cdot cov\left(R^{m^*}, R^i\right)}{\gamma \cdot var\left(R^{m^*}\right)} = \beta_{R^i,R^{m^*}}.$$

$\|$

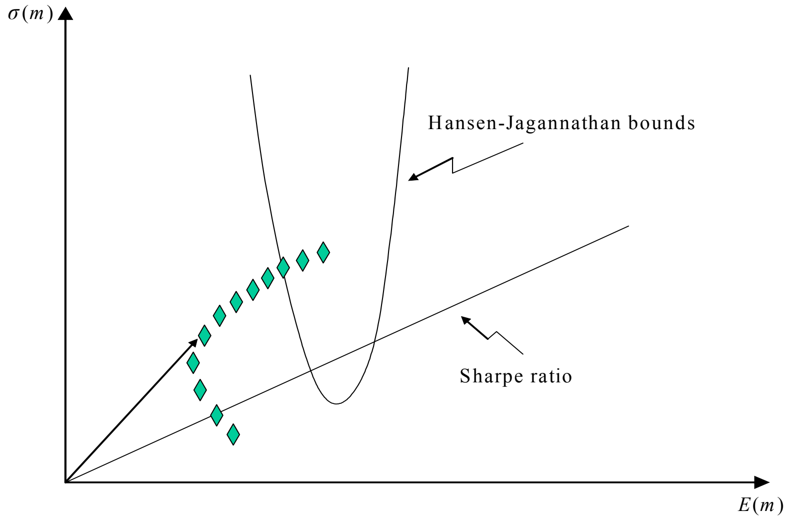
5.6.2 Final thoughts on the pricing kernel bounds

Figure 5.2 summarizes the typical situation that arises in practice. Points ♦ are the ones generated by the Lucas model in correspondence of different η. The model has to be such that points ♦ lie *above* the observed Sharpe ratio ($\sigma(m)/\,E(m) \geq$ greatest Sharpe ratio ever observed in the data—Sharpe ratio on the market portfolio) *and* inside the Hansen-Jagannathan bounds. Typically, very high values of η are required to enter the Hansen-Jagannathan bounds.

There is a beautiful connection between these things and the familiar mean-variance portfolio frontier described in chapter 2. As shown in figure 5.3, every asset or portfolio must lie inside the wedged region bounded by two straight lines with slopes $\mp\,\sigma(m)/\,E(m)$. This is so because, for any asset (or portfolio) that is priced with a kernel m, we have that

$$\left|E(R^i) - R\right| \leq \frac{\sigma(m)}{E(m)} \cdot \sigma\left(R^i\right).$$

As seen in the previous section, the equality is only achieved by asset (or portfolio) returns that are perfectly correlated with m. The point here is that a tangency portfolio such as T doesn't necessarily attain the kernel volatility bounds. Also, there is no reason for a market portfolio to lie on the kernel volatility bound. In the simple Lucas-Breeden economy considered in the previous section, for example, the (only existing) asset has a Sharpe ratio that doesn't lie on



the kernel volatility bounds. *In a sense, the CCAPM doesn't necessarily imply the CAPM*, i.e. there is no necessarily an asset acting at the same time as a market portfolio and β-CAPM generating *that is also priced consistently with the true kernel of the economy*. These conditions simultaneously hold if the (candidate) market portfolio is perfectly negatively correlated with the true kernel of the economy, but this is very particular (it is in this sense that one may say that the CAPM is a particular case of the CCAPM). A good research question is to find conditions on families of kernels consistent with the previous considerations.

On the other hand, we know that there exists another portfolio, the maximum correlation portfolio, that is also β-CAPM generating. In other terms, if $\exists R_* : R_* = -\gamma m$, for some positive constant γ, then the β-CAPM representation holds, but this doesn't necessarily mean that R_* is also a market portfolio. More generally, if there is a return R_* that is β-CAPM generating, then

$$\rho_{i,R^*} = \frac{\rho_{i,m}}{\rho_{R^*,m}}, \text{ all } i. \tag{5.12}$$

Therefore, we don't need an asset or portfolio return that is *perfectly* correlated with m to make the CCAPM shrink to the CAPM. In other terms, the existence of an asset return that is perfectly negatively correlated with the price kernel is a sufficient condition for the CCAPM to shrink to the CAPM, not a necessary condition. The proof of eq. (5.12) is easy. By the CCAPM,

$$E(R^i) - R = -\rho_{i,m}\frac{\sigma(m)}{E(m)}\sigma(R^i); \quad \text{and} \quad E(R^*) - R = -\rho_{R^*,m}\frac{\sigma(m)}{E(m)}\sigma(R^*).$$

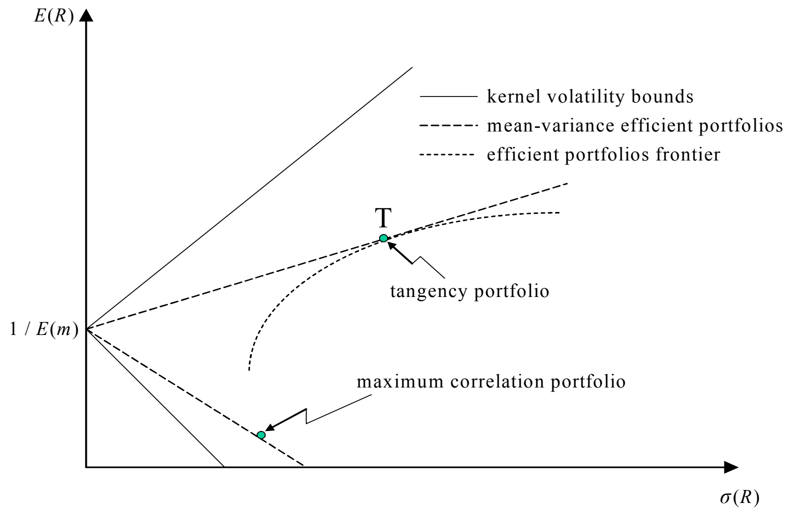
That is,

$$\frac{E(R^i) - R}{E(R^*) - R} = \frac{\rho_{i,m}}{\rho_{R^*,m}}\frac{\sigma(R^i)}{\sigma(R^*)} \tag{5.13}$$

But if R^* is β-CAPM generating,

$$\frac{E(R^i) - R}{E(R^*) - R} = \frac{cov(R^i, R^*)}{\sigma(R^*)^2} = \rho_{i,R^*}\frac{\sigma(R^i)}{\sigma(R^*)}. \tag{5.14}$$

Comparing eq. (5.13) with eq. (5.14) produces (5.12).



A final thought. Many recent applied research papers have important result but also a surprising motivation. They often state that because we observe time-varying Sharpe ratios on (proxies of) the market portfolio, one should also model the market risk-premium $\sqrt{Var_t(m_{t+1})}\Big/ E_t(m_{t+1})$ as time-varying. However, this is not rigorous motivation. The Sharpe ratio of the market portfolio is generally less than $\sqrt{Var_t(m_{t+1})}\Big/ E_t(m_{t+1})$. $\sqrt{Var_t(m_{t+1})}\Big/ E_t(m_{t+1})$ is only a bound. On a strictly theoretical point of view, $\sqrt{Var_t(m_{t+1})}\Big/ E_t(m_{t+1})$ time-varying is not a necessary nor a sufficient condition to observe time-varying Sharpe ratios. Figure 5.3 illustrates this point.

5.6.3 The Roll's critique

In applications and tests of the CAPM, proxies of the market portfolio such as the S&P 500 are used. However, the market portfolio is unobservable, and this prompted Roll (1977) to point out that the CAPM is inherently untestable. The argument is that a tangency portfolio (or in general, a β-CAPM generating portfolio) always exists. So, even if the CAPM is wrong, the proxy of the market portfolio will incorrectly support the model if such a proxy is more or less the same as the tangency portfolio. On the other hand, if the proxy is not mean-variance efficient, the CAPM can be rejected even if the CAPM is wrong. All in all, any test of the CAPM is a joint test of the model itself and of the closeness of the proxy to the market portfolio.

5.7 Appendix

Proof of the Equation, $\mathbf{1}_n = E\left[m_t^*(\bar{m}) \cdot (\mathbf{1}_n + R_t)\right]$. We have,

$$\begin{aligned}
E\left[m_t^*(\bar{m}) \cdot (\mathbf{1}_n + R_t)\right] &= E\left[\left(\bar{m} + (R_t - E(R_t))^\top \beta_{\bar{m}}\right)(\mathbf{1}_n + R_t)\right] \\
&= \bar{m}E(\mathbf{1}_n + R_t) + E\left[(R_t - E(R_t))^\top \beta_{\bar{m}} (\mathbf{1}_n + R_t)\right] \\
&= \bar{m}E(\mathbf{1}_n + R_t) + E\left[(\mathbf{1}_n + R_t)(R_t - E(R_t))^\top\right] \beta_{\bar{m}} \\
&= \bar{m}E(\mathbf{1}_n + R_t) + E\left[((\mathbf{1}_n + E(R_t)) + (R_t - E(R_t)))(R_t - E(R_t))^\top\right] \beta_{\bar{m}} \\
&= \bar{m}E(\mathbf{1}_n + R_t) + E\left[(R_t - E(R_t))(R_t - E(R_t))^\top\right] \beta_{\bar{m}} \\
&= \bar{m}E(\mathbf{1}_n + R_t) + \Sigma\beta_{\bar{m}} \\
&= \bar{m}E(\mathbf{1}_n + R_t) + \mathbf{1}_n - \bar{m}E(\mathbf{1}_n + R_t),
\end{aligned}$$

where the last line follows by the definition of $\beta_{\bar{m}}$.

Proof that R^{m^*} can be generated by a feasible portfolio

Proof of the Equation, $1 = E\left(m \cdot R^{m^*}\right)$. We have,

$$E(m \cdot R^{m^*}) = \frac{1}{E\left[(m^*)^2\right]} E(m \cdot m^*),$$

where

$$\begin{aligned}
E(m \cdot m^*) &= \bar{m}^2 + E\left[m(R_t - E(R_t))^\top \beta_{\bar{m}}\right] \\
&= \bar{m}^2 + E\left[m(1 + R_t)^\top\right] \beta_{\bar{m}} - E\left[m(1 + E(R_t))^\top\right] \beta_{\bar{m}} \\
&= \bar{m}^2 + \beta_{\bar{m}} - E(m)\left[1 + E(R_t)\right]^\top \beta_{\bar{m}} \\
&= \bar{m}^2 + \left[\mathbf{1}_n - \bar{m}(1 + E(R_t))^\top\right] \beta_{\bar{m}} \\
&= \bar{m}^2 + \left[\mathbf{1}_n - \bar{m}(1 + E(R_t))^\top\right] \Sigma^{-1}\left[\mathbf{1}_n - \bar{m}(\mathbf{1}_n + E(R_t))\right] \\
&= \bar{m}^2 + var(m^*),
\end{aligned}$$

where the last line is due to the definition of m^*.

Proof that R^{m^*} is mean-variance efficient. Let $p = (p_0, p_1, \cdots, p_n)^\top$ the vector of $n+1$ portfolio weights (here $p_i \equiv \pi^i / w$ is the portfolio weight of asset i, $i = 0, 1, \cdots, n$. We have,

$$p^\top \mathbf{1}_{n+1} = 1.$$

The returns we consider are $\underline{r}_t = \left(\bar{m}^{-1} - 1,\ r_{1,t}, \cdots, r_{n,t}\right)^\top$. We denote our "benchmark" portfolio return as $r_{bt} = r^{m^*} - 1$. Next, we build up an arbitrary portfolio yielding the same expected return $E(r_{bt})$ and then we show that this has a variance greater than the variance of r_{bt}. Since this portfolio is arbitrary, the proof will be complete. Let $r_{pt} = p^\top \underline{r}_t$ such that $E(r_{pt}) = E(r_{bt})$. We have:

$$\begin{aligned} cov\left(r_{bt}, r_{pt} - r_{bt}\right) &= E\left[r_{bt} \cdot \left(r_{pt} - r_{bt}\right)\right] \\ &= E\left[R_{bt} \cdot \left(R_{pt} - R_{bt}\right)\right] \\ &= E\left(R_{bt} \cdot R_{pt}\right) - E\left(R_{bt}^2\right) \\ &= \frac{1}{E\left(m^{*2}\right)} E\left[m^*\left(1 + p^\top \underline{r}_t\right)\right] - \frac{1}{\left[E\left(m^{*2}\right)\right]^2} E\left(m^{*2}\right) \\ &= \frac{1}{E\left(m^{*2}\right)}\left\{p^\top E\left[m^*\left(\mathbf{1}_{n+1} + \underline{r}_t\right)\right] - 1\right\} \\ &= 0. \end{aligned}$$

The first line follows by construction since $E(r_{pt}) = E(r_{bt})$. The last line follows because

$$p^\top E\left[m^*\left(\mathbf{1}_{n+1} + \underline{r}_t\right)\right] = p^\top \mathbf{1}_{n+1} = 1.$$

Given this, the claim follows directly from the fact that

$$var\left(R_{pt}\right) = var\left[R_{bt} + \left(R_{pt} - R_{bt}\right)\right] = var\left(R_{bt}\right) + var\left(R_{pt} - R_{bt}\right) \geq var\left(R_{bt}\right).$$

PROOF OF THE EQUATION, $E\left(R^{m^*}\right) - 1 = r - \frac{1+r}{1+Sh} Sh$. We have,

$$E(R^{m^*}) - 1 = \frac{\bar{m}}{E[(m^*)^2]} - 1.$$

In terms of the notation introduced in section 2.8 (p. 55), m^* is:

$$m^* = \bar{m} + (a\epsilon)^\top \beta_{\bar{m}}, \;\; \beta_{\bar{m}} = \sigma^{-1}\left(\mathbf{1}_n - \bar{m}\left\{\mathbf{1}_n + b\right\}\right).$$

We have,

$$\begin{aligned} E[(m^*)^2] &= \left[\bar{m} + (a\epsilon)^\top \beta_{\bar{m}}\right]^2 \\ &= \bar{m}^2 + E\left[(a\epsilon)^\top \beta_{\bar{m}}\right]^2 \\ &= \bar{m}^2 + E\left[(a\epsilon)^\top \beta_{\bar{m}} \cdot (a\epsilon)^\top \beta_{\bar{m}}\right]^2 \\ &= \bar{m}^2 + E\left[\left(\beta_{\bar{m}}^\top a\epsilon\right)\left(\epsilon^\top a^\top \beta_{\bar{m}}\right)\right] \\ &= \bar{m}^2 + \beta_{\bar{m}}^\top \cdot \sigma \cdot \beta_{\bar{m}} \\ &= \bar{m}^2 + \left[\mathbf{1}_n^\top - \bar{m}\left(\mathbf{1}_n^\top + b^\top\right)\right] \sigma^{-1}\left[\mathbf{1}_n - \bar{m}\left(\mathbf{1}_n + b\right)\right] \\ &= \bar{m}^2 + \mathbf{1}_n^\top \sigma^{-1} \mathbf{1}_n - \bar{m}\left(\mathbf{1}_n^\top \sigma^{-1} \mathbf{1}_n + \mathbf{1}_n^\top \sigma^{-1} b\right) \\ &\quad -\bar{m}\left\{\mathbf{1}_n^\top \sigma^{-1} \mathbf{1}_n + b^\top \sigma^{-1} \mathbf{1}_n - \bar{m}\left(\mathbf{1}_n^\top \sigma^{-1} \mathbf{1}_n + b^\top \sigma^{-1} \mathbf{1}_n + \mathbf{1}_n^\top \sigma^{-1} b + b^\top \sigma^{-1} b\right)\right\} \end{aligned}$$

Again in terms of the notation of section 2.8 ($\gamma \equiv \mathbf{1}_n^\top \sigma^{-1} \mathbf{1}_n$ and $\beta \equiv \mathbf{1}_n^\top \sigma^{-1} b$), this is:

$$E\left[(m^*)^2\right] = \gamma - 2\bar{m}(\gamma + \beta) + \bar{m}^2\left(1 + \gamma + 2\beta + b^\top \sigma^{-1} b\right).$$

The expected return is thus,

$$E\left(R^{m^*}\right) - 1 = \frac{E\left(m^*\right)}{E\left[(m^*)^2\right]} - 1 = \frac{\bar{m} - \gamma + 2\bar{m}(\gamma + \beta) - \bar{m}^2\left(1 + \gamma + 2\beta + b^\top \sigma^{-1} b\right)}{\gamma - 2\bar{m}(\gamma + \beta) + \bar{m}^2\left(1 + \gamma + 2\beta + b^\top \sigma^{-1} b\right)}.$$

Now recall two definitions:

$$\bar{m}=\frac{1}{1+r} \quad ; \quad Sh=\left(b-\mathbf{1}_m r\right)^{\top}\sigma^{-1}\left(b-\mathbf{1}_m r\right)=b^{\top}\sigma^{-1}b-2\beta r+\gamma r^2.$$

In terms of r and Sh, we have,

$$\begin{aligned}
E\left(R^{m^*}\right)-1 &= \frac{E\left(m^*\right)}{E\left[\left(m^*\right)^2\right]}-1 \\
&= -\frac{\gamma\left(1+r\right)^2-\left(1+r\right)\left(1+2\gamma+2\beta\right)+1+\gamma+2\beta+b^{\top}\sigma^{-1}b}{\gamma\left(1+r\right)^2-\left(1+r\right)\left(2\gamma+2\beta\right)+1+\gamma+2\beta+b^{\top}\sigma^{-1}b} \\
&= \frac{r-Sh}{1+Sh} \\
&= r-\frac{1+r}{1+Sh}Sh \\
&< r.
\end{aligned}$$

This is positive if $r-Sh>0$, i.e. if $b^{\top}\sigma^{-1}b-\left(2\beta+1\right)r+\gamma r^2<0$, which is possible for sufficiently low (or sufficiently high) values of r.

PROOF THAT R^{m^*} IS THE m-MAXIMUM CORRELATION PORTFOLIO. We have to show that for any price kernel m, $|corr(m,R_{bt})|\geq|corr(m,R_{pt})|$. Define a ℓ-parametrized portfolio such that:

$$E\left[(1-\ell)R_o+\ell R_{pt}\right]=E\left(R_{bt}\right), \quad R_o\equiv\bar{m}^{-1}.$$

We have

$$\begin{aligned}
corr\left(m,R_{pt}\right) &= corr\left[m,(1-\ell)R_o+\ell R_{pt}\right] \\
&= corr\left[m,R_{bt}+\left((1-\ell)R_o+\ell R_{pt}-R_{bt}\right)\right] \\
&= \frac{cov\left(m,R_{bt}\right)+cov\left(m,(1-\ell)R_o+\ell R_{pt}-R_{bt}\right)}{\sigma(m)\cdot\sqrt{var\left((1-\ell)R_o+\ell R_{pt}\right)}} \\
&= \frac{cov\left(m,R_{bt}\right)}{\sigma(m)\cdot\sqrt{var\left((1-\ell)R_o+\ell R_{pt}\right)}}
\end{aligned}$$

The first line follows because $(1-\ell)R_o+\ell R_{pt}$ is a nonstochastic affine translation of R_{pt}. The last equality follows because

$$\begin{aligned}
cov\left(m,(1-\ell)R_o+\ell R_{pt}-R_{bt}\right) &= E\left[m\cdot\left((1-\ell)R_o+\ell R_{pt}-R_{bt}\right)\right] \\
&= (1-\ell)\cdot\underbrace{E\left(mR_o\right)}_{=1}+\ \ell\cdot\underbrace{E\left(mR_{pt}\right)}_{=1}-\ \underbrace{E\left(m\cdot R_{bt}\right)}_{=1} \\
&= 0.
\end{aligned}$$

where the first line follows because $E((1-\ell)R_o+\ell R_{pt})=E(R_{bt})$.

Therefore,

$$corr\left(m,R_{pt}\right)=\frac{cov\left(m,R_{bt}\right)}{\sigma(m)\cdot\sqrt{var\left((1-\ell)R_o+\ell R_{pt}\right)}}\leq\frac{cov\left(m,R_{bt}\right)}{\sigma(m)\cdot\sqrt{var(R_{bt})}}=corr\left(m,R_{bt}\right),$$

where the inequality follows because R_{bt} is mean-variance efficient (i.e. $\nexists$ feasible portfolios with the same expected return as R_{bt} and variance less than $var(R_{bt})$), and then $var((1-\ell)R_o+\ell R_{pt})\geq var(R_{bt})$, all R_{pt}.

References

Campbell, J. Y. and J. Cochrane (1999). "By Force of Habit: A Consumption-Based Explanation of Aggregate Stock Market Behavior," *Journal of Political Economy* 107: 205-251.

Cecchetti, S., Lam, P-S. and N. C. Mark (1994). "Testing Volatility Restrictions on Intertemporal Rates of Substitution Implied by Euler Equations and Asset Returns," *Journal of Finance* 49: 123-152.

Cochrane, J. (1996). "," *Journal of Political Economy*.

Constantinides, G. M. and D. Duffie (1996). "Asset Pricing with Heterogeneous Consumers," *Journal of Political Economy* 104: 219–40.

Ferson, W.E. and A.F. Siegel (2002). "Stochastic Discount Factor Bounds with Conditioning Information," forthcoming in the *Review of Financial Studies*.

Gallant, R.A., L.P. Hansen and G. Tauchen (1990). "Using the Conditional Moments of Asset Payoffs to Infer the Volatility of Intertemporal Marginal Rates of Substitution," *Journal of Econometrics* 45: 141-179.

Hansen, L. P. and R. Jagannathan (1991). "Implications of Security Market Data for Models of Dynamic Economies," *Journal of Political Economy* 99: 225-262.

Mehra, R. and E.C. Prescott (1985). "The Equity Premium: a Puzzle," *Journal of Monetary Economics* 15: 145-161.

Roll 1977.

6
Aggregate stock-market fluctuations

6.1 Introduction

This chapter reviews the progress made to address the empirical puzzles relating to the neoclassical asset pricing model. We first provide a succinct overview of the main empirical regularities of aggregate stock-market fluctuations. For example, we emphasize that price-dividend ratios and returns are procyclical, and that returns volatility and risk-premia are both time-varying and countercyclical. Then, we discuss the extent to which these empirical features can be explained by *rational* models. For example, many models with state dependent preferences predict that Sharpe ratios are time-varying and that stock-market volatility is countercyclical. Are these appealing properties razor-edge? Or are they general properties of all conceivable models with state-dependent preferences? Moreoover, would we expect that these properties show up in other related models in which asset prices are related to the economic conditions? The final part of this chapter aims at providing answers to these questions, and develops theoretical test conditions on the pricing kernel and other primitive state processes that make the resulting models consistent with sets of qualitative predictions given in advance.

6.2 The empirical evidence

One fascinating aspect of stock-market behavior lies in a number of empirical regularities closely related to the business cycle. These empirical regularities seem to persist after controlling for the sample period, and are qualitatively the same in all industrialized countries. These pieces of evidence are extensively surveyed in Campbell (2003). In this section, we summarize the salient empirical features of aggregate US stock market behavior. We rely on the basic statistics and statistical models presented in Table 6.1. There exist three fundamental sets of stylized facts that deserve special attention.

Fact 1. P/D, P/E ratios and returns are strongly procylical, but variations in the general business cycle conditions do not seem to be the only force driving these variables.

For example, Figure 6.1 reveals that price-dividend ratios decline at *all* NBER recessions. But during NBER expansions, price-dividend ratios seem to be driven by additional factors not necessarily related to the business cycle conditions. As an example, during the "roaring" 1960s, price-dividend ratios experienced two major drops having the same magnitude as the decline at the very beginning of the "chaotic" 1970s. Ex-post returns follow approximately the same pattern, but they are more volatile than price-dividend ratios (see Figure 6.2).[1]

A second set of stylized facts is related to the first two moments of the returns distribution:

Fact 2. Returns volatility, the equity premium, risk-adjusted discount rates, and Sharpe ratios are strongly countercyclical. Again, business cycle conditions are not the only factor explaining both short-run and long-run movements in these variables.

Figures 6.3 through 6.5 are informally very suggestive of the previous statement. For example, volatility is markedly higher during recessions than during expansions. (It also appears that the volatility of volatility is countercyclical.) Yet it rocketed to almost 23% during the 1987 crash - a crash occurring during one of the most enduring post-war expansions period. As we will see later in this chapter, countercyclical returns volatility is a property that may emerge when the volatility of the P/D ratio *changes* is countercyclical. Table 1 reveals indeed that the P/D ratios variations are more volatile in bad times than in good times. Table 1 also reveals that the P/D ratio (in *levels*) is more volatile in good times than in bad times. Finally, P/E ratios behave in a different manner.

A third set of very intriguing stylized facts regards the asymmetric behavior of some important variables over the business cycle:

Fact 3. P/D ratio *changes*, risk-adjusted discount rates *changes*, equity premium *changes*, and Sharpe ratio *changes* behave asymmetrically over the business cycle. In particular, the deepest variations of these variables occur during the negative phase of the business cycle.

As an example, not only are risk-adjusted discount rates counter-cyclical. On average, risk-adjusted discount rates increase more during NBER recessions than they decrease during NBER expansions. Analogously, not only are P/D ratios procyclical. On average, P/D increase less during NBER expansions than they decrease during NBER recessions. Furthermore, the order of magnitude of this asymmetric behavior is very high. As an example, the average of P/D percentage (negative) changes during recessions is almost twice as the average of P/D percentage (positive) changes during expansions. It is one objective in this chapter to connect this sort of "concavity" of P/D ratios ("with respect to the business cycle") to "convexity" of risk-adjusted discount rates.[2]

[1] We use "smoothed" ex-post returns to eliminate the noise inherent to high frequency movements in the stock-market.

[2] Volatility of changes in risk-premia related objects appears to be higher during expansions. This is probably a conservative view because recessions have occurred only 16% of the time. Yet during recessions, these variables have moved on average more than they have done during good times. In other terms, "economic time" seems to move more fastly during a recession than during an expansion. For this reason, the "physical calendar time"-based standard deviations in Table 6.1 should be rescaled to reflect unfolding of "economic calendar time". In this case, a more appropriate concept of volatility would be the standard deviation ÷ average of expansions/recessions time.

	total		NBER expansions		NBER recessions	
	average	std dev	average	std dev	average	std dev
P/D	33.951	15.747	35.085	15.538	28.162	15.451
P/E	16.699	6.754	17.210	6.429	14.089	7.672
$\text{P/D}_{t+1} - \text{P/D}_t$	0.063	1.357	0.118	1.255	−0.219	1.758
$\log \frac{\text{P/D}_{t+1}}{\text{P/D}_t} \times 100 \times 12$	2.34	42.228	4.032	37.236	−6.264	60.876
$\text{P/E}_{t+1} - \text{P/E}_t$	0.033	0.762	−0.0003	0.709	0.201	0.967
$\log \frac{\text{P/E}_{t+1}}{\text{P/E}_t} \times 100 \times 12$	2.208	47.004	10.512	41.688	8.964	67.368
$\log \frac{D_{t+1}}{D_t} \times 100 \times 12$	4.901	5.655	5.210	5.710	3.324	5.049
returns (real)	7.810	52.777	9.414	48.423	−0.377	70.183
smooth returns (real)	8.143	16.303	11.997	13.342	−11.529	15.758
risk-free rate (nominal)	5.253	2.762	5.019	2.440	6.442	3.796
risk-free rate (real)	1.346	3.097	1.416	2.866	0.988	4.045
excess returns	6.464	52.387	7.998	48.125	−1.366	69.491
smooth excess returns	6.782	15.802	10.560	12.828	−12.501	15.395
equity premium (π)	8.773	3.459	8.138	3.215	12.012	2.758
$\pi_t - \pi_{t-1}$	0.002	0.544	−0.087	0.506	0.459	0.496
$\log \frac{\pi_t}{\pi_{t-1}} \times 100$	0.023	8.243	−0.788	8.537	4.163	4.644
risk-adjusted rate (Disc)	10.296	3.630	9.635	3.376	13.669	2.917
$\text{Disc}_t - \text{Disc}_{t-1}$	0.004	0.574	−0.091	0.530	0.493	0.536
$\log \frac{\text{Disc}_t}{\text{Disc}_{t-1}} \times 100$	0.048	6.553	−0.711	6.625	3.919	4.438
excess returns volatility	14.938	3.026	14.458	2.711	17.386	3.335
Sharpe ratio (Sh)	0.586	0.192	0.565	0.194	0.696	0.139
$\text{Sh}_t - \text{Sh}_{t-1}$	−0.0001	0.053	−0.003	0.051	0.017	0.061
$\log \frac{\text{Sh}_t}{\text{Sh}_{t-1}} \times 100$	−0.027	10.129	−0.548	10.260	2.630	8.910

TABLE 6.1. P/D and P/E are the S&P Comp. price-dividends and price-earnings ratios. Smooth returns as of time t are defined as $\sum_{i=1}^{12}(\tilde{R}_{t-i} - R_{t-i})$, where $\tilde{R}_t = \log(\frac{q_t + D_t}{q_{t-1}})$, and R is the risk-free rate. Volatility is the excess returns volatility. With the exception of the P/D and P/E ratios, all figures are annualized percent. Data are sampled monthly and cover the period from January 1954 through December 2002. Time series estimates of equity premium π_t (say), excess return volatility σ_t (say) and Sharpe ratios π_t/σ_t are obtained through Maximum Likelihood estimation (MLE) of the following model,

$$\tilde{R}_t - R_t = \pi_t + \varepsilon_t, \quad \varepsilon_t | F_{t-1} \sim N\left(0, \sigma_t^2\right)$$
$$\pi_t = \underset{(0.004)}{0.162} + \underset{(0.005)}{0.766}\, \pi_{t-1} - \underset{(0.015)}{0.146}\, \text{IP}^*_{t-1} \quad ; \quad \sigma_t = \underset{(0.089)}{0.218} + \underset{(0.010)}{0.106}\, |\varepsilon_{t-1}| + \underset{(0.029)}{0.868}\, \sigma_{t-1}$$

where (robust) standard errors are in parentheses; IP is the US real, seasonally adjusted industrial production rate; and IP* is generated by $\text{IP}^*_t = 0.2 \cdot \text{IP}_{t-1} + 0.8 \cdot \text{IP}^*_{t-1}$. Analogously, time series estimates of the risk-adjusted discount rate Disc (say) are obtained by MLE of the following model,

$$\tilde{R}_t - \text{infl}_t = \text{Disc}_t + u_t, \quad u_t | F_{t-1} \sim N\left(0, v_t^2\right)$$
$$\text{Disc}_t = \underset{(0.036)}{0.191} + \underset{(0.042)}{0.767}\, \text{Disc}_{t-1} - \underset{(0.081)}{0.152}\, \text{IP}^*_{t-1} \quad ; \quad v_t = \underset{(0.012)}{0.214} + \underset{(0.004)}{0.105}\, |u_{t-1}| + \underset{(0.003)}{0.869}\, v_{t-1}$$

where infl_t is the inflation rate as of time t.

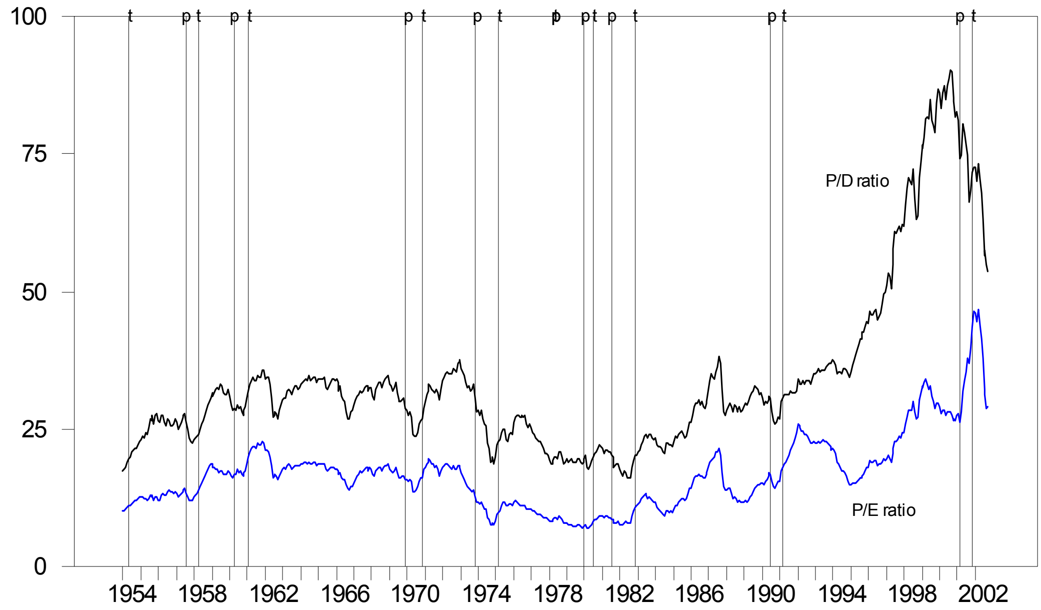


FIGURE 6.1. P/D and P/E ratios

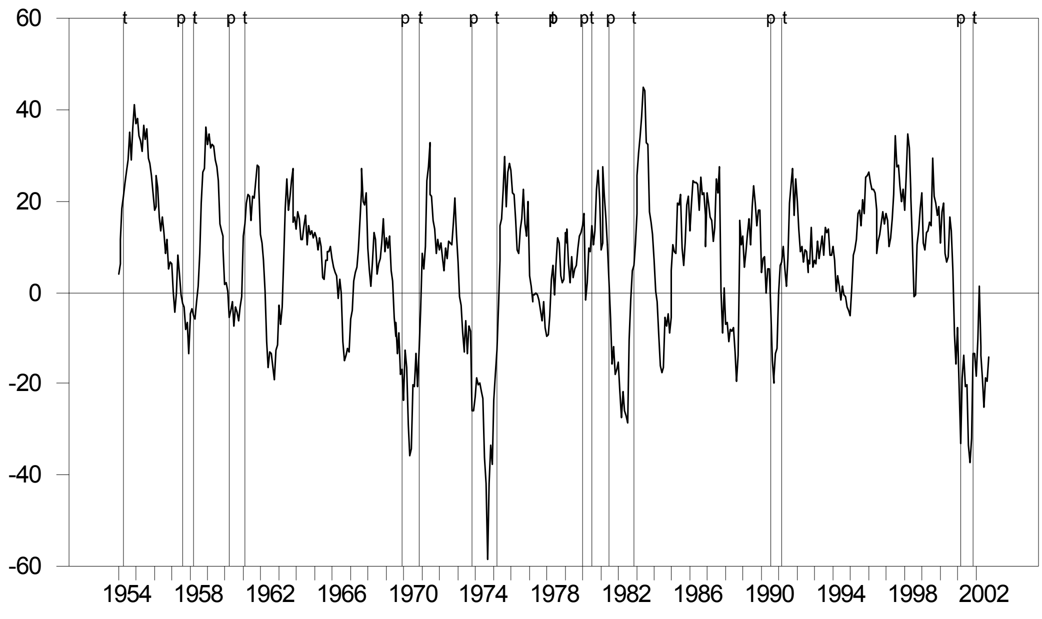


FIGURE 6.2. Monthly smoothed excess returns (%)

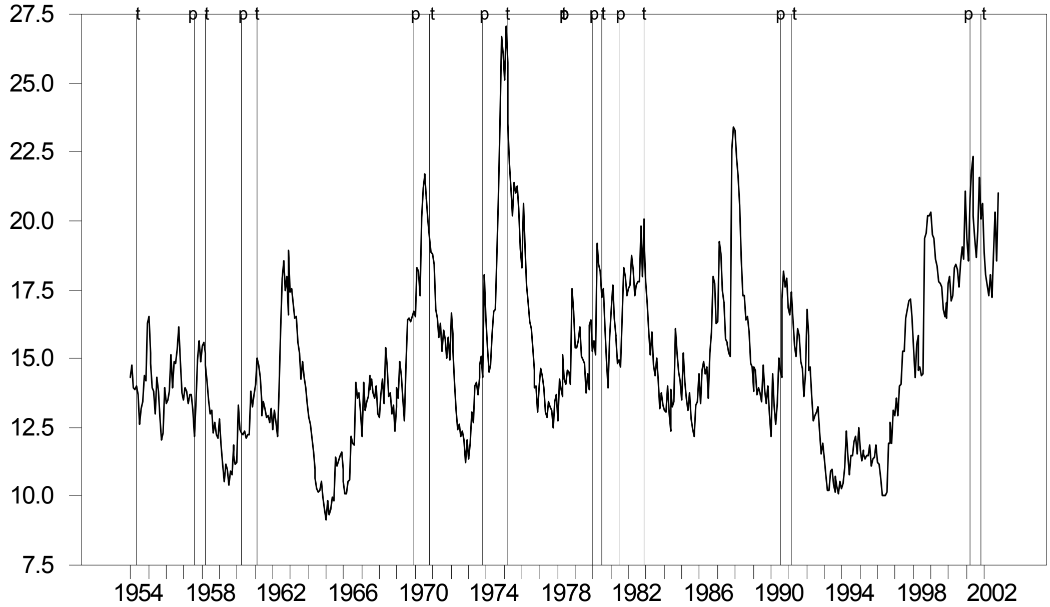


FIGURE 6.3. Excess returns volatility

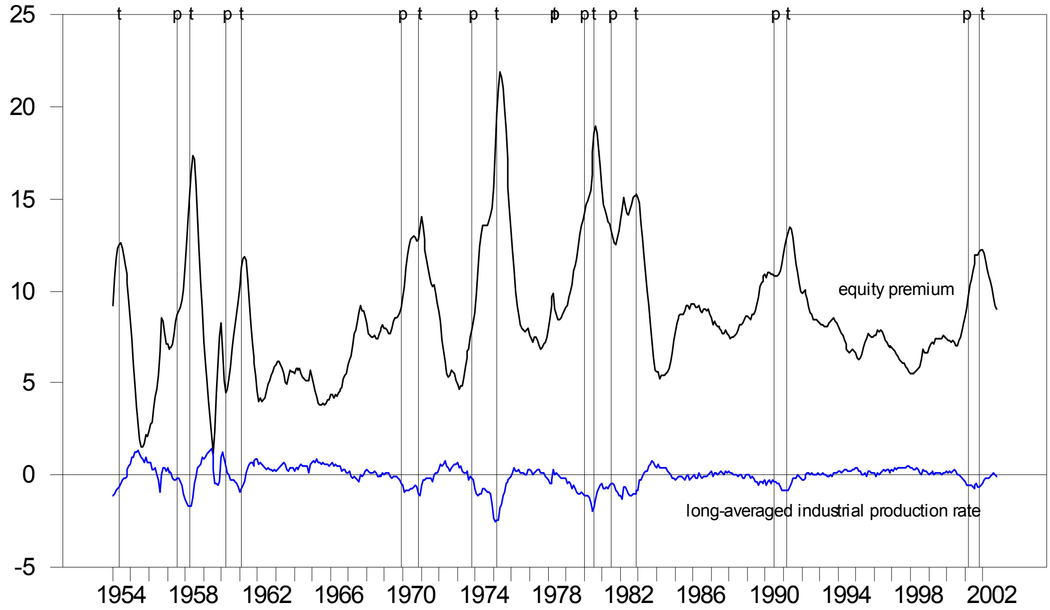


FIGURE 6.4. Equity premium and long-averaged real industrial production rate

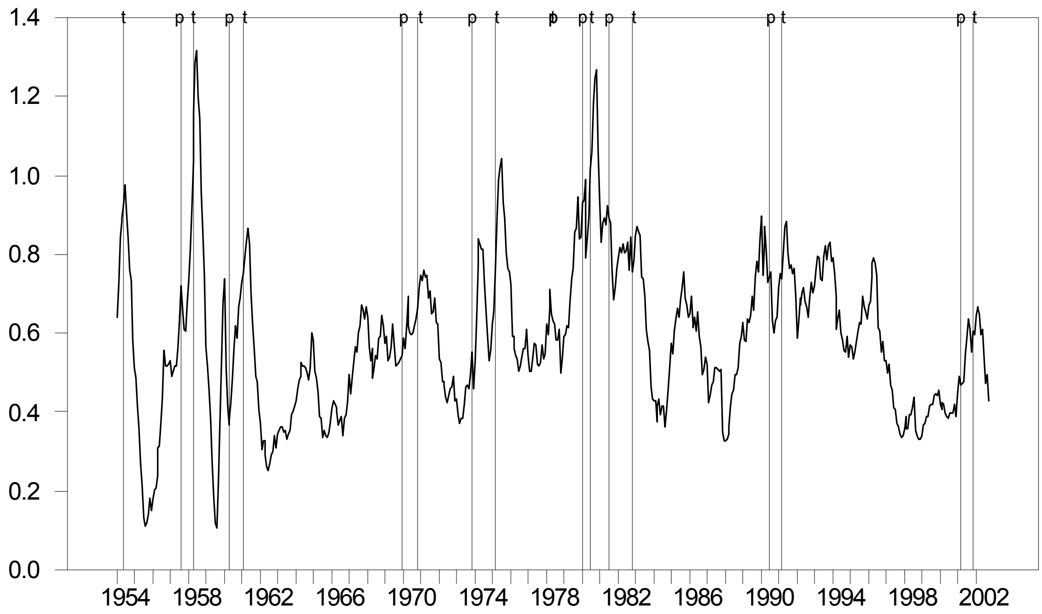


FIGURE 6.5. Sharpe ratio

Stylized fact 1 has a simple and very intuitive consequence: price-dividend ratios are somewhat related to, or "predict", future medium-term returns. The economic content of this prediction is simple. After all, expansions are followed by recessions. Therefore in good times the stock-market predicts that in the future, returns will be negative. Indeed, define the excess return as $\tilde{R}^e_t \equiv \tilde{R}_t - R_t$. Consider the following regressions,

$$\tilde{R}^e_{t+n} = a_n + b_n \times \mathrm{P/D}_t + u_{n,t}, \quad n \geq 1,$$

where u is a residual term. Typically, the estimates of b_n are significantly negative, and the R^2 on these regressions increases with n.[3] In turn, the previous regressions imply that $E[\tilde{R}^e_{t+n} \big| \mathrm{P/D}_t] = a_n + b_n \times \mathrm{P/D}_t$. They thus suggest that price-dividend ratios are driven by expected excess returns. In this restrictive sense, countercyclical expected returns (stated in stylized fact 2) and procyclical price-dividend ratios (stated in stylized fact 1) seem to be the two sides of the same coin.

There is also one apparently puzzling feature: price-dividend ratios do not predict future dividend growth. Let $g_t \equiv \log(D_t/D_{t-1})$. In regressions of the following form,

$$g_{t+n} = a_n + b_n \times \mathrm{P/D}_t + u_{n,t}, \quad n \geq 1,$$

the predictive content of price-dividend ratios is very poor, and estimates of b_n even come with a wrong sign.

The previous simple regressions thus suggest that: 1) price-dividend ratios are driven by time-varying expected returns (i.e. by time-varying risk-premia); and 2) the role played by expected

[3]See Volkanov JFE for a critique on long-run regressions.

dividend growth seems to be somewhat limited. As we will see later in this chapter, this view can however be challenged along several dimensions. First, it seems that expected *earning growth* does help predicting price-dividend ratios. Second, the fact that expected dividend growth doesn't seem to affect price-dividend ratios can in fact be a property to be expected in equilibrium.

6.3 Understanding the empirical evidence

Consider the following decomposition,

$$\log \tilde{R}_{t+1} \equiv \log \frac{q_{t+1} + D_{t+1}}{q_t} = g_{t+1} + \log \frac{1 + p_{t+1}}{p_t}; \qquad \text{where } g_t \equiv \log \frac{D_t}{D_{t-1}} \text{ and } p_t \equiv \frac{q_t}{D_t}.$$

The previous formula reveals that properties of returns can be understood through the corresponding properties of dividend growth g_t *and* price-dividend ratios p_t. The empirical evidence discussed in the previous section suggests that our models should take into account at least the following two features. First, we need volatile price-dividend ratios. Second we need that price-dividend ratios be on average more volatile in bad times than in good times. For example, consider a model in which prices are affected by some key state variables related to the business cycle conditions (see section 6.4 for examples of models displaying this property). A basic property that we should require from this particular model is that the price-dividend ratio be increasing and *concave* in the state variables related to the business cycle conditions. In particular, the concavity property ensures that returns volatility increases on the downside - which is precisely the very definition of countercyclical returns volatility. One of the ultimate scopes in this chapter is to search for classes of promising models ensuring this and related properties.

The Gordon's model in Chapter 4 and 5 predicts that price dividend ratios are constant - which is counterfactual. It is thus unsuitable for the scopes we are pursuing here. We need to think of multidimensional models. However, not all multidimensional models will work. As an example, in the previous chapter we showed how to arbitrarily increase the variance of the kernel of the Lucas model by adding more and more factors. We also showed that the resulting model is one in which price-dividend ratios are constant. We need to impose some discipline on how to increase the dimension of a model.

6.4 The asset pricing model

6.4.1 A multidimensional model

Consider a reduced-form rational expectation model in which the rational asset price q_i is a twice-differentiable function of a number of factors,

$$q_i = q_i(y), \qquad y \in \mathbb{R}^d, \qquad i = 1, \cdots, m \ (m \leq d),$$

where $y = [y_1, \cdots, y_d]^\top$ is the vector of factors affecting asset prices, and q_i is the rational pricing function. We assume that asset i pays off an instantaneous dividend rate D_i, $i = 1, \cdots, m$, and that $D_i = D_i(y)$, $i = 1, \cdots, m$. We also assume that y is a multidimensional diffusion process, viz

$$dy_t = \varphi(y_t)dt + v(y_t)dW_t,$$

where φ is d-valued, v is $d \times d$ valued, and W is a d-dimensional Brownian motion. By Itô's lemma,

$$\frac{dq_i}{q_i} = \frac{Lq_i}{q_i}dt + \frac{\overbrace{\nabla q_i}^{1\times d}}{q_i}\overbrace{v}^{d\times d}dW,$$

where Lq_i is the usual infinitesimal operator. Let $r = \{r_t\}_{t\geq 0}$ be the instantaneous short-term rate process. We now invoke the FTAP to claim that under mild regularity conditions, there exists a measurable d-vector process λ (unit prices of risk) such that,

$$\begin{bmatrix} \frac{Lq_1}{q_1} - r + \frac{D_1}{q_1} \\ \vdots \\ \frac{Lq_m}{q_m} - r + \frac{D_m}{q_m} \end{bmatrix} = \underbrace{\sigma}_{m\times d}\underbrace{\lambda}_{d\times 1}, \qquad \text{where } \sigma = \begin{bmatrix} \frac{\nabla q_1}{q_1} \\ \vdots \\ \frac{\nabla q_m}{q_m} \end{bmatrix} \cdot v. \tag{6.1}$$

The usual interpretation of λ is the vector of unit prices of risk associated with the fluctuations of the d factors. To simplify the structure of the model, we suppose that,

$$r \equiv r(y_t) \quad \text{and} \quad \lambda \equiv \lambda(y_t). \tag{6.2}$$

The previous assumptions impose a series of severe restrictions on the dimension of the model. We emphasize that these restrictions are arbitrary, and that they are only imposed for simplicity sake.

Eqs. (6.1) constitute a system of m uncoupled partial differential equations. The solution to it is an equilibrium price system. For example, the Gordon's model in Chapter 4 is a special case of this setting.[4] We do not discuss transversality conditions and bubbles in this chapter. Nor we discuss issues related to market completeness.[5] Instead, we implement a reverse-engineering approach and search over families of models guaranteeing that long-lived asset prices exhibit some properties given in advance. In particular, we wish to impose conditions on the primitives $\mathcal{P} \equiv (a, b, r, \lambda)$ such that the aggregate stock-market behavior exhibits the same patterns surveyed in the previous section. For example, model (6.1) predicts that returns volatility is,

$$volatility\left(\frac{dq_{i,t}}{q_{i,t}}\right) \equiv \mathcal{V}(y_t)dt \equiv \frac{1}{q_{i,t}^2}\left\|\nabla q_i(y_t)v(y_t)\right\|^2 dt.$$

In this model volatility is thus typically time-varying. But we also wish to answer questions such as, Which restrictions may we impose to $\mathcal{P}$ to ensure that volatility $\mathcal{V}(y_t)$ is countercyclical?

Naturally, an important and challenging subsequent step is to find models guaranteeing that the restrictions on $\mathcal{P}$ we are looking for are economically and quantitatively sensible. The most natural models we will look at are models which innovate over the Gordon's model due to time-variation in the expected returns and/or in the expected dividend growth. These issues are analyzed in a simplified version of the model.

6.4.2 A simplified version of the model

We consider a pure exchange economy endowed with a flow of a (single) consumption good. To make the presentation simple, we assume that consumption equals the dividends paid by a *single*

[4]Let $m = d = 1$, $\zeta = y$, $\varphi(y) = \mu y$ and $\xi(y) = \sigma_0 y$, and assume that λ and r are constant. By replacing these things into eq. (6.1) and assuming no-bubbles yields the (constant) price-dividend ratio predicted by the Gordon's model, $q_t/\zeta_t = (\mu - r - \sigma_0\lambda)^{-1}$.

[5]As we explained in chapter 4, in this setting markets are complete if and only if $m = d$.

long-lived asset (see below). We also assume that $d = 2$, and take as given the consumption endowment process z and a second state variable y. We assume that z, y are solution to,

$$\begin{cases} dz(\tau) = m\,(y(\tau))\,z\,(\tau)\,d\tau + \sigma_0 z(\tau) dW_1(\tau) \\ \\ dy(\tau) = \varphi(z(\tau), y(\tau))d\tau + v_1\,(y(\tau))\,dW_1(\tau) + v_2\,(y(\tau))\,dW_2(\tau) \end{cases}$$

where W_1 and W_2 are independent standard Brownian motions. By the connection between conditional expectations and solutions to partial differential equations (the Feynman-Kac representation theorem) (see Chapter 4), we may re-state the FTAP in (6.1) in terms of conditional expectations in the following terms. By (6.1), we know that q is solution to,

$$Lq + z = rq + (q_z \sigma_0 z + q_y v_1)\,\lambda_1 + q_y v_2 \lambda_2, \quad \forall (z, y) \in \mathbb{Z} \times \mathbb{Y}. \tag{6.3}$$

Under regularity conditions, the Feynman-Kac representation of the solution to Eq. (6.3) is:

$$q(z, y) = \int_0^\infty C(z, y, \tau) d\tau, \qquad C(z, y, \tau) \equiv \mathbb{E}\left[\exp\left(-\int_0^\tau r(z(t), y(t))dt\right) \cdot z(\tau) \bigg|\, z, y\right], \tag{6.4}$$

where $\mathbb{E}$ is the expectation operator taken under the risk-neutral probability Q (say).[6] Finally, (Z, Y) are solution to

$$\begin{cases} dz(\tau) = \hat{m}(y(\tau))z(\tau)d\tau + \sigma_0 z(\tau) d\hat{W}_1(\tau) \\ dy(\tau) = \hat{\varphi}(z(\tau), y(\tau))d\tau + v_1\,(y(\tau))\,d\hat{W}_1(\tau) + v_2\,(y(\tau))\,d\hat{W}_2(\tau) \end{cases} \tag{6.5}$$

where $\hat{W}_1$ and $\hat{W}_2$ are two independent Q-Brownian motions, and $\hat{m}$ and $\hat{\varphi}$ are risk-adjusted drift functions defined as $\hat{m}\,(z, y) \equiv m(y)z - \sigma_0 z \lambda_1\,(z, y)$ and $\hat{\varphi}\,(z, y) \equiv \varphi\,(z, y) - v_1\,(y)\,\lambda_1\,(z, y) - v_2\,(y)\,\lambda_2\,(z, y)$.[7] Naturally, Eq. (6.4) can also be rewritten under the physical measure. We have,

$$C(z, y, \tau) = \mathbb{E}\left[\exp\left(-\int_0^\tau r(z(t), y(t))dt\right) \cdot z(\tau) \bigg|\, z, y\right] = E\left[\mu\,(\tau) \cdot z(\tau)|\, z, y\right],$$

where μ is the stochastic discount factor of the economy:

$$\mu\,(\tau) = \frac{\xi\,(\tau)}{\xi\,(0)}; \quad \xi\,(0) = 1.$$

Given the previous assumptions on the information structure of the economy, ξ necessarily satisfies,

$$\frac{d\xi(\tau)}{\xi\,(\tau)} = -\left[r(z(\tau), y(\tau))d\tau + \lambda_1(z(\tau), y(\tau))dW_1(\tau) + \lambda_2(z(\tau), y(\tau))dW_2(\tau)\right]. \tag{6.6}$$

In the appendix ("Markov pricing kernels"), we provide an example of pricing kernel generating interest rates and risk-premia having the same functional form as in (6.2).

---

[6] See, for example, Huang and Pagès (1992) (thm. 3, p. 53) or Wang (1993) (lemma 1, p. 202), for a series of regularity conditions underlying the Feynman-Kac theorem in infinite horizon settings arising in typical financial applications.

[7] See, for example, Huang and Pagès (1992) (prop. 1, p. 41) for mild regularity conditions ensuring that Girsanov's theorem holds in infinite horizon settings.

6.4.3 Issues

We analyze general properties of long-lived asset prices that can be streamlined into three categories: "monotonicity properties", "convexity properties", and "dynamic stochastic dominance properties". We now produce examples illustrating the economic content of such a categorization.

- *Monotonicity.* Consider a model predicting that $q(z, y) = z \cdot p(y)$, for some positive function $p \in \mathcal{C}^2(\mathbb{Y})$. By Itô's lemma, returns volatility is $\text{vol}(z) + \frac{p'(y)}{p(y)}\text{vol}(y)$, where $\text{vol}(z) > 0$ is consumption growth volatility and $\text{vol}(y)$ has a similar interpretation. As explained in the previous chapter, actual returns volatility is too high to be explained by consumption volatility. Naturally, additional state variables may increase the overall returns volatility. In this simple example, state variable y inflates returns volatility whenever the price-dividend ratio p is increasing in y. At the same time, such a monotonicity property would ensure that asset returns volatility be strictly positive. Eventually, strictly positive volatility is one crucial condition guaranteeing that dynamic constraints of optimizing agents are well-defined.

- *Concavity.* Next, suppose that y is some state variable related to the business cycle conditions. Another robust stylized fact is that stock-market volatility is countercyclical. If $q(z, y) = z \cdot p(y)$ and $\text{vol}(y)$ is constant, returns volatility is countercyclical whenever p is a *concave* function of y. Even in this simple example, second-order properties (or "nonlinearities") of the price-dividend ratio are critical to the understanding of time variation in returns volatility.

- *Convexity.* Alternatively, suppose that expected dividend growth is positively affected by a state variable g. If p is increasing and *convex* in $y \equiv g$, price-dividend ratios would typically display "overreaction" to small changes in g. The empirical relevance of this point was first recognized by Barsky and De Long (1990, 1993). More recently, Veronesi (1999) addressed similar convexity issues by means of a fully articulated equilibrium model of learning.

- *Dynamic stochastic dominance.* An old issue in financial economics is about the relation between long-lived asset prices and volatility of fundamentals.[8] The traditional focus of the literature has been the link between dividend (or consumption) volatility and stock prices. Another interesting question is the relationship between the volatility of additional state variables (such as the dividend growth rate) and stock prices. In some models, volatility of these additional state variables is endogenously determined. For example, it may be inversely related to the quality of signals about the state of the economy.[9] In many other circumstances, producing a probabilistic description of y is as arbitrary as specifying the preferences of a representative agent. In fact, y is in many cases related to the dynamic specification of agents' preferences. The issue is then to uncover stochastic dominance properties of dynamic pricing models where state variables are possible nontradable.

In the next section, we provide a simple characterization of the previous properties. To achieve this task, we extend some general ideas in the recent option pricing literature. This literature

[8]See, for example, Malkiel (1979), Pindyck (1984), Poterba and Summers (1985), Abel (1988) and Barsky (1989).

[9]See, for example, David (1997) and Veronesi (1999, 2000)

attempts to explain the qualitative behavior of a contingent claim price function $C(z, y, \tau)$ with as few assumptions as possible on z and y. Unfortunately, some of the conceptual foundations in this literature are not well-suited to pursue the purposes of this chapter. As an example, many available results are based on the assumption that at least one state variable is tradable. This is not the case of the "European-type option" pricing problem (6.4). In the next section, we introduce an abstract asset pricing problem which is appropriate to our purposes. Many existing results are specific cases of the general framework developed in the next section (see theorems 1 and 2). In sections 6.6 and 6.7, we apply this framework of analysis to study basic model examples of long-lived asset prices.

6.5 Analyzing qualitative properties of models

Consider a two-period, risk-neutral environment in which there is a right to receive a cash premium ψ at the second period. Assume that interest rates are zero, and that the cash premium is a function of some random variable $\tilde{x}$, viz $\psi = \psi(\tilde{x})$. Finally, let $\bar{c} \equiv \mathbb{E}[\psi(\tilde{x})]$ be the price of this right. What is the relationship between the volatility of $\tilde{x}$ and $\bar{c}$? By classical second-order stochastic dominance arguments (see chapter 2), $\bar{c}$ is inversely related to *mean preserving* spreads in $\tilde{x}$ if ψ is concave. Intuitively, this is so because a concave function "exaggerates" poor realizations of $\tilde{x}$ and "dampens" the favorable ones.

Do stochastic dominance properties still hold in a dynamic setting? Consider for example a multiperiod, continuous time extension of the previous risk-neutral environment. Assume that the cash premium ψ is paid off at some future date T, and that $\tilde{x} = x(T)$, where $X = \{x(\tau)\}_{\tau\in[0,T]}$ $(x(0) = x)$ is some underlying state process. If the yield curve is flat at zero, $c(x) \equiv \mathbb{E}[\psi(x(T))|\, x]$ is the price of the right. Clearly, the pricing problem $\mathbb{E}[\psi(x(T))|\, x]$ is different from the pricing problem $\mathbb{E}[\psi(\tilde{x})]$. Analogies exist, however. First, if X is a proportional process (one for which the risk-neutral distribution of $x(T)/x$ is independent of x),

$$c(x) = \mathbb{E}\left[\psi(x \cdot G(T))\right], \quad G(T) \equiv \frac{x(T)}{x}, \quad x > 0.$$

As this simple formula reveals, standard stochastic dominance arguments still apply: c decreases (increases) after a mean-preserving spread in G whenever ψ is concave (convex) - consistently for example with the prediction of the Black and Scholes (1973) formula. This point was first made by Jagannathan (1984) (p. 429-430). In two independent papers, Bergman, Grundy and Wiener (1996) (BGW) and El Karoui, Jeanblanc-Picqué and Shreve (1998) (EJS) generalized these results to *any* diffusion process (i.e., not necessarily a proportional process).[10,11] But one crucial assumption of these extensions is that X must be the price of a traded asset that does not pay dividends. This assumption is crucial because it makes the risk-neutralized drift function of X proportional to x. As a consequence of this fact, c inherits convexity properties of ψ, as in the proportional process case. As we demonstrate below, the presence of nontradable

[10] The proofs in these two articles are markedly distinct but are both based on *price function* convexity. An alternate proof directly based on *payoff function* convexity can be obtained through a direct application of Hajek's (1985) theorem. This theorem states that if ψ is increasing and convex, and X_1 and X_2 are two diffusion processes (both starting off from the same origin) with integrable drifts b_1 and b_2 and volatilities a_1 and a_2, then $E[\psi(x_1(T))] \leq E[\psi(x_2(T))]$ whenever $m_2(\tau) \leq m_1(\tau)$ and $a_2(\tau) \leq a_1(\tau)$ for all $\tau \in (0, \infty)$. Note that this approach is more general than the approach in BGW and EJS insofar as it allows for shifts in both m and a. As we argue below, both shifts are important to account for when X is nontradable.

[11] Bajeux-Besnainou and Rochet (1996) (section 5) and Romano and Touzi (1997) contain further extensions pertaining to stochastic volatility models.

state variables makes interesting nonlinearites emerge. As an example, Proposition 6.1 reveals that in general, convexity of ψ is neither a necessary or a sufficient condition for convexity of c.[12] Furthermore, "dynamic" stochastic dominance properties are more intricate than in the classical second order stochastic dominance theory (see Proposition 6.1).

To substantiate these claims, we now introduce a simple, abstract pricing problem.

Canonical pricing problem. *Let X be the (strong) solution to:*

$$dx(\tau) = b\,(x(\tau))\,d\tau + a\,(x(\tau))\,d\bar{W}(\tau),$$

where $\bar{W}$ is a multidimensional $\bar{P}$-Brownian motion (for some $\bar{P}$), and b, a are some given functions. Let ψ and ρ be two twice continuously differentiable positive functions, and define

$$c(x,T) \equiv \bar{\mathbb{E}}\left[\exp\left(-\int_0^T \rho(x(t))dt\right)\cdot\psi(x(T))\bigg|\, x\right] \tag{6.7}$$

to be the price of an asset which promises to pay $\psi(x(T))$ at time T.

In this pricing problem, X can be the price of a traded asset. In this case $b(x) = x\rho(x)$. If in addition, $\rho' = 0$, the problem collapses to the classical European option pricing problem with constant discount rate. If instead, X is not a traded risk, $b(x) = b_0(x) - a(x)\lambda(x)$, where b_0 is the physical drift function of X and λ is a risk-premium. The previous framework then encompasses a number of additional cases. As an example, set $\psi(x) = x$. Then, one may 1) interpret X as consumption process; 2) restrict a *long-lived* asset price q to be driven by consumption only, and set $q = \int_0^\infty c(x,\tau)d\tau$. As another example, set $\psi(x) = 1$ and $\rho(x) = x$. Then, c is a zero-coupon bond price as predicted by a simple univariate short-term rate model. The importance of these specific cases will be clarified in the following sections.

In the appendix (see proposition 6.A.1), we provide a result linking the volatility of the state variable x to the price c. Here I characterize slope (c_x) and convexity (c_{xx}) properties of c. We have:

Proposition 6.1. *The following statements are true:*

a) If $\psi' > 0$, then c is increasing in x whenever $\rho' \leq 0$. Furthermore, if $\psi' = 0$, then c is decreasing (resp. increasing) whenever $\rho' > 0$ (resp. < 0).

b) If $\psi'' \leq 0$ (resp. $\psi'' \geq 0$) and c is increasing (resp. decreasing) in x, then c is concave (resp. convex) in x whenever $b'' < 2\rho'$ (resp. $b'' > 2\rho'$) and $\rho'' \geq 0$ (resp. $\rho'' \leq 0$). Finally, if $b'' = 2\rho'$, c is concave (resp. convex) whenever $\psi'' < 0$ (resp. > 0) and $\rho'' \geq 0$ (resp. ≤ 0).

Proposition 6.1-a) generalizes previous monotonicity results obtained by Bergman, Grundy and Wiener (1996). By the so-called "no-crossing property" of a diffusion, X is not decreasing in its initial condition x. Therefore, c inherits the same monotonicity features of ψ if discounting does not operate adversely. While this observation is relatively simple, it explicitly allows to address monotonicity properties of long-lived asset prices (see the next section).

Proposition 6.1-b) generalizes a number of existing results on option price convexity. First, assume that ρ is constant and that X is the price of a traded asset. In this case, $\rho' = b'' = 0$.

[12] Kijima (2002) recently produced a counterexample in which option price convexity may break down in the presence of convex payoff functions. His counterexample was based on an extension of the Black-Scholes model in which the underlying asset price had a concave drift function. (The source of this concavity was due to the presence of dividend issues.) Among other things, the proof of proposition 2 reveals the origins of this counterexample.

The last part of Proposition 6.1-b) then says that convexity of ψ propagates to convexity of c. This result reproduces the findings in the literature that surveyed earlier. Proposition 6.1-b) characterizes option price convexity within more general contingent claims models. As an example, suppose that $\psi'' = \rho' = 0$ and that X is not a traded risk. Then, Proposition 6.1-b) reveals that c inherits the same convexity properties of the instantaneous drift of X. As a final example, Proposition 6.1-b) extends one (scalar) bond pricing result in Mele (2003). Precisely, let $\psi(x) = 1$ and $\rho(x) = x$; accordingly, c is the price of a zero-coupon bond as predicted by a standard short-term rate model. By Proposition 6.1-b), c is convex in x whenever $b''(x) < 2$. This corresponds to Eq. (8) (p. 688) in Mele (2003).[13] In analyzing properties of long-lived asset prices, both discounting and drift nonlinearities play a prominent role.

An intuition of the previous result can be obtained through a Taylor-type expansion of $c(x,T)$ in Eq. (6.7). To simplify, suppose that in Eq. (6.7), $\psi \equiv 1$, and that

$$-\rho(x) = g(x) - \text{Disc}(x).$$

The economic interpretation of the previous decomposition is that g is the growth rate of some underlying "dividend process" and Disc is some "risk-adjusted" discount rate. Consider the following discrete-time counterpart of Eq. (6.7):

$$c(x_0, N) \equiv \bar{\mathbb{E}}\left\{ e^{\sum_{i=0}^{N}[g(x_i) - \text{Disc}(x_i)]} \middle| x_0 \right\}.$$

For small values of the exponent,

$$c(x, N) \approx 1 + [g(x) - \text{Disc}(x)] \times N + \sum_{i=0}^{N} \sum_{j=1}^{i} \bar{\mathbb{E}}\left[\Delta g(x_j) - \Delta\text{Disc}(x_j) \middle| x\right], \tag{6.8}$$

where $\Delta g(x_j) \equiv g(x_j) - g(x_{j-1})$. The second term of the r.h.s of Eq. (6.8) make clear that convexity of g can potentially translate to convexity of c w.r.t x; and that convexity of Disc can potentially translate to concavity of c w.r.t x. But Eq. (6.8) reveals that higher order terms are important too. Precisely, the expectation $\bar{\mathbb{E}}[\Delta g(x_j) - \Delta\text{Disc}(x_j)| x]$ plays some role. Intuitively, convexity properties of c w.r.t x also depend on convexity properties of this expectation. In discrete time, these things are difficult to see. But in continuous time, this simple observation translates to a joint restriction on the law of movement of x. Precisely, convexity properties of c w.r.t x will be somehow inherited by convexity properties of the drift function of x. In continuous time, Eq. (6.8) becomes, for small T,[14]

$$\begin{aligned} c(x,T) &\approx 1 + [g(x) - \text{Disc}(x)] \times T \\ &+ \left\{ [g(x) - \text{Disc}(x)]^2 + b(x) \cdot \frac{d}{dx}[g(x) - \text{Disc}(x)] + \frac{1}{2} a(x)^2 \cdot \frac{d^2}{dx^2}[g(x) - \text{Disc}(x)] \right\} \times T^2. \end{aligned} \tag{6.9}$$

Naturally, this formula is only an approximation. Importantly, it doesn't work very well for large T. Proposition 6.1 gives the exact results.

[13] In the appendix, we have developed further intuition on this bounding number.

[14] Eq. (6.9) can derived with operation-theoretic arguments based on the functional iteration of the infinitesimal generator for Markov processes.

6.6 Time-varying discount rates and equilibrium volatility

Campbell and Cochrane (1999) model of external habit formation is certainly one of the most well-known attempts at explaining some of the empirical features outlined in section 6.2. Consider an infinite horizon, complete markets economy. There is a (representative) agent with undiscounted instantaneous utility given by

$$u(c,x)=\frac{(c-x)^{1-\eta}-1}{1-\eta}, \tag{6.10}$$

where c is consumption and x is a (time-varying) habit, or (exogenous) "subsistence level". The total endowment process $Z=\{z(\tau)\}_{\tau\geq 0}$ satisfies,

$$\frac{dz(\tau)}{z(\tau)}=g_0 d\tau+\sigma_0 dW(\tau). \tag{6.11}$$

In equilibrium, $C=Z$. Let $s\equiv(z-x)/z$, the "surplus consumption ratio". By assumption, $S=\{s(\tau)\}_{\tau\geq 0}$ is solution to:

$$ds(\tau)=s(\tau)\left[(1-\phi)(\bar{s}-\log s(\tau))+\frac{1}{2}\sigma_0^2 l(s(\tau))^2\right]d\tau+\sigma_0 s(\tau)l(s(\tau))dW(\tau), \tag{6.12}$$

where l is a positive function given in appendix 6.2. This function l turns out to be decreasing in s; and convex in s for the empirically relevant range of variation of s. The Sharpe ratio predicted by the model is:

$$\lambda(s)=\eta\sigma_0\left[1+l(s)\right] \tag{6.13}$$

(see appendix 6.2 for details). Finally, Campbell and Cochrane choose function l so as to make the short-term rate constant.

The economic interpretation of the model is relatively simple. Consider first the instantaneous utility in (6.10). It is readily seen that CRRA $=\eta s^{-1}$. That is, risk aversion is countercyclical in this model. Intuitively, during economic downturns, the surplus consumption ratio s decreases and agents become more risk-averse. As a result, prices decrease and expected returns increase. Very nice, but this is still a model of high risk-aversion. Furthermore, Barberis, Huang and Santos (2001) have a similar mechanism based on behavioral preferences.

The rationale behind Eq. (6.12) is simply that the log of s is a mean reverting process. By taking logs, we are sure that s remains positive. Finally, $\log s$ is also conditionally heteroskedastic since its instantaneous volatility is $\sigma_0 l$. Because l is decreasing in s and s is clearly pro-cyclical, the volatility of $\log s$ is countercyclical. This feature is responsible of many interesting properties of the model (such as countercyclical returns volatility).

Finally, the Sharpe ratio λ in (6.13) is made up of two components. The first component is $\eta\sigma_0$ and coincides with the Sharpe ratio predicted by the standard Gordon's model. The additional component $\eta\sigma_0 l(s)$ arises as a compensation related to the stochastic fluctuations of $x=z(1-s)$. By the functional form of l picked up by the authors, λ is therefore countercyclical. Since ϕ is high, this generates a slowly-varying, countercyclical risk-premium in stock-returns. Finally, numerical results revealed that one important prediction of the model is that the price-dividend ratio is a concave function of s. In appendix 6.3, we provide an algorithm that one may use to solve this and related models numerically (we only consider discrete time models in appendix 6.3). Here we now clarify the theoretical link between convexity of l and concavity of the price-dividend ratio in this and related models.

We aim at writing the solution in the canonical pricing problem format of section 6.5, and then at applying Proposition 6.1. Our starting point is the evaluation formula (6.4). To apply it here, we might note that interest rate are constant. Yet to gain in generality we continue to assume that they are state dependent, but that they only depend on s. Therefore Eq. (6.4) becomes,

$$\frac{q(z,s)}{z}=\int_0^\infty \frac{C(z,s,\tau)}{z}d\tau=\int_0^\infty \mathbb{E}\left[e^{-\int_0^\tau r(s(u))du}\cdot\frac{z(\tau)}{z}\right| z,s\left.\right] d\tau. \tag{6.14}$$

To compute the inner expectation, we have to write the dynamics of Z under the risk-neutral probability measure. By Girsanov theorem,

$$\frac{z(\tau)}{z}=e^{-\frac{1}{2}\sigma_0^2\tau+\sigma_0\hat{W}(\tau)}\cdot e^{g_0\tau-\int_0^\tau \sigma_0\lambda(s(u))du},$$

where $\hat{W}$ is a Brownian motion under the risk-neutral measure. By replacing this into Eq. (6.14),

$$\frac{q(z,s)}{z}=\int_0^\infty e^{g_0\tau}\cdot\mathbb{E}\left[e^{-\frac{1}{2}\sigma_0^2\tau+\sigma_0\hat{W}(\tau)}\cdot e^{-\int_0^\tau \mathrm{Disc}(s(u))du}\right| z,s\left.\right] d\tau, \tag{6.15}$$

where

$$\mathrm{Disc}\,(s)\equiv r\,(s)+\sigma_0\lambda\,(s)$$

is the "risk-adjusted" discount rate. Note also, that under the risk-neutral probability measure,

$$ds(\tau)=\hat{\varphi}\,(s(\tau))\,d\tau+v(s\,(\tau))d\hat{W}(\tau),$$

where $\hat{\varphi}\,(s)=\varphi\,(s)-v\,(s)\,\lambda\,(s)$, $\varphi\,(s)=s[(1-\phi)(\bar{s}-\log s)+\frac{1}{2}\sigma_0^2 l(s)^2]$ and $v\,(s)=\sigma_0 s l(s)$.

Eq. (6.15) reveals that the price-dividend ratio $p\,(z,s)\equiv q\,(z,s)/\,z$ is independent of z. Therefore, $p\,(z,s)=p\,(s)$. To obtain a neat formula, we should also get rid of the $e^{-\frac{1}{2}\sigma_0^2\tau+\sigma_0\hat{W}(\tau)}$ term. Intuitively this term arises because consumption and habit are correlated. A convenient change of measure will do the job. Precisely, define a new probability measure $\bar{P}$ (say) through the Radon-Nikodym derivative $d\bar{P}/\,d\hat{P}=e^{-\frac{1}{2}\sigma_0^2\tau+\sigma_0\hat{W}(\tau)}$. Under this new probability measure, the price-dividend ratio $p\,(s)$ satisfies,

$$p\,(s)=\int_0^\infty e^{g_0\tau}\cdot\bar{\mathbb{E}}\left[e^{-\int_0^\tau \mathrm{Disc}(s(u))du}\right| s\left.\right] d\tau, \tag{6.16}$$

and

$$ds(\tau)=\bar{\varphi}\,(s(\tau))\,d\tau+v(s\,(\tau))d\bar{W}(\tau),$$

where $\bar{W}\,(\tau)=\hat{W}\,(\tau)-\sigma_0\tau$ is a $\bar{P}$-Brownian motion, and $\bar{\varphi}\,(s)=\varphi\,(s)-v\,(s)\,\lambda\,(s)+\sigma_0 v\,(s)$.

The inner expectation in Eq. (6.16) comes in exactly the same format as in the canonical pricing problem of Section 6.5. Therefore, we are now ready to apply Proposition 6.1. We have,

1. *Suppose that risk-adjusted discount rates are countercyclical, viz* $\frac{d}{ds}\mathrm{Disc}(s)\leq 0$. *Then price-dividend ratios are procyclical, viz* $\frac{d}{ds}p\,(s)>0$.

2. *Suppose that price-dividend ratios are procyclical. Then price-dividend ratios are concave in s whenever risk-adjusted discount rates are convex in s, viz* $\frac{d^2}{ds^2}\mathrm{Disc}(s)>0$, *and* $\frac{d^2}{ds^2}\bar{\varphi}\,(s)\leq 2\frac{d}{ds}\mathrm{Disc}(s)$.

So we have found joint restrictions on the primitives such that the pricing function p is consistent with certain properties given in advance. What is the *economic interpretation* related to the convexity of risk-adjusted discount rates? If price-dividend ratios are concave in some state variable Y tracking the business cycle condititions, returns volatility increases on the downside, and it is thus countercyclical (see Figure 6.6.) According to the previous predictions, price-dividend ratios are concave in Y whenever risk-adjusted discount rates are decreasing and *sufficiently* convex in Y. The economic significance of convexity in this context is that in good times, risk-adjusted discount rates are substantially stable; consequently, the evaluation of future dividends does not vary too much, and price-dividend ratios are relatively stable. And in bad times risk-adjusted discount rates increase sharply, thus making price-dividend ratios more responsive to changes in the economic conditions.

Heuristically, the *mathematics* behind the previous results can be explained as follows. For small τ, Eq. (6.9) is,

$$p(s,\tau) \equiv \bar{\mathbb{E}}\left[e^{\int_0^\tau [g_0 - \text{Disc}(s(u))]du} \middle| s \right] \approx 1 + [g_0 - \text{Disc}(s)] \times \tau + h.o.t.$$

Hence convexity of Disc(s) translates to concavity of $p(s,\tau)$. But as pointed out earlier, the additional higher order terms matter too. The problem with these heuristic arguments is how well the approximation works for small τ. Furthermore $p(s,\tau)$ is not the price-dividend ratio. The price dividend ratio is instead $p(s) = \int_0^\infty p(s,\tau)\,d\tau$. Anyway the previous predictions confirm that the intuition is indeed valid.

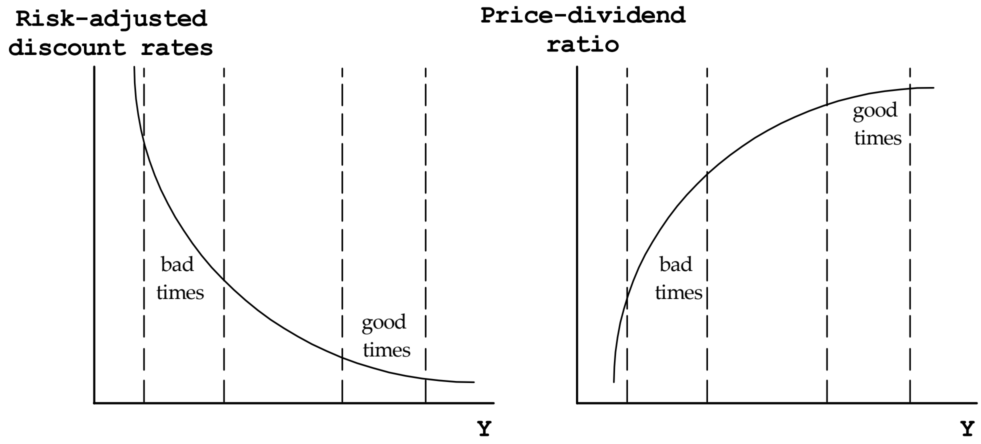


FIGURE 6.6. Countercyclical return volatility

What does empirical evidence suggest? To date no empirical work has been done on this. Here is a simple exploratory analysis. First it seems that real risk-adjusted discount rates Disc_t are convex in some very natural index summarizing the economic conditions (see Figure 6.7). In Table 6.2 , we also run Least Absolute Deviations (LAD) regressions to explore whether P/D dividend ratios are concave functions in IP.[15] And we run LAD regressions in correspondence of three sample periods to better understand the role of the exceptional (yet persistent) increase in the P/D ratio during the late 1990s. Figure 6.8 depicts scatter plots of data (along with fitted regressions) related to these three sampling periods.

[15]We run LAD regressions because this methodology is known to be more robust to the presence of outliers than Ordinary Least Squares.

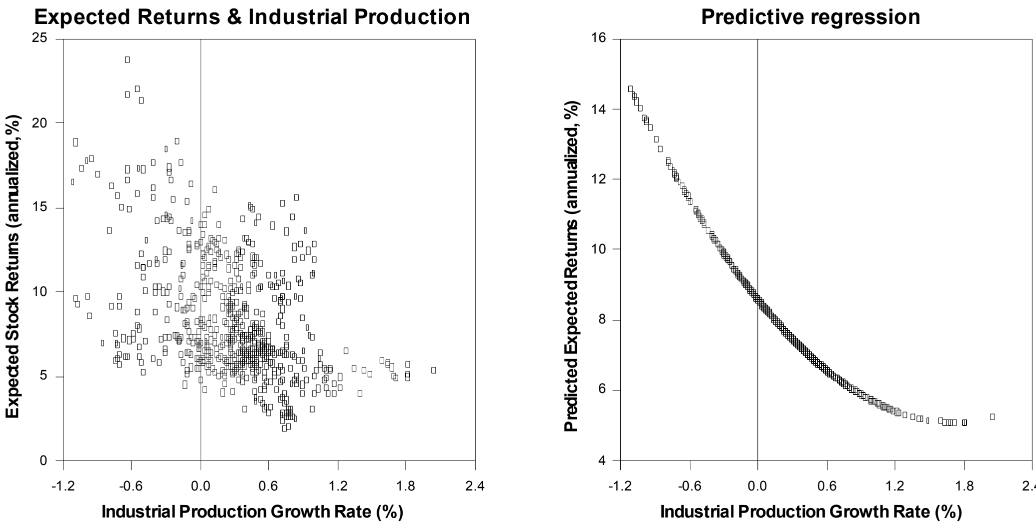


FIGURE 6.7.The left-hand side of this picture plots estimates of the expected returns (annualized, percent) ($\hat{\mathcal{E}}_t$ say) against one-year moving averages of the industrial production growth (IP_t). The expected returns are estimated through the predictive regression of S&P returns on to default-premium, term-premium and return volatility defined as $\widehat{\text{Vol}}_t \equiv \sqrt{\frac{\pi}{2}} \sum_{i=1}^{12} \frac{|\text{Exc}_{t+1-i}|}{\sqrt{12}}$, where Exc_t is the return in excess of the 1-month bill return as of month t. The one-year moving average of the industrial production growth is computed as $\text{IP}_t \equiv \frac{1}{12} \sum_{i=1}^{12} \text{Ind}_{t+1-i}$, where Ind_t is the real, seasonally adjusted industrial production growth as of month t. The right-hand side of this picture depicts the prediction of the static Least Absolute Deviations regression: $\hat{\mathcal{E}}_t = \underset{(0.15)}{8.56} - \underset{(0.30)}{4.05} \cdot \text{IP}_t + \underset{(0.31)}{1.18} \cdot \text{IP}_t^2 + w_t$, where w_t is a residual term, and standard errors are in parenthesis. Data are sampled monthly, and span the period from January 1948 to December 2002.

TABLE 6.2. Price-dividend ratios and economic conditions. Results of the LAD regression $\text{P/D} = a + b \cdot \text{IP} + c \cdot \text{IP}^2 + w$, where P/D is the S&P Comp. price-dividend ratio, $\text{IP}_t = (I_t + \cdots + I_{t-11}) / 12$; I_t is the real, seasonally adjusted US industrial production growth rate, and w_t is a residual term. Data are sampled monthly, and cover the period from January 1948 through December 2002.

	1948:01 - 1991:12		1948:01 - 1996:12		1948:01 - 2002:12	
	estimate	std dev	estimate	std dev	estimate	std dev
a	27.968	0.311	29.648	0.329	30.875	0.709
b	2.187	0.419	2.541	0.475	3.059	1.074
c	−2.428	0.429	−3.279	0.480	−3.615	1.091

6.7 Large price swings as a learning induced phenomenon

Consider a static scenario in which consumption z is generated by $z = \theta + w$, where θ and w are independently distributed, with $p \equiv \Pr(\theta = A) = 1 - \Pr(\theta = -A)$, and $\Pr(w = A) = \Pr(w = -A) = \frac{1}{2}$. Suppose that the "state" θ is unobserved. How would we update our prior probability p of the "good" state upon observation of z? A simple application of the Bayes' Theorem gives the posterior probabilities $\Pr(\theta = A|\, z_i)$ displayed in Table 6.3. Considered as a random variable defined over observable states z_i, the posterior probability $\Pr(\theta = A|\, z_i)$ has expectation $E\left[\Pr\left(\theta = A|\, z\right)\right] = p$ and variance $var\left[\Pr\left(\theta = A|\, z\right)\right] = \frac{1}{2}p(1-p)$. Clearly, this variance is zero exactly where there is a degenerate prior on the state. More generally, it is a $\cap$-shaped function of the a priori probability p of the good state. Since the "filter",

$$g \equiv E\left(\theta = A|\, z\right)$$

is linear in $\Pr\left(\theta = A|\, z\right)$, the same qualitative conclusions are also valid for g.

	z_i (observable state)		
	$z_1 = 2A$	$z_2 = 0$	$z_3 = -2A$
$\Pr(z_i)$	$\frac{1}{2}p$	$\frac{1}{2}$	$\frac{1}{2}(1-p)$
$\Pr\left(\theta = A|\, z = z_i\right)$	1	p	0

TABLE 6.3. Randomization of the posterior probabilities $\Pr\left(\theta = A|\, z\right)$.

To understand in detail how we computed the values in Table 6.3, let us recall Bayes' Theorem. Let $(E_i)_i$ be a partition of the state space Ω. (This partition can be finite or uncountable, i.e. the set of indexes i can be finite or uncountable - it really doesn't matter.) Then Bayes' Theorem says that,

$$\Pr\left(E_i|\, F\right) = \Pr\left(E_i\right) \cdot \frac{\Pr\left(F|\, E_i\right)}{\Pr\left(F\right)} = \Pr\left(E_i\right) \cdot \frac{\Pr\left(F|\, E_i\right)}{\sum_j \Pr\left(F|\, E_j\right)\Pr\left(E_j\right)}. \tag{6.17}$$

By applying Eq. (6.17) to our example,

$$\Pr\left(\theta = A|\, z = z_1\right) = \Pr\left(\theta = A\right)\frac{\Pr\left(z = z_1|\, \theta = A\right)}{\Pr\left(z = z_1\right)} = p\frac{\Pr\left(z = z_1|\, \theta = A\right)}{\Pr\left(z = z_1\right)}.$$

But $\Pr\left(z = z_1|\, \theta = A\right) = \Pr\left(w = z_1 - A\right) = \Pr\left(w = A\right) = \frac{1}{2}$. On the other hand, $\Pr\left(z = z_1\right) = \frac{1}{2}p$. This leaves $\Pr\left(\theta = A|\, z = z_1\right) = 1$. Naturally this is trivial, but one proceeds similarly to compute the other probabilities.

The previous example conveys the main ideas underlying nonlinear filtering. However, it leads to a nonlinear filter g which is somewhat distinct from the ones usually encountered in the literature.[16] In the literature, the instantaneous variance of the posterior probability changes $d\pi$ (say) is typically proportional to $\pi^2(1-\pi)^2$, not to $\pi(1-\pi)$. As we now heuristically demonstrate, the distinction is merely technical. Precisely, it is due to the assumption that w is a discrete random variable. Indeed, assume that w has zero mean and unit variance, and that

[16] See, e.g., Liptser and Shiryaev (2001a) (chapters 8 and 9).

it is absolutely continuous with *arbitrary* density function ϕ. Let $\pi(z) \equiv \Pr(\theta = A|\, z \in dz)$. By an application of the Bayesian "learning" mechanism in Eq. (6.17),

$$\pi(z) = \Pr(\theta = A) \cdot \frac{\Pr(z \in dz|\, \theta = A)}{\Pr(z \in dz|\, \theta = A)\Pr(\theta = A) + \Pr(z \in dz|\, \theta = -A)\Pr(\theta = -A)}.$$

But $\Pr(z \in dz|\, \theta = A) = \Pr(w = z - A) = \phi(z - A)$ and similarly, $\Pr(z \in dz|\, \theta = -A) = \Pr(w = z + A) = \phi(z + A)$. Simple computations then leave,

$$\pi(z) - p = p(1-p)\frac{\phi(z-A) - \phi(z+A)}{p\phi(z-A) + (1-p)\phi(z+A)}. \tag{6.18}$$

That is, the variance of the "probability changes" $\pi(z) - p$ is proportional to $p^2(1-p)^2$.

To add more structure to the problem, we now assume that w is Brownian motion and set $A \equiv A d\tau$. Let $z_0 \equiv z(0) = 0$. In appendix, we show that by an application of Itô's lemma to $\pi(z)$,

$$d\pi(\tau) = 2A \cdot \pi(\tau)(1 - \pi(\tau))dW(\tau), \quad \pi(z_0) \equiv p, \tag{6.19}$$

where $dW(\tau) \equiv dz(\tau) - g(\tau)d\tau$ and $g(\tau) \equiv E(\theta|\, z(\tau)) = [A\pi(\tau) - A(1 - \pi(\tau))]$. Naturally, this construction is heuristic. Nevertheless, the result is correct.[17] Importantly, it is possible to show that W is a Brownian motion with respect to the agents' information set $\sigma(z(t),\, t \leq \tau)$.[18] Therefore, the equilibrium in the original economy with incomplete information is isomorphic in its pricing implications to the equilibrium in a full information economy in which,

$$\begin{cases} dz(\tau) = [g(\tau) - \lambda(\tau)\sigma_0]\, d\tau + \sigma_0 d\hat{W}(\tau) \\ \\ dg(\tau) = -\lambda(\tau)v(g(\tau))d\tau + v(g(\tau))d\hat{W}(\tau) \end{cases} \tag{6.20}$$

where $\hat{W}$ is a Q-Brownian motion, λ is a risk-premium process, $v(g) \equiv (A - g)(g + A)/\sigma_0$ and $\sigma_0 \equiv 1$.[19] In fact, if the variance per unit of time of w is σ_0^2, eqs. (6.20) hold for any $\sigma_0 > 0$.[20] Furthermore, a similar result would hold had drift and diffusion of Z been assumed to be proportional to z. In this case, (Z, G) would be solution to

$$\begin{cases} \dfrac{dz(\tau)}{z(\tau)} = [g(\tau) - \sigma_0\lambda]\, d\tau + \sigma_0 d\hat{W}(\tau) \\ \\ dg(\tau) = \varphi(g(\tau))\, d\tau + v(g(\tau))\, d\hat{W}(\tau) \end{cases} \tag{6.21}$$

where $\varphi = -\lambda v$ and v is as before. In all cases, the instantaneous volatility of G is $\cap$-shaped. Under positive risk-aversion, this makes the risk-neutralized drift of Z a convex function of g. The economic implications of this result are very important, and will be analyzed with the help of Proposition 6.1.

System (6.20) is related to a model studied by Veronesi (1999). This model regards an infinite horizon economy in which a representative agent with CARA $= \gamma$. This agent observes realizations of Z generated by:

$$dz(\tau) = \theta d\tau + \sigma_0 dw_1(\tau), \tag{6.22}$$

[17] See, for example, Liptser and Shiryaev (2001*a*) (theorem 8.1 p. 318; and example 1 p. 371).

[18] See Liptser and Shiryaev (2001*a*) (theorem 7.12 p. 273).

[19] Such an isomorphic property has been pointed out for the first time by Veronesi (1999) in a related model.

[20] More precisely, we have $dW(\tau) = \sigma_0^{-1}\left(dz(\tau) - E\left(\theta|\, z(t)_{t \leq \tau}\right) d\tau\right) = \sigma_0^{-1}(dz(\tau) - g(\tau)\, d\tau)$.

where w_1 is a Brownian motion, and θ is a two-states $(\bar{\theta}, \underline{\theta})$ Markov chain. θ is unobserved, and the agent implements a Bayesian procedure to learn whether she lives in the "good" state $\bar{\theta} > \underline{\theta}$. All in all, such a Bayesian procedure is similar to the one in Eq. (6.17). Therefore, it would be relatively simple to show that all equilibrium prices in this economy are isomorphic to all equilibrium prices in an economy in which (Z, G) are solution to:

$$\begin{cases} dz(\tau) = \left[g(\tau) - \gamma\sigma_0^2\right] d\tau + \sigma_0 d\hat{W}(\tau) \\ \\ dg(\tau) = \left[k(\bar{g} - g(\tau)) - \gamma\sigma_0 v\left(g(\tau)\right)\right] d\tau + v\left(g(\tau)\right) d\hat{W}(\tau) \end{cases}$$

where $\hat{W}$ is a P^0-Brownian motion, $v(g) = (\bar{\theta} - g)(g - \underline{\theta})/\sigma_0$, $k, \bar{g}$ are some positive constants. Veronesi (1999) also assumed that the riskless asset is infinitely elastically supplied, and therefore that the interest rate r is a constant. It is instructive to examine the price implications of the resulting economy. In terms of the representation in Eq. (6.4), this model predicts that $q(z, g) = \int_0^\infty C(z, g, \tau) d\tau$, where

$$C(z, g, \tau) = e^{-r\tau}\left(z - \sigma_0\gamma\tau\right) + D\left(g, \tau\right), \text{ and } D\left(g, \tau\right) \equiv e^{-r\tau} \int_0^\tau \mathbb{E}\left[g(u)|\, g\right] du, \quad \tau \geq 0. \quad (6.23)$$

We may now apply Proposition 6.1 to study convexity properties of D. Precisely, function $\mathbb{E}\left[g(u)|\, g\right]$ is a special case of the auxiliary pricing function (6.7) (namely, for $\rho \equiv 1$ and $\psi(g) = g$). By Proposition 6.1-b), $\mathbb{E}\left[g(u)|\, g\right]$ is convex in g whenever the drift of G in (6.20) is convex. This condition is automatically guaranteed by $\gamma > 0$. Technically, Proposition 6.1 implies that the conditional expectation of a diffusion process inherits the very same second order properties (concavity, linearity, and convexity) of the drift function.

The economic implications of this result are striking. In this economy prices are convex in the expected dividend growth. This means that in good times, prices may well rocket to very high values with relatively small movements in the underlying fundamentals.

The economic interpretation of this convexity property is that risk-aversion correction is nil during extreme situations (i.e. when the dividend growth rate is at its boundaries), and it is the highest during relatively more "normal" situations. More formally, the risk-adjusted drift of g is $\hat{\varphi}\left(y\right) = \varphi\left(g\right) - \gamma\sigma_0 v\left(g\right)$, and it is convex in g because v is concave in g.

Finally, we examine model (6.21). Also, please notice that this model has been obtained as a result of a specific learning mechanism. Yet alternative learning mechanisms can lead to a model having the same structure, but with different coefficients φ and v. For example, a model related to Brennan and Xia (2001) information structure is one in which a single infinitely lived agent observes Z, where Z is solution to:

$$\frac{dz(\tau)}{z(\tau)} = \hat{g}(\tau) d\tau + \sigma_0 dw_1(\tau),$$

where $\hat{G} = \{\hat{g}(\tau)\}_{\tau > 0}$ is unobserved, but now it does not evolve on a countable number of states. Rather, it follows an Ornstein-Uhlenbeck process:

$$d\hat{g}(\tau) = k(\bar{g} - \hat{g}(\tau))d\tau + \sigma_1 dw_1(\tau) + \sigma_2 dw_2(\tau)$$

where $\bar{g}$, σ_1 and σ_2 are positive constants. Suppose now that the agent implements a learning procedure similar as before. If she has a Gaussian prior on $\hat{g}(0)$ with variance γ_*^2 (defined below), the nonarbitrage price takes the form $q(z,g)$, where (Z,G) are now solution to Eq. (6.6), with $m_0(z,g) = gz$, $\sigma(z) = \sigma_0 z$, $\varphi_0(z,g) = k(\bar{g} - g)$, $v_2 = 0$, and $v_1 \equiv v_1(\gamma_*) = (\sigma_1 + \frac{1}{\sigma_0}\gamma_*)^2$, where γ_* is the positive solution to $v_1(\gamma) = \sigma_1^2 + \sigma_2^2 - 2k\gamma$.[21]

Finally, models making expected consumption another *observed* diffusion may have an interest in their own (see for example, Campbell (2003) and Bansal and Yaron (2004)).

Now let's analyze these models. Once again, we may make use of Proposition 6.1. We need to set the problem in terms of the notation of the canonical pricing problem in section 6.5. To simplify the exposition, we suppose that λ is constant. By the same kind of reasoning leading to Eq. (6.16), one finds that the price-dividend ratio is independent of z here too, and is given by function p below,

$$p(g) = \int_0^\infty \bar{\mathbb{E}}\left[e^{\int_0^\tau [g(u) - r(g(u))]du - \sigma_0\lambda\tau} \Big| g \right] d\tau, \tag{6.24}$$

where,

$$\begin{cases} \dfrac{dz(\tau)}{z(\tau)} = [g(\tau) - \sigma_0\lambda]\, d\tau + \sigma_0 d\bar{W}(\tau) \\ \\ dg(\tau) = [\varphi(g(\tau)) + \sigma_0 v(g(\tau))]\, d\tau + v(g(\tau))\, d\bar{W}(\tau) \end{cases}$$

and $\bar{W}(\tau) = \hat{W}(\tau) - \sigma_0\tau$ is a $\bar{P}$-Brownian motion. Under regularity conditions, monotonicity and convexity properties are inherited by the inner expectation in Eq. (6.24). Precisely, in the notation of the canonical pricing problem,

$$\rho(g) = -g + R(g) + \sigma_0\lambda \quad \text{and} \quad b(g) = \varphi_0(g) + (\sigma_0 - \lambda)\, v(g),$$

where φ_0 is the physical probability measure. Therefore,

1. *The price-dividend ratio is increasing in the dividend growth rate whenever* $\frac{d}{dg}R(g) < 1$.

2. *Suppose that the price-dividend ratio is increasing in the dividend growth rate. Then it is convex whenever* $\frac{d^2}{dg^2}R(g) > 0$, *and* $\frac{d^2}{dg^2}[\varphi_0(g) + (\sigma_0 - \lambda)\, v(g)] \geq -2 + 2\frac{d}{dg}R(g)$.

For example, if the riskless asset is constant (because for example it is infinitely elastically supplied), then the price-dividend ratio is always increasing and it is convex whenever,

$$\frac{d^2}{dg^2}[\varphi_0(g) + (\sigma_0 - \lambda)\, v(g)] \geq -2.$$

The reader can now use these conditions to check predictions made by all models with stochastic dividend growth presented before.

[21] In their article, Brennan and Xia considered a slightly more general model in which consumption and dividends differ. They obtain a reduced-form model which is identical to the one in this example. In the calibrated model, Brennan and Xia found that the variance of the filtered $\hat{g}$ is higher than the variance of the expected dividend growth in an economy with complete information. The results on γ^* in this example can be obtained through an application of theorem 12.1 in Liptser and Shiryaev (2001) (Vol. II, p. 22). They generalize results in Gennotte (1986) and are a special case of results in Detemple (1986). Both Gennotte and Detemple did not emphasize the impact of learning on the pricing function.

6.8 Appendix 6.1

6.8.1 *Markov pricing kernels*

Let

$$\xi(\tau) \equiv \xi(z(\tau), y(\tau), \tau) = e^{-\int_0^\tau \delta(z(s),y(s))ds} \Upsilon(z(\tau), y(\tau)), \tag{6.25}$$

for some bounded positive function δ, and some positive function $\Upsilon(z, y) \in \mathcal{C}^{2,2}(\mathbb{Z}\times\mathbb{Y})$. By the assumed functional form for ξ, and Itô's lemma,

$$\begin{aligned} R(z, y) &= \delta(z, y) - \frac{L\Upsilon(z, y)}{\Upsilon(z, y)} \\ \lambda_1(z, y) &= -\sigma_0 z \frac{\partial}{\partial z} \log \Upsilon(z, y) - v_1(z, y) \frac{\partial}{\partial y} \log \Upsilon(z, y) \\ \lambda_2(z, y) &= -v_2(z, y) \frac{\partial}{\partial y} \log \Upsilon(z, y) \end{aligned}$$

Example A1 below is an important special case of this setting. Finally, to derive Eq. (6.3) in this setting, let us define the (undiscounted) "Arrow-Debreu adjusted" asset price process as:

$$w(z, y) \equiv \Upsilon(z, y) \cdot q(z, y).$$

By the results in section 6.4.2, we know that the following price representation holds true:

$$q(\tau)\xi(\tau) = E\left[\int_\tau^\infty \xi(s) z(s) ds\right], \quad \tau \geq 0.$$

Under usual regularity conditions, the previous equation can then be understood as the unique Feynman-Kac stochastic representation of the solution to the following partial differential equation

$$Lw(z, y) + f(z, y) = \delta(z, y) w(z, y), \quad \forall (z, y) \in \mathbb{Z} \times \mathbb{Y},$$

where $f \equiv \Upsilon z$. Eq. (6.3) then follows by the definition of $Lw(\tau) \equiv \frac{d}{ds} E\left[\Upsilon q\right]\big|_{s=\tau}$.

Example A1 (Infinite horizon, complete markets economy.) Consider an infinite horizon, complete markets economy in which total consumption Z is solution to Eq. (6.6), with $v_2 \equiv 0$. Let a (single) agent's program be:

$$\max E\left[\int_0^\infty e^{-\delta\tau} u(c(\tau), x(\tau)) d\tau\right] \text{ s.t. } V_0 = E\left[\int_0^\infty \xi(\tau) c(\tau) d\tau\right], \quad V_0 > 0,$$

where $\delta > 0$, the instantaneous utility u is continuous and thrice continuously differentiable in its arguments, and x is solution to

$$dx(\tau) = \beta(z(\tau), g(\tau), x(\tau)) d\tau + \gamma(z(\tau), g(\tau), x(\tau)) dW_1(\tau).$$

In equilibrium, $C = Z$, where C is optimal consumption. In terms of the representation in (6.25), we have that $\delta(z, x) = \delta$, and $\Upsilon(z(\tau), x(\tau)) = u_1(z(\tau), x(\tau)) / u_1(z(0), x(0))$. Consequently, $\lambda_2 = 0$,

$$\begin{aligned} R(z, g, x) = \delta &- \frac{u_{11}(z, x)}{u_1(z, x)} m_0(z, g) - \frac{u_{12}(z, x)}{u_1(z, x)} \beta(z, g, x) \\ &- \frac{1}{2}\sigma(z, g)^2 \frac{u_{111}(z, x)}{u_1(z, x)} - \frac{1}{2}\gamma(z, g, x)^2 \frac{u_{122}(z, x)}{u_1(z, x)} - \gamma(z, g, x)\sigma(z, g) \frac{u_{112}(z, x)}{u_1(z, x)} \end{aligned} \tag{6.26}$$

$$\lambda(z, g, x) = -\frac{u_{11}(z, x)}{u_1(z, x)} \sigma(z, g) - \frac{u_{12}(z, x)}{u_1(z, x)} \gamma(z, g, x). \tag{6.27}$$

6.8.2 The maximum principle

Suppose we are given the differential equation:

$$\frac{dx(\tau)}{d\tau} = \phi(\tau),\ \tau \in (t, T),$$

where ϕ satisfies some basic regularity conditions (essentially an integrability condition: see below). Suppose we know that

$$x(T) = 0,$$

and that

$$\text{sign}\,(\phi(\tau)) = \text{constant on } \tau \in (t, T).$$

We wish to determine the sign of $x(t)$. Under the previous assumptions on $x(T)$ and the sign of ϕ, we have that:

$$\text{sign}\,(x(t)) = -\ \text{sign}\,(\phi)\,.$$

The proof of this basic result can be grasped very simply from Figure 6.11, and it also follows easily analytically. We obviously have,

$$0 = x(T) = x(t) + \int_t^T \phi(\tau)d\tau \quad \Leftrightarrow \quad x(t) = -\int_t^T \phi(\tau)d\tau.$$

Next, suppose that,

$$\frac{dx(\tau)}{d\tau} = \phi(\tau),\ \tau \in (t, T),$$

where

$$x(\tau) = f\left(y(\tau), \tau\right),\ \tau \in (t, T),$$

and

$$\frac{dy(\tau)}{d\tau} = D(\tau),\ \tau \in (t, T).$$

With enough regularity conditions on ϕ, f, D, we have that

$$\frac{dx}{d\tau} = \left(\frac{\partial}{\partial \tau} + L\right) f,\ \tau \in (t, T),$$

where $Lf = \frac{\partial f}{\partial y} \cdot D$. Therefore,

$$\left(\frac{\partial}{\partial \tau} + L\right) f = \phi,\ \tau \in (t, T), \tag{6.28}$$

and the previous conclusions hold here as well: if $f(y, T) = 0\ \forall y$, and $\text{sign}(\phi(\tau)) = \text{constant}$ on $\tau \in (t, T)$, then,

$$\text{sign}\,(f(t)) = -\ \text{sign}\,(\phi)\,.$$

Again, this is so because

$$f(y(t), t) = -\int_t^T \phi(\tau)d\tau.$$

Such results can be extended in a straightforward manner in the case of stochastic differential equations. Consider the more elaborate operator-theoretic format version of (6.28) which typically emerges in many asset pricing problems with Brownian information:

$$0 = \left(\frac{\partial}{\partial \tau} + L - k\right) u + \zeta,\ \tau \in (t, T). \tag{6.29}$$

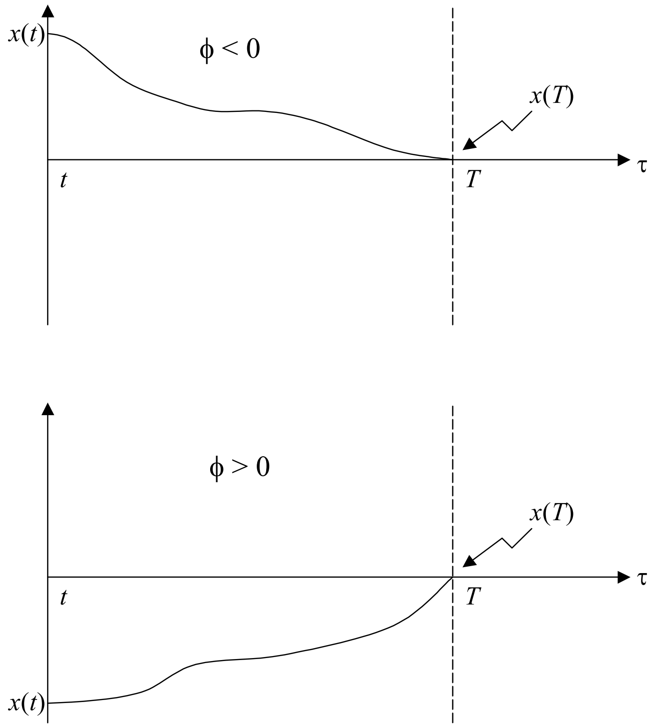


FIGURE 6.9. Illustration of the maximum principle for ordinary differential equations

Let

$$y(\tau) \equiv e^{-\int_t^\tau k(u)du} u(\tau) + \int_t^\tau e^{-\int_t^u k(s)ds} \zeta(u) du.$$

I claim that if (6.29) holds, then y is a martingale under some regularity conditions. Indeed,

$$\begin{aligned} dy(\tau) &= -k(\tau)e^{-\int_t^\tau k(u)du} u(\tau)d\tau + e^{-\int_t^\tau k(u)du} du(\tau) + e^{-\int_t^\tau k(u)du} \zeta(\tau)d\tau \\ &= -k(\tau)e^{-\int_t^\tau k(u)du} u(\tau) + e^{-\int_t^\tau k(u)du} \left[\left(\frac{\partial}{\partial \tau} + L \right) u(\tau) \right] d\tau + e^{-\int_t^\tau k(u)du} \zeta(\tau) d\tau \\ &\quad + \text{local martingale} \\ &= e^{-\int_t^\tau k(u)du} \left[-k(\tau)u(\tau) + \left(\frac{\partial}{\partial \tau} + L \right) u(\tau) + \zeta(\tau) \right] d\tau + \text{ local martingale} \\ &= \text{local martingale - because } \left(\frac{\partial}{\partial \tau} + L - k \right) u + \zeta = 0. \end{aligned}$$

Suppose that y is also a martingale. Then

$$y(t) = u(t) = E\left[y(T)\right] = E\left[e^{-\int_t^T k(u)du} u(T)\right] + E\left[\int_t^T e^{-\int_t^u k(s)ds} \zeta(u) du\right],$$

and starting from this relationship, you can adapt the previous reasoning on deterministic differential equations to the stochastic differential case. The case with jumps is entirely analogous.

6.8.3 Dynamic Stochastic Dominance

We have,

Proposition 6.A.1. (Dynamic Stochastic Dominance) *Consider two economies A and B with two fundamental volatilities a_A and a_B and let $\pi_i(x) \equiv a_i(x) \cdot \lambda^i(x)$ and $\rho_i(x)$ $(i = A, B)$ the corresponding risk-premium and discount rate. If $a_A > a_B$, the price c^A in economy A is lower than the price price c^B in economy B whenever for all $(x, \tau) \in \mathbb{R} \times [0, T]$,*

$$V(x,\tau) \equiv -\left[\rho_A(x) - \rho_B(x)\right] c^B(x,\tau) - \left[\pi_A(x) - \pi_B(x)\right] c^B_x(x,\tau) + \frac{1}{2}\left[a_A^2(x) - a_B^2(x)\right] c^B_{xx}(x,\tau) < 0. \tag{6.30}$$

If X is the price of a traded asset, $\pi_A = \pi_B$. If in addition ρ is constant, c is decreasing (increasing) in volatility whenever it is concave (convex) in x. This phenomenon is tightly related to the "convexity effect" discussed earlier. If X is not a traded risk, two additional effects are activated. The first one reflects a discounting adjustment, and is apparent through the first term in the definition of V. The second effect reflects risk-premia adjustments and corresponds to the second term in the definition of V. Both signs at which these two terms show up in Eq. (6.30) are intuitive.

6.8.4 Proofs

Proof of proposition 6.A.1. Function $c(x, T-s) \equiv \mathbb{E}[\exp(-\int_s^T \rho(x(t))dt) \cdot \psi(x(T)) \Big| x(s) = x]$ is solution to the following partial differential equation:

$$\begin{cases} 0 = -c_2(x, T-s) + L^* c(x, T-s) - \rho(x) c(x, T-s), & \forall (x,s) \in \mathbb{R} \times [0, T) \\ c(x, 0) = \psi(x), & \forall x \in \mathbb{R} \end{cases} \tag{6.31}$$

where $L^* c(x,u) = \frac{1}{2} a(x)^2 c_{xx}(x,u) + b(x) c_x(x,u)$ and subscripts denote partial derivatives. Clearly, c^A and c^B are both solutions to the partial differential equation (6.31), but with different coefficients. Let $b_A(x) \equiv b_0(x) - \pi_A(x)$. The price difference $\Delta c(x,\tau) \equiv c^A(x,\tau) - c^B(x,\tau)$ is solution to the following partial differential equation: $\forall (x,s) \in \mathbb{R} \times [0, T)$,

$$0 = -\Delta c_2(x, T-s) + \frac{1}{2}\sigma^B(x)^2 \Delta c_{xx}(x, T-s) + b_A(x) \Delta c_x(x, T-s) - \rho_A(x) \Delta c(x, T-s) + V(x, T-s),$$

with $\Delta c(x, 0) = 0$ for all $x \in \mathbb{R}$, and V is as in Eq. (6.30) of the proposition. The result follows by the maximum principle for partial differential equations. ■

Proof of proposition 6.1. By differentiating twice the partial differential equation (6.31) with respect to x, I find that $c^{(1)}(x,\tau) \equiv c_x(x,\tau)$ and $c^{(2)}(x,\tau) \equiv c_{xx}(x,\tau)$ are solutions to the following partial differential equations: $\forall (x,s) \in \mathbb{R}_{++} \times [0, T)$,

$$\begin{aligned} 0 &= -c_2^{(1)}(x, T-s) + \frac{1}{2} a(x)^2 c_{xx}^{(1)}(x, T-s) + [b(x) + \frac{1}{2}(a(x)^2)'] c_x^{(1)}(x, T-s) \\ &\quad - \left[\rho(x) - b'(x)\right] c^{(1)}(x, T-s) - \rho'(x) c(x, T-s), \end{aligned}$$

with $c^{(1)}(x, 0) = \psi'(x)$ $\forall x \in \mathbb{R}$; and $\forall (x,s) \in \mathbb{R} \times [0, T)$,

$$\begin{aligned} 0 &= -c_2^{(2)}(x, T-s) + \frac{1}{2} a(x)^2 c_{xx}^{(2)}(x, T-s) + [b(x) + (a(x)^2)'] c_x^{(2)}(x, T-s) \\ &\quad - \left[\rho(x) - 2b'(x) - \frac{1}{2}(a(x)^2)''\right] c^{(2)}(x, T-s) \\ &\quad - \left[2\rho'(x) - b''(x)\right] c^{(1)}(x, T-s) - \rho''(x) c(x, T-s), \end{aligned}$$

with $c^{(2)}(x, 0) = \psi''(x)$ $\forall x \in \mathbb{R}$. By the maximum principle for partial differential equations, $c^{(1)}(x, T-s) > 0$ (resp. < 0) $\forall (x,s) \in \mathbb{R} \times [0, T)$ whenever $\psi'(x) > 0$ (resp. < 0) and $\rho'(x) < 0$ (resp. > 0) $\forall x \in \mathbb{R}$. This completes the proof of part a) of the proposition. The proof of part b) is obtained similarly. ■

Derivation of Eq. (6.19). We have,

$$dz = gd\tau + dW,$$

and, by Eq. (6.18),

$$\pi(z) = \frac{p\phi(z-A)}{p\phi(z-A)+(1-p)\phi(z+A)} = \frac{pe}{p+(1-p)e^{-2Az}}$$

where the second equality follows by the Gaussian distribution assumption $\phi(x) \propto e^{-\frac{1}{2}x^2}$, and straight forward simplifications. By simple computations,

$$\frac{1-\pi(z)}{\pi(z)} = \frac{(1-p)\,e^{-2Az}}{p}; \quad \pi'(z) = 2A\pi(z)^2\frac{(1-p)\,e^{-2Az}}{p}; \quad \pi''(z) = 2A\pi'(z)\left[1-2\pi(z)\right]. \quad (6.32)$$

By construction,

$$g = \pi(z)\,A + [1-\pi(z)](-A) = A\left[2\pi(z)-1\right].$$

Therefore, by Itô's lemma,

$$d\pi = \pi' dz + \frac{1}{2}\pi'' d\tau = \pi' dz + A\pi'(1-2\pi)\,d\tau = \pi'\left[g + A(1-2\pi)\right]d\tau + \pi' dW = \pi' dW.$$

By using the relations in (6.32) once again,

$$d\pi = 2A\pi(1-\pi)\,dW. \qquad \blacksquare$$

6.8.5 On bond prices convexity

Consider a short-term rate process $\{r(\tau)\}_{\tau\in[0,T]}$ (say), and let $u(r_0,T)$ be the price of a bond expiring at time T when the current short-term rate is r_0:

$$u(r_0,T) = \mathbb{E}\left[\exp\left(-\int_0^T r(\tau)d\tau\right)\middle| r_0\right].$$

As pointed out in section 6, a *restricted version* of proposition 1-b) implies that in all scalar (diffusion) models of the short-term rate, $u_{11}(r_0,T) < 0$ whenever $b'' < 2$, where b is the risk-netraulized drift of r. This specific result was originally obtained in Mele (2003). Both the theory in Mele (2003) and the proof of proposition 1-b) rely on the Feynman-Kac representation of u_{11}. Here we provide a more intuitive derivation under a set of simplifying assumptions.

By Mele (2003) (Eq. (6) p. 685),

$$u_{11}(r_0,T) = \mathbb{E}\left\{\left[\left(\int_0^T \frac{\partial r}{\partial r_0}(\tau)d\tau\right)^2 - \int_0^T \frac{\partial^2 r}{\partial r_0^2}(\tau)d\tau\right]\exp\left(-\int_0^T r(\tau)d\tau\right)\right\}.$$

Hence $u_{11}(r_0,T) > 0$ whenever

$$\int_0^T \frac{\partial^2 r}{\partial r_0^2}(\tau)d\tau < \left(\int_0^T \frac{\partial r}{\partial r_0}(\tau)d\tau\right)^2. \quad (6.33)$$

To keep the presentation as simple as possible, we assume that r is solution to:

$$dr(\tau) = b(r(\tau))dt + a_0 r(\tau)dW(\tau),$$

where a_0 is a constant. We have,

$$\frac{\partial r}{\partial r_0}(\tau) = \exp\left[\int_0^\tau b'(r(u))du - \frac{1}{2}a_0^2\tau + a_0 W(\tau)\right]$$

and

$$\frac{\partial^2 r}{\partial r_0^2}(\tau) = \frac{\partial r}{\partial r_0}(\tau) \cdot \left[\int_0^\tau b''(r(u))\frac{\partial r(u)}{\partial r_0}du\right].$$

Therefore, if $b'' < 0$, then $\partial^2 r(\tau)/\partial r_0^2 < 0$, and by inequality (11.5), $u_{11} > 0$. But this result can considerably be improved. Precisely, suppose that $b'' < 2$ (instead of simply assuming that $b'' < 0$). By the previous equality,

$$\frac{\partial^2 r}{\partial r_0^2}(\tau) < 2 \cdot \frac{\partial r}{\partial r_0}(\tau) \cdot \left(\int_0^\tau \frac{\partial r(u)}{\partial r_0}du\right),$$

and consequently,

$$\int_0^T \frac{\partial^2 r}{\partial r_0^2}(\tau)d\tau < 2\int_0^T \frac{\partial r}{\partial r_0}(\tau) \cdot \left(\int_0^\tau \frac{\partial r(u)}{\partial r_0}du\right)d\tau = \left(\int_0^T \frac{\partial r(u)}{\partial r_0}du\right)^2,$$

which is inequality (11.5).

6.9 Appendix 6.2

In their original article, Campbell and Cochrane considered a discrete-time model in which consumption is a Gaussian process. The diffusion limit of their model is simply Eq. (7.10) given in the main text. By example A1 (Eq. (6.27)),

$$\lambda(z,x) = \frac{\eta}{s}\left[\sigma_0 - \frac{1}{z}\gamma(z,x)\right]. \tag{6.34}$$

To find the diffusion function γ of x, notice that $x = z(1-s)$, where s solution to Eq. (6.12). By Itô's lemma, then, $\gamma = [1 - s - sl(s)]\, z\sigma_0$. Finally, we replace this function into (6.34), and obtain $\lambda(s) = \eta\sigma_0 [1 + l(s)]$, as we claimed in the main text. (This result holds approximately in the original discrete time framework.) Finally, the real interest rate is found by an application of formula (6.26),

$$R(s) = \delta + \eta\left(g_0 - \frac{1}{2}\sigma_0^2\right) + \eta(1-\phi)(\bar{s} - \log s) - \frac{1}{2}\eta^2\sigma_0^2 [1 + l(s)]^2.$$

Campbell and Cochrane choose l so as to make the real interest rate constant. They took $l(s) = \bar{S}^{-1}\sqrt{1 + 2(\bar{s} - \log s)} - 1$, where $\bar{S} = \sigma_0\sqrt{\eta/(1-\phi)} = \exp(\bar{s})$, which leaves $R = \delta + \eta\left(g_0 - \frac{1}{2}\sigma_0^2\right) - \frac{1}{2}\eta(1-\phi)$.

6.10 Appendix 6.3: Simulation of discrete-time pricing models

The pricing equation is

$$q = E\left[m \cdot (q' + z')\right], \quad m = \beta\frac{u_c(z', x')}{u_c(z, x)} = \beta\left(\frac{s'}{s}\right)^{-\eta}\left(\frac{z'}{z}\right)^{-\eta}.$$

Hence, the price-dividend ratio $p \equiv q/z$ satisfies:

$$p = E\left[m\frac{z'}{z}(1 + p')\right], \quad \frac{z'}{z} = e^{g_0 + w}.$$

This is a functional equation having the form,

$$p(s) = E\left\{ g\left(s', s\right)\left[1 + p\left(s'\right)\right] \middle| \, s\right\}, \quad g\left(s', s\right) = \beta \left(\frac{s'}{s}\right)^{-\eta} \left(\frac{z'}{z}\right)^{1-\eta}.$$

A numerical solution can be implemented as follows. Create a grid and define $p_j = p\left(s_j\right)$, $j = 1, \cdots, N$, for some N. We have,

$$\begin{bmatrix} p_1 \\ \vdots \\ p_N \end{bmatrix} = \begin{bmatrix} b_1 \\ \vdots \\ b_N \end{bmatrix} + \begin{bmatrix} a_{11} & \cdots & a_{N1} \\ \vdots & \ddots & \vdots \\ a_{1N} & \cdots & a_{NN} \end{bmatrix} \begin{bmatrix} p_1 \\ \vdots \\ p_N \end{bmatrix},$$

$$b_i = \sum_{j=1}^{N} a_{ji}, \quad a_{ji} = g_{ji} \cdot p_{ji}, \quad g_{ji} = g\left(s_j, s_i\right), \quad p_{ji} = \Pr\left(s_j \middle| \, s_i\right) \cdot \Delta s,$$

where Δs is the integration step; $s_1 = s_{\min}$, $s_N = s_{\max}$; $s_{\min}$ and $s_{\max}$ are the boundaries in the approximation; and $\Pr\left(s_j \middle| \, s_i\right)$ is the transition density from state i to state j - in this case, a Gaussian transition density. Let $p = [p_1 \ \cdots \ p_N]^\top$, $b = [b_1 \ \cdots \ b_N]^\top$, and let A be a matrix with elements a_{ji}. The solution is,

$$p = (I - A)^{-1} b. \tag{6.35}$$

The model can be simulated in the following manner. Let $\underline{s}$ and $\bar{s}$ be the boundaries of the underlying state process. Fix $\Delta s = \dfrac{\bar{s} - \underline{s}}{N}$. Draw states. State s^* is drawn. Then,

1. If $\min\left(s^* - \underline{s}, \bar{s} - s^*\right) = s^* - \underline{s}$, let k be the smallest integer close to $\frac{s^* - \underline{s}}{\Delta s}$. Let $s_{\min} = s^* - k\Delta s$, and $s_{\max} = s_{\min} + N \cdot \Delta s$.

2. If $\min\left(s^* - \underline{s}, \bar{s} - s^*\right) = \bar{s} - s^*$, let k be the biggest integer close to $\frac{s^* - \underline{s}}{\Delta s}$. Let $s_{\max} = s^* + k\Delta s$, and $s_{\min} = s_{\max} - N \cdot \Delta s$.

The previous algorithm avoids interpolations. Importantly, it ensures that during the simulations, p is computed in correspondence of exactly the state s^* that is drawn. Precisely, once s^* is drawn, *1*) create the corresponding grid $s_1 = s_{\min}$, $s_2 = s_{\min} + \Delta s, \cdots, s_N = s_{\max}$ according to the previous rules; *2*) compute the solution from Eq. (6.35). In this way, one has $p\left(s^*\right)$ at hand - the simulated P/D ratio when state s^* is drawn.

References

Abel, A. B. (1988). "Stock Prices under Time-Varying Dividend Risk: an Exact Solution in an Infinite-Horizon General Equilibrium Model," *Journal of Monetary Economics*: 22, 375-393.

Bajeux-Besnainou, I. and J.-C. Rochet (1996). "Dynamic Spanning: Are Options an Appropriate Instrument?," *Mathematical Finance*: 6, 1-16.

Bansal, R. and A. Yaron (2004). "Risks for the Long Run: A Potential Resolution of Asset Pricing Puzzles," *Journal of Finance*: 59, 1481-1509.

Barberis, Huang and Santos (2001).

Barsky, R. B. (1989). "Why Don't the Prices of Stocks and Bonds Move Together?" *American Economic Review*: 79, 1132-1145.

Barsky, R. B. and J. B. De Long (1990). "Bull and Bear Markets in the Twentieth Century," *Journal of Economic History*: 50, 265-281.

Barsky, R. B. and J. B. De Long (1993). "Why Does the Stock Market Fluctuate?" *Quarterly Journal of Economics*: 108, 291-311.

Bergman, Y. Z., B. D. Grundy, and Z. Wiener (1996). "General Properties of Option Prices," *Journal of Finance*: 51, 1573-1610.

Black, F. and M. Scholes (1973). "The Pricing of Options and Corporate Liabilities," *Journal of Political Economy*: 81, 637-659.

Brennan, M. J. and Y. Xia (2001). "Stock Price Volatility and Equity Premium," *Journal of Monetary Economics*: 47, 249-283.

Campbell, J. Y. (2003). "Consumption-Based Asset Pricing," in Constantinides, G.M., M. Harris and R. M. Stulz (Editors): *Handbook of the Economics of Finance* (Volume 1B: chapter 13), 803-887.

Campbell, J. Y., and J. H. Cochrane (1999). "By Force of Habit: A Consumption-Based Explanation of Aggregate Stock Market Behavior," *Journal of Political Economy*: 107, 205-251.

David, A. (1997). "Fluctuating Confidence in Stock Markets: Implications for Returns and Volatility," *Journal of Financial and Quantitative Analysis*: 32, 427-462.

Detemple, J. B. (1986). "Asset Pricing in a Production Economy with Incomplete Information," *Journal of Finance*: 41, 383-391.

El Karoui, N., M. Jeanblanc-Picqué and S. E. Shreve (1998). "Robustness of the Black and Scholes Formula," *Mathematical Finance*: 8, 93-126.

Fama, E. F. and K. R. French (1989). "Business Conditions and Expected Returns on Stocks and Bonds," *Journal of Financial Economics*: 25, 23-49.

Gennotte, G. (1986). "Optimal Portfolio Choice Under Incomplete Information," *Journal of Finance*: 41, 733-746.

Huang, C.-F. and Pagès, H. (1992). "Optimal Consumption and Portfolio Policies with an Infinite Horizon: Existence and Convergence," *Annals of Applied Probability*: 2, 36-64.

Hajek, B. (1985). "Mean Stochastic Comparison of Diffusions," *Zeitschrift fur Wahrscheinlichkeitstheorie und Verwandte Gebiete*: 68, 315-329.

Jagannathan, R. (1984). "Call Options and the Risk of Underlying Securities," *Journal of Financial Economics*: 13, 425-434.

Kijima, M. (2002). "Monotonicity and Convexity of Option Prices Revisited," *Mathematical Finance*: 12, 411-426.

Liptser, R. S. and A. N. Shiryaev (2001). *Statistics of Random Processes*: (Second Edition), Berlin, Springer-Verlag. [2001*a*: *Vol.* I (*General Theory*). 2001*b*: *Vol.* II (*Applications*).]

Malkiel, B. (1979). "The Capital Formation Problem in the United States," *Journal of Finance*: 34, 291-306.

Mele, A. (2003). "Fundamental Properties of Bond Prices in Models of the Short-Term Rate," *Review of Financial Studies*: 16, 679-716.

Mele, A. (2005). "Rational Stock-Market Fluctuations," WP FMG-LSE.

Pindyck, R. (1984). "Risk, Inflation and the Stock Market," *American Economic Review*: 74, 335-351.

Poterba, J. and L. Summers (1985). "The Persistence of Volatility and Stock-Market Fluctuations," *American Economic Review*: 75, 1142-1151.

Romano, M. and N. Touzi (1997). "Contingent Claims and Market Completeness in a Stochastic Volatility Model," *Mathematical Finance*: 7, 399-412.

Rothschild, M. and J. E. Stiglitz (1970). "Increasing Risk: I. A Definition," *Journal of Economic Theory*: 2, 225-243.

Veronesi, P. (1999). "Stock Market Overreaction to Bad News in Good Times: A Rational Expectations Equilibrium Model," *Review of Financial Studies*: 12, 975-1007.

Veronesi, P. (2000). "How Does Information Quality Affect Stock Returns?" *Journal of Finance*: 55, 807-837.

Volkanov

Wang, S. (1993). "The Integrability Problem of Asset Prices," *Journal of Economic Theory*: 59, 199-213.

7
Tackling the puzzles

7.1 Non-expected utility

The standard intertemporal additively separable utility function confounds intertemporal substitution effects from attitudes towards risk. This fact is problematic. Epstein and Zin (1989, 1991) and Weil (1989) consider a class of recursive, but not necessarily expected utility, preferences. In this section, we present the details of this approach, without insisting on the theoretic underpinnings which the reader will find in Epstein and Zin (1989). We provide a basic definition and derivation of this class of preferences, a heuristic analysis of the resulting asset pricing properties, and finally some continuous time extensions first considered by Duffie and Epstein (1992*a*, *b*).

7.1.1 The recursive formulation

Let utility as of time t be v_t. We have,

$$v_t = W(c_t, \hat{v}_{t+1}),$$

where W is the "aggregator function" and $\hat{v}_{t+1}$ is the certainty-equivalent utility at $t+1$ defined as,

$$h(\hat{v}_{t+1}) = E[h(v_{t+1})],$$

where h is a von Neumann - Morgenstern utility function. That is, the certainty equivalent depends on some agent's risk-attitudes encoded in h. Therefore,

$$v_t = W\left(c_t, h^{-1}[E(h(v_{t+1}))]\right).$$

The analytical example used in the asset pricing literature is,

$$W(c, \hat{v}) = \left(c^\rho + e^{-\delta}\hat{v}^\rho\right)^{1/\rho} \quad \text{and} \quad h(\hat{v}) = \hat{v}^{1-\eta},$$

for three positive constants ρ, η and δ. In this formulation, risk-attitudes for static wealth gambles have still the classical CRRA flavor. More precisely, we say that η is the RRA for static wealth gambles and $(1-\rho)^{-1}$ is the IES.

We have,

$$\hat{v}_{t+1} = h^{-1}\left[E\left(h\left(v_{t+1}\right)\right)\right] = h^{-1}\left[E\left(v_{t+1}^{1-\eta}\right)\right] = \left[E\left(v_{t+1}^{1-\eta}\right)\right]^{\frac{1}{1-\eta}}.$$

The previous parametrization of the aggregator function then implies that,

$$v_t = \left[c_t^\rho + e^{-\delta}\left(E(v_{t+1}^{1-\eta})\right)^{\frac{\rho}{1-\eta}}\right]^{1/\rho}. \tag{7.1}$$

This collapses to the standard intertemporal additively separable case when $\rho = 1-\eta \Leftrightarrow$ RRA $=$ IES^{-1}. Indeed, it is straight forward to show that in this case,

$$v_t = \left[E\left(\sum_{n=0}^{\infty} e^{-\delta n} c_{t+n}^{1-\eta}\right)\right]^{\frac{1}{1-\eta}}.$$

Let us go back to Eq. (7.1). Function $V = v^{1-\eta}/(1-\eta)$ is obviously ordinally equivalent to v, and satisfies,

$$V_t = \frac{1}{1-\eta}\left[c_t^\rho + e^{-\delta}\left((1-\eta)\,E(V_{t+1})\right)^{\frac{\rho}{1-\eta}}\right]^{\frac{1-\eta}{\rho}}. \tag{7.2}$$

The previous formulation makes even more transparent that these utils collapse to standard intertemporal additive utils as soon as RRA $=$ IES^{-1}.

7.1.2 *Optimality*

Let us define cum-dividend wealth as $x_t \equiv \sum_{i=1}^m \left(p_{it} + D_{it}\right)\theta_{it}$. In the Appendix, we show that wealth x_t accumulates according to,

$$x_{t+1} = (x_t - c_t)\,\omega_t^\top\left(\mathbf{1}_m + \mathbf{r}_{t+1}\right) \equiv (x_t - c_t)\left(1 + r_{M,t+1}\right), \tag{7.3}$$

where ω is the vector of proportions of wealth invested in the m assets. Let us consider a Markov economy in which the underlying state is some process y. We consider stationary consumption and investment plans. Accordingly, let the stationary util be a function $v(x,y)$ when current wealth is x and the state is y. By Eq. (7.2),

$$V(x,y) = \max_{c,\omega} \mathcal{W}\left(c, E\left(V\left(x',y'\right)\right)\right) \equiv \frac{1}{1-\eta}\max_{c,\omega}\left\{c^\rho + e^{-\delta}\left[(1-\eta)\,E(V(x',y'))\right]^{\frac{\rho}{1-\eta}}\right\}^{\frac{1-\eta}{\rho}} \tag{7.4}$$

The first order condition for c yields,

$$\mathcal{W}_1\left(c, E\left(V\left(x',y'\right)\right)\right) = \mathcal{W}_2\left(c, E\left(V\left(x',y'\right)\right)\right)\cdot E\left[V_1\left(x',y'\right)\left(1 + r_M\left(y'\right)\right)\right], \tag{7.5}$$

where subscripts denote partial derivatives. Thus, optimal consumption is some function $c(x,y)$. Hence,

$$x' = (x - c(x,y))\left(1 + r_M\left(y'\right)\right)$$

We have,

$$V(x,y) = \mathcal{W}\left(c(x,y), E\left(V\left(x',y'\right)\right)\right).$$

By differentiating the value function with respect to x,

$$\begin{aligned} V_1(x,y) &= \mathcal{W}_1\left(c(x,y), E\left(V\left(x',y'\right)\right)\right) c_1(x,y) \\ &\quad + \mathcal{W}_2\left(c(x,y), E\left(V\left(x',y'\right)\right)\right) E\left[V_1\left(x',y'\right)\left(1 + r_M\left(y'\right)\right)\right]\left(1 - c_1(x,y)\right), \end{aligned}$$

where subscripts denote partial derivatives. By replacing Eq. (7.5) into the previous equation we get the Envelope Equation for this dynamic programming problem,

$$V_1(x,y)=\mathcal{W}_1\left(c(x,y),E\left(V\left(x',y'\right)\right)\right). \tag{7.6}$$

By replacing Eq. (7.6) into Eq. (7.5), and rearranging terms,

$$E\left[\frac{\mathcal{W}_2(c(x,y),\nu(x,y))}{\mathcal{W}_1(c(x,y),\nu(x,y))}\mathcal{W}_1\left(c\left(x',y'\right),\nu\left(x',y'\right)\right)\left(1+r_M\left(y'\right)\right)\right]=1,\quad \nu(x,y)\equiv E\left(V\left(x',y'\right)\right). \tag{7.7}$$

In the Appendix, we show that by a similar argument the same Euler equation applies to any asset i,

$$E\left[\frac{\mathcal{W}_2(c(x,y),\nu(x,y))}{\mathcal{W}_1(c(x,y),\nu(x,y))}\mathcal{W}_1\left(c\left(x',y'\right),\nu\left(x',y'\right)\right)\left(1+r_i\left(y'\right)\right)\right]=1,\quad i=1,\cdots,m. \tag{7.8}$$

7.1.3 Testable restrictions

To make Eq. (7.8) operational, we need to compute explicitly the stochastic discount factor,

$$m\left(x,y;x'y'\right)=\frac{\mathcal{W}_2(c(x,y),\nu(x,y))}{\mathcal{W}_1(c(x,y),\nu(x,y))}\mathcal{W}_1\left(c\left(x',y'\right),\nu\left(x',y'\right)\right).$$

In the appendix, we show that

$$m\left(x,y;x'y'\right)=\left[e^{-\delta}\left(\frac{c\left(x',y'\right)}{c(x,y)}\right)^{\rho-1}\right]^{\frac{1-\eta}{\rho}}\left(1+r_M\left(y'\right)\right)^{\frac{1-\eta}{\rho}-1}. \tag{7.9}$$

Therefore,

$$E\left\{\left[e^{-\delta}\left(\frac{c\left(x',y'\right)}{c(x,y)}\right)^{\rho-1}\right]^{\frac{1-\eta}{\rho}}\left(1+r_M\left(y'\right)\right)^{\frac{1-\eta}{\rho}-1}\left(1+r_i\left(y'\right)\right)\right\}=1,\quad i=1,\cdots,m.$$

7.1.4 Some examples

Models for long-run risks.

7.1.5 Continuous time extensions

Let us derive heuristically the continuous time limit of the previous utils. We have,

$$v_t=\left(c_t^{\rho}\Delta t+e^{-\delta\Delta t}\left(E(v_{t+\Delta t}^{1-\eta})\right)^{\frac{\rho}{1-\eta}}\right)^{1/\rho}.$$

[...]

Continuation utility v_t solves the stochastic differential equation,

$$\begin{cases} dv_t=\left[-f\left(c_t,v_t\right)-\frac{1}{2}A\left(v_t\right)\left\|\sigma_{vt}\right\|^2\right]dt+\sigma_{vt}dB_t \\ v_T=0\end{cases}$$

Here (f, A) is the aggregator. A is a variance multiplier - it places a penalty proportional to utility volatility $\|\sigma_{vt}\|^2$. (f, A) somehow corresponds to $(W, \hat{v})$ in the discrete time case.

Solution to the previous "stochastic differential utility" is,

$$v_t = E\left\{\int_t^T \left[f(c_s, v_s) + \frac{1}{2}A(v_s)\|\sigma_{vs}\|^2\right]\right\}.$$

The standard additive utility case is obtained by setting,

$$f(c, v) = u(c) - \beta v \quad \text{and} \quad A = 0.$$

7.2 "Catching up with the Joneses" in a heterogeneous agents economy

Chan and Kogan (2002) study an economy with heterogeneous agents and "catching up with the Joneses" preferences. In this economy, there is a continuum of agents indexed by a parameter $\eta \in [1, \infty)$ of the instantaneous utility function,

$$u_\eta(c, x) = \frac{\left(\frac{c}{x}\right)^{1-\eta}}{1-\eta},$$

where c is consumption, and x is the "standard living of others", to be defined below.

The total endowment in the economy, D, follows a geometric Brownian motion,

$$\frac{dD(\tau)}{D(\tau)} = g_0 d\tau + \sigma_0 dW(\tau). \tag{7.10}$$

By assumption, then, the standard of living of others, $x(\tau)$, is a weighted geometric average of the past realizations of the aggregate consumption D, viz

$$\log x(\tau) = \log x(0)\, e^{-\theta\tau} + \theta \int_0^\tau e^{-\theta(\tau-s)} \log D(s)\, ds, \quad \text{with } \theta > 0.$$

Therefore, $x(\tau)$ satisfies,

$$dx(\tau) = \theta s(\tau)\, x(\tau)\, d\tau, \quad \text{where } s(\tau) \equiv \log\left(D(\tau)/x(\tau)\right). \tag{7.11}$$

By Eqs. (7.10) and (7.11), $s(\tau)$ is solution to,

$$ds(\tau) = \left[g_0 - \frac{1}{2}\sigma_0^2 - \theta s(\tau)\right] d\tau + \sigma_0 dW(\tau).$$

In this economy with complete markets, the equilibrium price process is the same as the price process in an economy with a representative agent with the following utility function,

$$u(D, x) \equiv \max_{c_\eta} \int_1^\infty u_\eta(c_\eta, x) f(\eta)\, d\eta \quad \text{s.t.} \int_1^\infty c_\eta d\eta = D, \tag{P1}$$

where $f(\eta)^{-1}$ is the marginal utility of income of the agent η. (See Chapter 2 in Part I, for the theoretical foundations of this program.)

In the appendix, we show that the solution to the static program [P1] leads to the following expression for the utility function $u(D, x)$,

$$u(D,x) = \int_1^\infty \frac{1}{1-\eta} f(\eta)^{\frac{1}{\eta}} V(s)^{\frac{\eta-1}{\eta}} d\eta,$$

where V is a Lagrange multiplier, which satisfies,

$$e^s = \int_1^\infty f(\eta)^{\frac{1}{\eta}} V(s)^{-\frac{1}{\eta}} d\eta.$$

The appendix also shows that the unit risk-premium predicted by this model is,

$$\lambda(s) = \sigma_0 \frac{\exp(s)}{\int_1^\infty \frac{1}{\eta} f(\eta)^{\frac{1}{\eta}} V(s)^{-\frac{1}{\eta}} d\eta}. \tag{7.12}$$

This economy collapses to an otherwise identical homogeneous economy if the social weighting function $f(\eta) = \delta(\eta - \eta^0)$, the Dirac's mass at η^0. In this case, $\lambda(s) = \sigma_0 \eta^0$, a constant.

A crucial assumption in this model is that the standard of living X is a process with *bounded* variation (see Eq. (7.11)). By this assumption, the standard living of others is not a risk which agents require to be compensated for. The unit risk-premium in Eq. (7.12) is driven by s through nonlinearities induced by agents heterogeneity. By calibrating their model to US data, Chan and Kogan find that the risk-premium, $\lambda(s)$, is decreasing and convex in s.[1] The mechanism at the heart of this result is an endogenous wealth redistribution in the economy. Clearly, the less risk-averse individuals put a higher proportion of their wealth in the risky assets, compared to the more risk-avers agents. In the poor states of the world, stock prices decrease, the wealth of the less risk-averse lowers more than that of the more risk-averse agents, which reduces the fraction of wealth held by the less risk-averse individuals in the whole economy. Thus, in bad times, the contribution of these less risk-averse individuals to aggregate risk-aversion decreases and, hence, the aggregate risk-aversion increases in the economy.

7.3 Incomplete markets

7.4 Limited stock market participation

Basak and Cuoco (1998) consider a model with two agents. One of these agents does not invest in the stock-market, and has logarithmic instantaneous utility, $u_n(c) = \log c$. From his perspective, markets are incomplete. The second agent, instead, invests in the stock market, and has an instantaneous utility equal to $u_p(c) = (c^{1-\eta} - 1)/(1-\eta)$. Both agents are infinitely lived.

Clearly, in this economy, the competitive equilibrium is Pareto inefficient. Yet, Basak and Cuoco show how aggregation obtains in this economy. Let $\hat{c}_i(\tau)$ be the general equilibrium allocation of the agent i, $i = p, n$. The first order conditions of the two agents are,

$$u_p'(\hat{c}_p(\tau)) = w_p e^{\delta\tau} \xi(\tau); \quad \hat{c}_n(\tau)^{-1} = w_n e^{\delta\tau - \int_0^\tau R(s)ds} \tag{7.13}$$

[1] Their numerical results also revealed that in their model, the *log* of the price-dividend ratio is increasing and *concave* in s. Finally, their lemma 5 (p. 1281) establishes that in a homogeneous economy, the price-dividend ratio is increasing and *convex* in s.

where w_p, w_n are two constants, and ξ is the usual pricing kernel process, solution to,

$$\frac{d\xi(\tau)}{\xi(\tau)} = -R(\tau)\,dt - \lambda(\tau)\cdot dW. \tag{7.14}$$

Let

$$u(D,x) \equiv \max_{c_p+c_n=D}\left[u_p(c_p) + x\cdot u_n(c_n)\right],$$

where

$$x \equiv \frac{u_p'(\hat{c}_p)}{u_n'(\hat{c}_n)} = u_p'(\hat{c}_p)\hat{c}_n \tag{7.15}$$

is a *stochastic social weight.* By the definition of ξ, $x(\tau)$ is solution to,

$$dx(\tau) = -x(\tau)\,\lambda(\tau)\,dW(\tau), \tag{7.16}$$

where λ is the unit risk-premium, which as shown in the appendix equals,

$$\lambda(s) = \eta\sigma_0 s^{-1}, \quad \text{where } s(\tau) \equiv \hat{c}_p(\tau)/\,D(\tau).$$

Then, the equilibrium price system in this economy is supported by a fictitious representative agent with utility $u(D,x)$. Intuitively, the representative agent "allocations" satisfy, by construction,

$$\frac{u_p'(c_p^*(\tau))}{u_n'(c_n^*(\tau))} = \frac{u_p'(\hat{c}_p(\tau))}{u_n'(\hat{c}_n(\tau))} = x(\tau),$$

where starred allocations are the representative agent's "allocations". In other words, the trick underlying this approach is to find a stochastic social weight process $x(\tau)$ such that the first order conditions of the representative agent leads to the market allocations. This is shown more rigorously in the Appendix.

Guvenen (2005) makes an interesting extension of the Basak and Cuoco model. He consider two agents in which only the "rich" invests in the stock-market, and is such that $\text{ISE}_{\text{rich}} > \text{IES}_{\text{poor}}$. He shows that for the rich, a low IES is needed to match the equity premium. However, US data show that the rich have a high IES, which can not do the equity premium. A natural extension of this model is one in which we can disentangle IES and CRRA for the rich. Another issue is that in Guvenen, the market participant is really not the "rich", since he has the same wage schedule as the "poor".

7.5 Appendix on non-expected utility

7.5.1 Derivation of selected relations in the main text

DERIVATION OF EQ. (7.3). We have,

$$\begin{aligned} x_{t+1} &= \sum (p_{it+1} + D_{it+1})\,\theta_{it+1} \\ &= \sum (p_{it+1} + D_{it+1} - p_{it})\,\theta_{it+1} + \sum p_{it}\theta_{it+1} \\ &= \left(1 + \sum \frac{p_{it+1} + D_{it+1} - p_{it}}{p_{it}} \frac{p_{it}\theta_{it+1}}{\sum p_{it}\theta_{it+1}}\right) \sum p_{it}\theta_{it+1} \\ &= \left(1 + \sum r_{it+1}\omega_{it}\right)(x_t - c_t) \end{aligned}$$

where the last line follows by the standard budget constraint $c_t + \sum p_{it}\theta_{it+1} = x_t$, the definition $r_{it} \equiv \frac{p_{it+1}+D_{it+1}-p_{it}}{p_{it}}$, and the definition,

$$\omega_{it} \equiv \frac{p_{it}\theta_{it+1}}{\sum p_{it}\theta_{it+1}}. \qquad \blacksquare$$

DERIVATION OF EQ. (7.8). We have,

$$V(x,y) = \max_{c,\omega} \mathcal{W}\left(c, E\left(V\left(x', y'\right)\right)\right) = \max_{\theta'} \mathcal{W}\left(x - \sum p_i\theta_i', E\left(V\left(x', y'\right)\right)\right); \quad x' = \sum \left(p_i' + D_i'\right)\theta_i'.$$

The set of first order conditions is,

$$\theta_i' : 0 = -\mathcal{W}_1(\cdot)\,p_i + \mathcal{W}_2(\cdot)\,E\left[V_1\left(x', y'\right)\left(p_i' + D_i'\right)\right], \quad i = 1, \cdots, m.$$

Optimal consumption is $c(x,y)$. Let $\nu(x,y) \equiv E(V(x', y'))$, as in the main text. By replacing Eq. (7.6) into the previous equation,

$$E\left[\frac{\mathcal{W}_2(c(x,y), \nu(x,y))}{\mathcal{W}_1(c(x,y), \nu(x,y))}\mathcal{W}_1\left(c\left(x', y'\right), \nu\left(x', y'\right)\right)\frac{p_i' + D_i'}{p_i}\right] = 1, \quad i = 1, \cdots, m. \qquad \blacksquare$$

DERIVATION OF EQ. (7.9). We have,

$$\mathcal{W}(c, \nu) = \frac{1}{1-\eta}\left\{c^\rho + e^{-\delta}\left[(1-\eta)\,\nu\right]^{\frac{\rho}{1-\eta}}\right\}^{\frac{1-\eta}{\rho}}.$$

From this it follows that,

$$\begin{aligned} \mathcal{W}_1(c,\nu) &= \left\{c^\rho + e^{-\delta}\left[(1-\eta)\,\nu\right]^{\frac{\rho}{1-\eta}}\right\}^{\frac{1-\eta}{\rho}-1} c^{\rho-1} \\ \mathcal{W}_2(c,\nu) &= \left\{c^\rho + e^{-\delta}\left[(1-\eta)\,\nu\right]^{\frac{\rho}{1-\eta}}\right\}^{\frac{1-\eta}{\rho}-1} e^{-\delta}\left[(1-\eta)\,\nu\right]^{\frac{\rho}{1-\eta}-1} \end{aligned}$$

and,

$$\mathcal{W}_1\left(c', \nu'\right) = \left\{c'^\rho + e^{-\delta}\left[(1-\eta)\,\nu'\right]^{\frac{\rho}{1-\eta}}\right\}^{\frac{1-\eta}{\rho}-1} c'^{\rho-1} = \mathcal{W}\left(c', \nu'\right)^{\frac{1-\eta-\rho}{1-\eta}}(1-\eta)^{\frac{1-\eta-\rho}{1-\eta}} c'^{\rho-1} \tag{7.17}$$

where $\nu' \equiv \nu(x', y')$. Therefore,

$$m\left(x, y; x'y'\right) = \frac{\mathcal{W}_2(c,\nu)}{\mathcal{W}_1(c,\nu)}\mathcal{W}_1\left(c', \nu'\right) = e^{-\delta}\left(\frac{\nu}{\mathcal{W}(c', \nu')}\right)^{\frac{\rho}{1-\eta}-1}\left(\frac{c'}{c}\right)^{\rho-1}.$$

Along any optimal consumption path, $V(x,y)=\mathcal{W}(c(x,y),\nu(x,y))$. Therefore,

$$m\left(x,y;x'y'\right)=e^{-\delta}\left[\frac{E\left(V\left(x',y'\right)\right)}{V\left(x',y'\right)}\right]^{\frac{\rho}{1-\eta}-1}\left(\frac{c'}{c}\right)^{\rho-1}. \tag{7.18}$$

We are left with evaluating the term $\frac{E(V(x',y'))}{V(x',y')}$. The conjecture to make is that $v(x,y)=\psi(y)^{1/(1-\eta)}x$, for some function ψ. From this, it follows that $V(x,y)=\psi(y)x^{1-\eta}/(1-\eta)$. We have,

$$\begin{aligned}&V_1(x,y)\\=\;&\mathcal{W}_1\left(c(x,y),E\left(V\left(x',y'\right)\right)\right)=\mathcal{W}(c,\nu)^{\frac{1-\eta-\rho}{1-\eta}}(1-\eta)^{\frac{1-\eta-\rho}{1-\eta}}c^{\rho-1}=V(x,y)^{\frac{1-\eta-\rho}{1-\eta}}(1-\eta)^{\frac{1-\eta-\rho}{1-\eta}}c^{\rho-1}.\end{aligned}$$

where the first equality follows by Eq. (7.6), the second equality follows by Eq. (7.17), and the last equality follows by optimality. By making use of the conjecture on V, and rearraning terms,

$$c(x,y)=a(y)x,\quad a(y)\equiv\psi(y)^{\frac{\rho}{(1-\eta)(\rho-1)}}. \tag{7.19}$$

Hence, $V(x',y')=\psi(y')x'^{1-\eta}/(1-\eta)$, where

$$x'=(1-a(y))x\left(1+r_M\left(y'\right)\right), \tag{7.20}$$

and

$$\frac{E\left(V\left(x',y'\right)\right)}{V\left(x',y'\right)}=\frac{E\left[\psi\left(y'\right)\left(1+r_M\left(y'\right)\right)^{1-\eta}\right]}{\psi\left(y'\right)\left(1+r_M\left(y'\right)\right)^{1-\eta}}. \tag{7.21}$$

Along any optimal path, $V(x,y)=\mathcal{W}(c(x,y),E(V(x',y')))$. By plugging in $\mathcal{W}$ (from Eq. 7.4)) and the conjecture for V,

$$E\left[\psi\left(y'\right)\left(1+r_M\left(y'\right)\right)^{1-\eta}\right]=\left(e^{-\delta}\right)^{-\frac{1-\eta}{\rho}}\left(\frac{a(y)}{1-a(y)}\right)^{\frac{(1-\eta)(\rho-1)}{\rho}}. \tag{7.22}$$

Moreover,

$$\psi\left(y'\right)\left(1+r_M\left(y'\right)\right)^{1-\eta}=\left[a\left(y'\right)\left(1+r_M\left(y'\right)\right)^{\frac{\rho}{\rho-1}}\right]^{\frac{(1-\eta)(\rho-1)}{\rho}}. \tag{7.23}$$

By plugging Eqs. (7.22)-(7.23) into Eq. (7.21),

$$\begin{aligned}\frac{E\left(V\left(x',y'\right)\right)}{V\left(x',y'\right)}&=\left(e^{-\delta}\right)^{-\frac{1-\eta}{\rho}}\left[\frac{a(y)}{(1-a(y))\,a(y')\left(1+r_M(y')\right)^{\frac{\rho}{\rho-1}}}\right]^{\frac{(1-\eta)(\rho-1)}{\rho}}\\&=\left(e^{-\delta}\right)^{-\frac{1-\eta}{\rho}}\left[\left(\frac{c'}{c}\right)^{-1}\frac{x'}{(1-a(y))\,x\left(1+r_M(y')\right)^{\frac{\rho}{\rho-1}}}\right]^{\frac{(1-\eta)(\rho-1)}{\rho}}\\&=\left(e^{-\delta}\right)^{-\frac{1-\eta}{\rho}}\left[\left(\frac{c'}{c}\right)^{-1}\frac{1}{\left(1+r_M(y')\right)^{\frac{1}{\rho-1}}}\right]^{\frac{(1-\eta)(\rho-1)}{\rho}}\end{aligned}$$

where the first equality follows by Eq. (7.19), and the second equality follows by Eq. (7.20). The result follows by replacing this into Eq. (7.18). ■

7.5.2 Derivation of interest rates and risk-premia in multifactor models

We have,

QUI QUI QUI

7.6 Appendix on economies with heterogenous agents

When every agent faces a system of complete markets, the equilibrium can be computed along the lines of Huang (1987), who generalizes the classical approach described in Chapter 2 of Part I of these Lectures. To keep the presentation as close as possible to some of the models of this chapter, we first consider the case of a continuum of agents indexed by the instantaneous utility function $u_a(c, x)$, where c is consumption, a is some parameter belonging to some set A, and x is some variable. (For example, x is the "standard of living of others" in the Chan and Kogan (2002) model.)

In the context of this chapter, the equilibrium allocation is Pareto efficient if every agent $a \in A$ faces a system of complete markets. But by the second welfare theorem, we know that for each Pareto allocation $(c_a)_{a\in A}$, there exists a social weighting function f such that the Pareto allocation can be "implemented" by means of the following program,

$$u(D, x) = \max_{c_a} \int_{a\in A} u_a(c_a, x) f(a) \, da, \qquad \text{s.t.} \int_{a\in A} c_a da = D,$$

where D is the aggregate endowment in the economy. Then, the equilibrium price system is the Arrow-Debreu state price density in an economy with a single agent endowed with the aggregate endowment D, instantaneous utility function $u(c, x)$, and where the social weighting function f equals the reciprocal of the marginal utility of income of the agents, i.e. for $a \in A$, we have that $f(a) =$ marginal utility of income of the agent a.

The practical merit of this approach is that while the marginal utility of income is unobservable, the thusly constructed Arrow-Debreu state price density depends on the "infinite dimensional parameter", f, which can be easily calibrated to reproduce the main quantitative features of consumption and asset price data.

We now apply this approach to indicate how to derive the Chan and Kogan (2002) equilibrium conditions.

The "catching up with the Joneses" economy in Chan and Kogan (2002). In the contex of this model, markets are complete, and we have that $A = [1, \infty]$ and $u_\eta(c_\eta, x) = (c_\eta/x)^{1-\eta}/(1-\eta)$. The static optimization problem can be written as,

$$u(D, x) = \max_{c_\eta} \int_1^\infty \frac{(c_\eta/x)^{1-\eta}}{1-\eta} f(\eta) \, d\eta, \qquad \text{s.t.} \int_1^\infty (c_\eta/x) \, d\eta = D/x. \qquad \text{[Soc-Pl]}$$

The first order conditions of the social planner problem [Soc-Pl] lead to,

$$(c_\eta/x)^{-\eta} f(\eta) = V(D/x), \tag{7.24}$$

where V is a Lagrange multiplier, a function of the aggregate endowment D, normalized by x. It is determined by the equation,

$$\int_1^\infty V(D/x)^{-\frac{1}{\eta}} f(\eta)^{\frac{1}{\eta}} \, d\eta = D/x,$$

which is obtained by replacing Eq. (7.24) into the integrand of the social planner utility function. The expression for the unit risk-premium in Eq. (7.12) follows by,

$$\lambda(s) = -\left(\frac{\partial^2 u(D, x)}{\partial D^2} \Big/ \frac{\partial u(D, x)}{\partial D} \right) \sigma_0 D,$$

and lenghty computations, after setting $D/x = e^s$. The short-term rate can be computed by calculating the expectation of the pricing kernel in this fictitious representative agent economy.

It is instructive to compare the first order conditions of the social planner in Eq. (7.24) with those of the decentralized economy. Since markets are complete, we have that the first order conditions in the decentralized economy satisfy:

$$e^{-\delta t}\left(c_{\eta}(t)/x(t)\right)^{-\eta}=\kappa(\eta)\,\xi(t)\,x(t), \tag{7.25}$$

where $\kappa(\eta)$ is the marginal utility of income for the agent η, and $\xi(t)$ is the usual pricing kernel. By eliminating $(c_{\eta}/x)^{-\eta}$ from Eq. (7.24) and Eq. (7.25) we obtain,

$$\frac{e^{\delta t}x(t)\,\xi(t)}{V(D(t)/x(t))}=\frac{1}{f(\eta)\,\kappa(\eta)},\quad (t,\eta)\in[0,\infty)\times[1,\infty).$$

In this equation, the left hand side does not depend on η, and the right hand side does not depend on time t. Therefore, there must exist a constant *const* >0 such that both sides of this equation are equal to. Moreover, we have

$$const=\frac{x(0)}{V(D(0)/x(0))}.$$

Hence, $f(\eta)$ is inversely related to the marginal utility of income for the agent η. [**Remains to be shown that** $const=1$.]

Incomplete markets in the Basak and Cuoco (1998) economy with restricted stock market participation. We first show that Eq. (7.16) holds true. Indeed, by the definition of the stochastic social weight in Eq. (7.15), we have that

$$x(\tau)=u_p'(\hat{c}_p(\tau))\hat{c}_n(\tau)=\frac{w_p}{w_n}\xi(\tau)\,e^{\int_0^{\tau}R(s)ds}$$

where the second line follows by the first order conditions in Eq. (7.13). Eq. (7.16) follows by the previous expression for x and the dynamics for the pricing kernel in Eq. (7.14).

QUI QUI QUI

By Chapter 6 (Appendix 1), the unit risk premium λ satisfies,

$$\lambda(D,x)=-\frac{u_{11}(D,x)}{u_1(D,x)}\sigma_0 D+\frac{u_{12}(D,x)x}{u_1(D,x)}\lambda(D,x).$$

This is:

$$\begin{aligned}\lambda(D,x)&=-\frac{u_1(D,x)u_{11}(D,x)}{u_1(D,x)-u_{12}(D,x)x}\cdot\frac{\sigma_0 D}{u_1(D,x)}\\&=-\frac{u_a''(\hat{c}_a)}{u_1(D,x)}\sigma_0 D\\&=-\frac{u_a''(\hat{c}_a)\hat{c}_a}{u_a'(\hat{c}_a)}\sigma_0 s^{-1}.\end{aligned}$$

where the second line follows by Basak and Cuoco (identity (33), p. 331) and the third line follows by the definition of $u(D,x)$ and s. The Sharpe ratio reported in the main text follows by the definition of u_a. The interest rate is also found through Chapter 6 (appendix 1). We have,

$$R(s)=\delta+\frac{\eta g_0}{\eta-(\eta-1)s}-\frac{1}{2}\frac{\eta(\eta+1)\sigma_0^2}{s(\eta-(\eta-1)s)}.$$

Finally, by applying Itô's lemma to $s = \frac{c_a}{D}$, and using the optimality conditions for agent a, I find that drift and diffusion functions of s are given by:

$$\phi(s) = g_0 \left[\frac{(1-\eta)(1-s)}{\eta + (1-\eta)s} \right] s - \frac{1}{2} \frac{(\eta+1)\sigma_0^2}{\eta + (1-\eta)\, s} + \frac{1}{2} \frac{(\eta+1)\sigma_0^2}{s} + \sigma_0 (s-1),$$

and $\xi(s) = \sigma_0(1-s)$.

References

Basak, S. and D. Cuoco (1998). "An Equilibrium Model with Restricted Stock Market Participation," *Review of Financial Studies* 11: 309-341.

Chan, Y.L. and L. Kogan (2002). "Catching Up with the Joneses: Heterogeneous Preferences and the Dynamics of Asset Prices." *Journal of Political Economy* 110: 1255-1285.

Duffie, D. and L.G. Epstein (1992*a*). "Asset Pricing with Stochastic Differential Utility." *Review of Financial Studies* 5: 411-436.

Duffie, D. and L.G. Epstein (with C. Skiadas) (1992*b*). "Stochastic Differential Utility." *Econometrica* 60: 353-394.

Epstein, L.G. and S.E. Zin (1989). "Substitution, Risk-Aversion and the Temporal Behavior of Consumption and Asset Returns: A Theoretical Framework." *Econometrica* 57: 937-969.

Epstein, L.G. and S.E. Zin (1991). "Substitution, Risk-Aversion and the Temporal Behavior of Consumption and Asset Returns: An Empirical Analysis." *Journal of Political Economy* 99: 263-286.

Guvenen, F. (2005). "A Parsimonious Macroeconomic Model for Asset Pricing: Habit formation or Cross-Sectional Heterogeneity." Working paper, University of Rochester.

Huang, C.-F. (1987). "An Intertemporal General Equilibrium Asset Pricing Model: the Case of Diffusion Information." *Econometrica*, 55, 117-142.

Weil, Ph. (1989). "The Equity Premium Puzzle and the Risk-Free Rate Puzzle." *Journal of Monetary Economics* 24: 401-421.

Part III

Applied asset pricing theory

8
Derivatives

8.1 Introduction

This chapter is under construction. I shall include material on futures, American options, exotic options, and evaluation of contingent through trees. I shall illustrate more systematically the general ideas underlying implied trees, and cover details on how to deal with market imperfections.

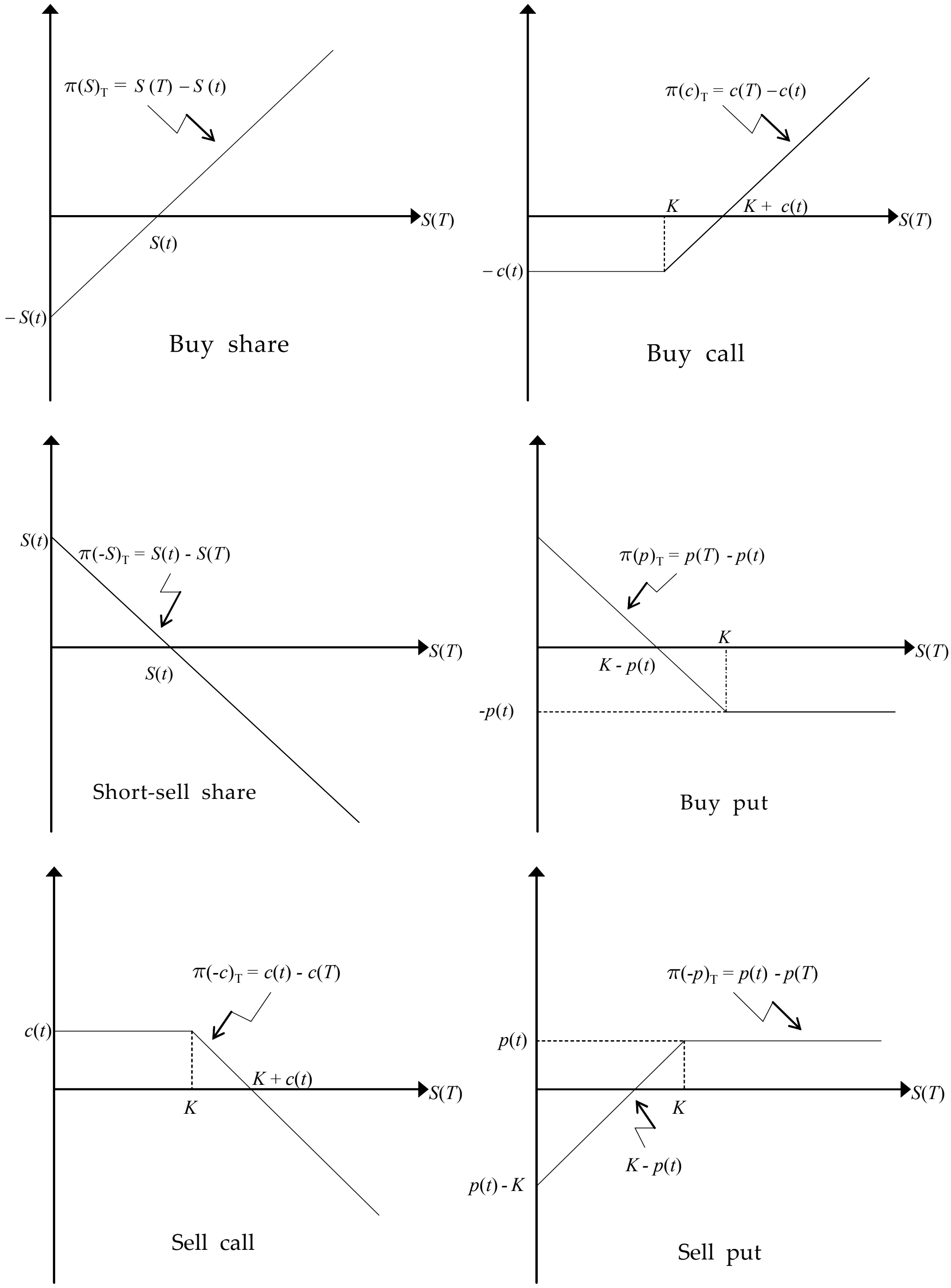
8.2 General properties of derivative prices

A *European call* (*put*) *option* is a contract by which the buyer has the right, but not the obligation, to *buy* (*sell*) a given asset at some price, called the strike, or exercise price, at some future date. Let C and p be the prices of the call and the put option. Let S be the price of the asset underlying the contract, and K and T be the exercise price and the expiration date. Finally, let t be the current evaluation time. The following relations hold true,

$$C(T) = \begin{cases} 0 & \text{if } S(T) \leq K \\ S(T) - K & \text{if } S(T) > K \end{cases} \qquad p(T) = \begin{cases} K - S(T) & \text{if } S(T) \leq K \\ 0 & \text{if } S(T) > K \end{cases}$$

or more succinctly, $C(T) = (S(T) - K)^+$ and $p(T) = (K - S(T))^+$.

Figure 8.1 depicts the net profits generated by a number of portfolios that include one asset (a share, say) and/or one option written on the share: buy a share, buy a call, short-sell a share, buy a put, etc. To simplify the exposition, we assume that the short-term rate $r = 0$. The first two panels in Figure 8.1 illustrate instances in which the exposure to losses is less important when we purchase an option than the share underlying the option contract. For example, consider the first panel, in which we depicts the two net profits related to (i) the purchase of the share and (ii) the purchase of the call option written on the share. Both cases generate positive net profits when $S(T)$ is high. However, the call option provides "protection" when $S(T)$ is low, at least insofar as $C(t) < S(t)$, a no-arbitrage condition we shall demonstrate later. It is this protecting feature to make the option contract economically valuable.



π(S)T = S (T) − S (t)
S(T)
S(t)
− S(t)
Buy share
π(c)T = c(T) − c(t)
K
K + c(t)
S(T)
− c(t)
Buy call
S(t)
π(-S)T = S(t) - S(T)
S(T)
S(t)
Short-sell share
π(p)T = p(T) - p(t)
K
S(T)
K - p(t)
-p(t)
Buy put
π(-c)T = c(t) - c(T)
c(t)
K + c(t)
S(T)
K
Sell call
π(-p)T = p(t) - p(T)
p(t)
S(T)
K
K - p(t)
p(t) - K
Sell put

FIGURE 8.1.

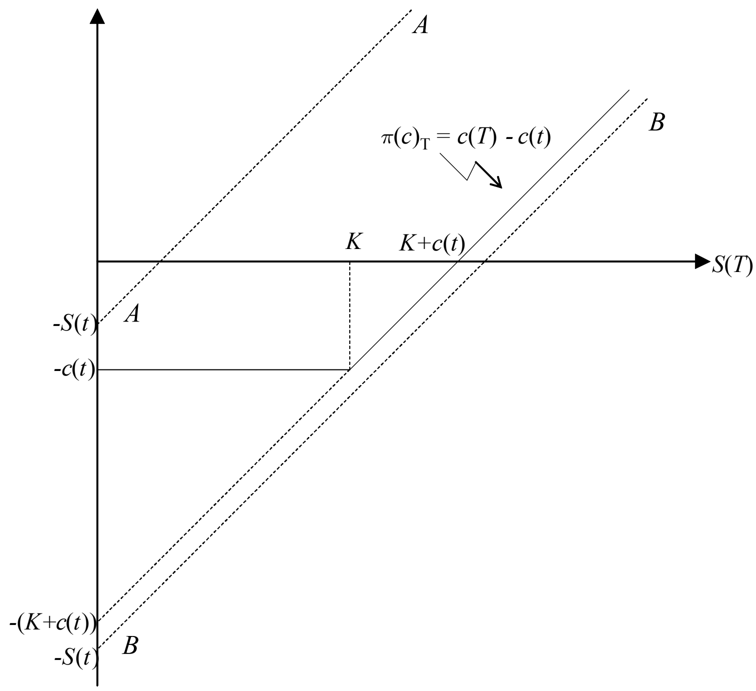


FIGURE 8.2.

Figure 8.2 illustrates this. It depicts two price configurations. The first configuration is such that the share price is less than the option price. In this case, the net profits generated by the purchase of the share, the AA line, would strictly dominate the net profits generated by the purchase of the option, $\pi(C)_T$, an arbitrage opportunity. Therefore, we need that $C(t) < S(t)$ to rule this arbitrage opportunity. Next, let us consider a second price configuration. Such a second configuration arises when the share and option prices are such that the net profits generated by the purchase of the share, the BB line, is always dominated by the net profits generated by the purchase of the option, $\pi(C)_T$, an arbitrage opportunity. Therefore, the price of the share and the price of the option must be such that $S(t) < K + C(t)$, or $C(t) > S(t) - K$. Finally, note that the option price is always strictly positive as the option payoff is nonnegative. Therefore, we have that $(S(t) - K)^+ < C(t) < S(t)$. Proposition 8.1 below provides a formal derivation of this result in the more general case in which the short-term rate is positive.

Let us go back to Figure 8.1. This figure suggests that the price of the call and the price of the put are intimately related. Indeed, by overlapping the first two panels, and using the same arguments used in Figure 8.2, we see that for two call and put contracts with the same exercise price K and expiration date T, we must have that $C(t) > p(t) - K$. The *put-call parity* provides the exact relation between the two prices $C(t)$ and $p(t)$. Let $P(t,T)$ be the price as of time t of a pure discount bond expiring at time T. We have:

PROPOSITION 8.1 (Put-call parity). *Consider a put and a call option with the same exercise price K and the same expiration date T. Their prices $p(t)$ and $C(t)$ satisfy,* $p(t) = C(t) - S(t) + KP(t,T)$.

PROOF. Consider two portfolios: (A) Long one call, short one underlying asset, and invest $KP(t,T)$; (B) Long one put. The table below gives the value of the two portfolios at time t

and at time T.

	Value at t	Value at T: $S(T) \leq K$	Value at T: $S(T) > K$
Portfolio A	$C(t) - S(t) + KP(t,T)$	$-S(T) + K$	$S(T) - K - S(T) + K$
Portfolio B	$p(t)$	$K - S(T)$	0

The two portfolios have the same value in each state of nature at time T. Therefore, their values at time t must be identical to rule out arbitrage. ∎

By proposition 8.1, the properties of European put prices can be mechanically deduced from those of the corresponding call prices. We focus the discussion on European call options. The following proposition gathers some basic properties of European call option prices *before* the expiration date, and generalizes the reasoning underlying Figure 8.2.

Proposition 8.2. *The rational option price* $C(t) = C(S(t); K; T-t)$ *satisfies the following properties: (i)* $C(S(t); K; T-t) \geq 0$*; (ii)* $C(S(t); K; T-t) \geq S(t) - KP(t,T)$*; and (iii)* $C(S(t); K; T-t) \leq S(t)$.

Proof. Part (i) holds because $\Pr\{C(S(T); K; 0) > 0\} > 0$, which implies that C must be nonnegative at time t to preclude arbitrage opportunities. As regards Part (ii), consider two portfolios: Portfolio A, buy one call; and Portfolio B, buy one underlying asset and issue debt for an amount of $KP(t,T)$. The table below gives the value of the two portfolios at time t and at time T.

	Value at t	Value at T: $S(T) \leq K$	Value at T: $S(T) > K$
Portfolio A	$C(t)$	0	$S(T) - K$
Portfolio B	$S(t) - KP(t,T)$	$S(T) - K$	$S(T) - K$

At time T, Portfolio A dominates Portfolio B. Therefore, in the absence of arbitrage, the value of Portfolio A must dominate the value of Portfolio B at time t. To show Part (iii), suppose the contrary, i.e. $C(t) > S(t)$, which is an arbitrage opportunity. Indeed, at time t, we could sell m options (m large) and buy m of the underlying assets, thus making a *sure* profit equal to $m \cdot (C(t) - S(t))$. At time T, the option will be exercized if $S(T) > K$, in which case we shall sell the underlying assets and obtain $m \cdot K$. If $S(T) < K$, the option will not be exercized, and we will still hold the asset or sell it and make a profit equal to $m \cdot S(T)$. ∎

Proposition 8.2 can be summarized as follows:

$$\max\{0, S(t) - KP(t,T)\} \leq C(S(t); K; T-t) \leq S(t). \tag{8.1}$$

The next corollary follows by Eq. (8.1).

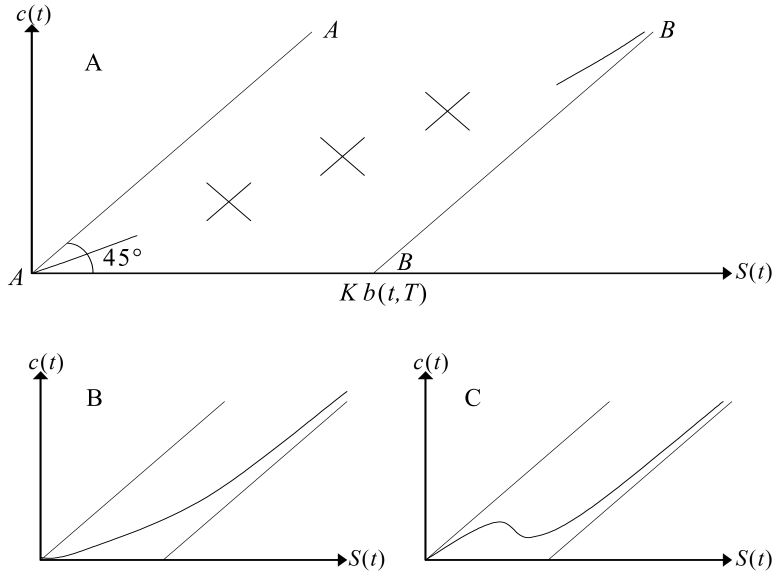


FIGURE 8.3.

COROLLARY 8.3. *We have, (i)* $\lim_{S\to 0} C(S;K;T-t) \to 0$*; (ii)* $\lim_{K\to 0} C(S;K;T-t) \to S$*; (iii)* $\lim_{T\to\infty} C(S;K;T-t) \to S$.

The previous results provide the basic properties, or arbitrage bounds, that option prices satisfy. First, consider the top panel of Figure 8.3. Eq. (8.1) tells us that $C(t)$ must lie inside the AA and the BB lines. Second, Corollary 8.3(i) tells us that the rational option price starts from the origin. Third, Eq. (8.1) also reveals that as $S \to \infty$, the option price also goes to infinity; but because C cannot lie outside the the region bounded by the AA line and the BB lines, C will go to infinity by "sliding up" through the BB line.

How does the option price behave within the bounds AA and BB? That is impossible to tell. Given the boundary behavior of the call option price, we only know that if the option price is strictly convex in S, then it is also increasing in S, In this case, the option price could be as in the left-hand side of the bottom panel of Figure 8.3. This case is the most relevant, empirically. It is predicted by the celebrated Black and Scholes (1973) formula. However, this property is not a *general* property of option prices. Indeed, Bergman, Grundy and Wiener (1996) show that in one-dimensional *diffusion* models, the price of a contingent claim written on a tradable asset is convex in the underlying asset price if the payoff of the claim is convex in the underlying asset price (as in the case of a Europen call option). In our context, the boundary conditions guarantee that the price of the option is then increasing and convex in the price of the underlying asset. However, Bergman, Grundy and Wiener provide several counter-examples in which the price of a call option can be decreasing over some range of the price of the asset underlying the option contract. These counter-examples include models with jumps, or the models with stochastic volatility that we shall describe later in this chapter. Therefore, there are no reasons to exclude that the option price behavior could be as that in right-hand side of the bottom panel of Figure 8.3.

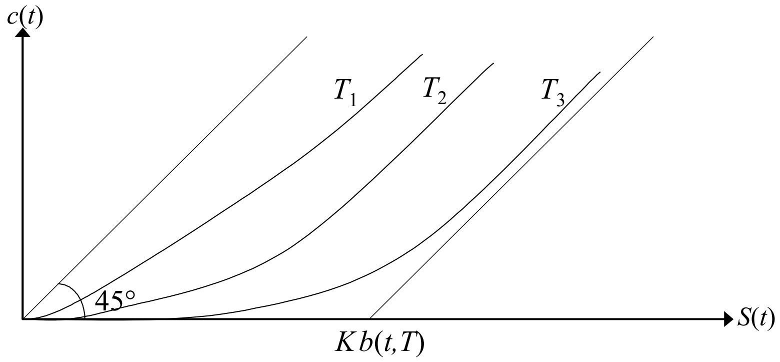


FIGURE 8.4.

The economic content of convexity in this context is very simple. When the price S is small, it is unlikely that the option will be exercized. Therefore, changes in the price S produce little effect on the price of the option, C. However, when S is large, it is likely that the option will be exercized. In this case, an increase in S is followed by almost the same increase in C. Furthermore, the elasticity of the option price with respect to S is larger than one,

$$\epsilon \equiv \frac{dC}{dS} \cdot \frac{S}{C} > 1,$$

as for a convex function, the first derivative is always higher than the secant. Overall, the option price is more volatile than the price of the asset underlying the contract. Finally, call options are also known as "wasting assets", as their value decreases over time. Figure 8.4 illustrate this property, in correspondence of the maturity dates $T_1 > T_2 > T_3$.

These properties illustrate very simply the general principles underlying a portfolio that "mimicks" the option price. For example, investment banks sell options that they want to hedge against, to avoid the exposure to losses illustrated in Figure 8.1. At a very least, the portfolio that "mimicks" the option price must exhibit the previous general properties. For example, suppose we wish our portfolio to exhibit the behavior in the left-hand side of the bottom panel of Figure 8.3, which is the most relevant, empirically. We require the portfolio to exhibit a number of properties.

P1) The portfolio value, V, must be increasing in the underlying asset price, S.

P2) The sensitivity of the portfolio value with respect to the underlying asset price must be strictly positive and bounded by one, $0 < \frac{dV}{dS} < 1$.

P3) The elasticity of the portfolio value with respect to the underlying asset price must be strictly greater than one, $\frac{dV}{dS} \cdot \frac{S}{V} > 1$.

The previous properties hold under the following conditions:

C1) The portfolio includes the asset underlying the option contract.

C2) The number of assets underlying the option contract is less than one.

C3) The portfolio includes debt to create a sufficiently large elasticity. Indeed, let $V = \theta S - D$, where θ is the number of assets underlying the option contract, with $\theta \in (0,1)$, and D is debt. Then, $\frac{dV}{dS} > \theta$ and $\frac{dV}{dS} \cdot \frac{S}{V} = \theta \cdot \frac{S}{V} > 1 \Leftrightarrow \theta S > V = \theta S - D$, which holds if and only if $D > 0$.

In fact, the hedging problem is dynamic in nature, and we would expect θ to be a function of the underlying asset price, S, and time to expiration. Therefore, we require the portfolio to display the following additional property:

P4) The number of assets underlying the option contract must increase with S. Moreover, when S is low, the value of the portfolio must be virtually insensitive to changes in S. When S is high, the portfolio must include mainly the assets underlying the option contract, to make the portfolio value "slide up" through the BB line in Figure 8.3.

The previous property holds under the following condition:

C4) θ is an increasing function of S, with $\lim_{S\to 0} \theta(S) \to 0$ and $\lim_{S\to\infty} \theta(S) \to 1$.

Finally, the purchase of the option does not entail any additional inflows or outflows until time to expiration. Therefore, we require that the "mimicking" portfolio display a similar property:

P5) The portfolio must be implemented as follows: (i) any purchase of the asset underlying the option contract must be financed by issue of new debt; and (ii) any sells of the asset underlying the option contract must be used to shrink the existing debt:

The previous property of the portfolio just says that the portfolio has to be self-financing, in the sense described in the first Part of these lectures.

C5) The portfolio is implemented through a *self-financing* strategy.

We now proceed to add more structure to the problem.

8.3 Evaluation

We consider a continuous-time model in which asset prices are driven by a d-dimensional Brownian motion W.[1] We consider a multivariate state process

$$dY^{(h)}(t) = \varphi_h(y(t))\, dt + \sum_{j=1}^{d} \ell_{hj}(y(t))\, dW^{(j)}(t),$$

for some functions φ_h and $\ell_{hj}(y)$, satisfying the usual regularity conditions.

The price of the primitive assets satisfy the regularity conditions in Chapter 4. The value of a portfolio strategy, V, is $V(t) = \theta(t) \cdot S_+(t)$. We consider a self-financing portfolio. Therefore, V is solution to

$$dV(t) = \left[\pi(t)^\top (\mu(t) - \mathbf{1}_m r(t)) + r(t) V(t) - C(t)\right] dt + \pi(t)^\top \sigma(t)\, dW(t), \tag{8.2}$$

where $\pi \equiv (\pi^{(1)}, ..., \pi^{(m)})^\top$, $\pi^{(i)} \equiv \theta^{(i)} S^{(i)}$, $\mu \equiv (\mu^{(1)}, ..., \mu^{(m)})^\top$, $S^{(i)}$ is the price of the i-th asset, $\mu^{(i)}$ is its drift and $\sigma(t)$ is the volatility matrix of the price process. We impose that V satisfy the same regularity conditions in Chapter 4.

[1] As usual, we let $\{\mathcal{F}(t)\}_{t\in[0,T]}$ be the P-augmentation of the natural filtration $\mathcal{F}^W(t) = \sigma(W(s), s \leq t)$ generated by W, with $\mathcal{F} = \mathcal{F}(T)$.

8.3.1 On spanning and cloning

Chapter 4 emphasizes how "spanning" is important to solve for optimal consumption-portfolio choices using martingale techniques. This section emphasizes how spanning helps defining replicating strategies that lead to price "redundant" assets. Heuristically, a set of securities "spans" a given vector space if any point in that space can be generated by a linear combination of the security prices. In our context, "spanning" means that a portfolio of securities can "replicate" any $\mathcal{F}(T)$-measurable random variable. The random variable can be the payoff promised by a contingent claim, such as that promised by a European call. Alternatively, we may think of this variable to be net consumption, as in the two-dates economy with a continuum of intermediate financial transaction dates considered by Harrison and Kreps (1979) and Duffie and Huang (1985) (see Chapter 4).

Formally, let $V^{x,\pi}(t)$ denote the solution to Eq. (8.2) when the initial wealth is x, the portfolio policy is π, and the intermediate consumption is $C \equiv 0$. We say that the portfolio policy π spans $\mathcal{F}(T)$ if $V^{x,\pi}(T) = \tilde{X}$ almost surely, where $\tilde{X}$ is any square-integrable $\mathcal{F}(T)$-measurable random variable.

Chapter 4 provides a characterization of the spanning property relying on the behavior of asset prices under the risk-adjusted probability measure, Q. In the context of this chapter, it is convenient to analyze the behavior of the asset prices under the physical probability, P. In the diffusion environment of this chapter, the asset prices are semimartingales under P. More generally, let consider the following representation of a $\mathcal{F}(t)$-P semimartingale,

$$dA(t) = dF(t) + \tilde{\gamma}(t)dW(t),$$

where F is a process with finite variation, and $\tilde{\gamma} \in L^2_{0,T,d}(\Omega, \mathcal{F}, P)$. We wish to replicate A through a portfolio. First, then, we must look for a portfolio π satisfying

$$\tilde{\gamma}(t) = \left(\pi^\top \sigma\right)(t). \tag{8.3}$$

Second, we equate the drift of V to the drift of F, obtaining,

$$\frac{dF(t)}{dt} = \pi(t)^\top (\mu(t) - \mathbf{1}_m r(t)) + r(t) V(t) = \pi(t)^\top (\mu(t) - \mathbf{1}_m r(t)) + r(t) F(t). \tag{8.4}$$

The second equality holds because if drift and diffusion terms of F and V are identical, then $F(t) = V(t)$.

Clearly, if $m < d$, there are no solutions for π in Eq. (8.3). The economic interpretation is that in this case, the number of assets is so small that we can not create a portfolio able to replicate all possible events in the future. Mathematically, if $m < d$, then $V^{x,\pi}(T) \in M \subset L^2(\Omega, \mathcal{F}, P)$. As Chapter 4 emphasizes, there is also a converse to this result, which motivates the definition of market incompleteness given in Chapter 4 (Definition 4.5).

8.3.2 Option pricing

Eq. (8.4) is all we need to price any derivative asset which promises to pay off some $\mathcal{F}(T)$-measurable random variable. The first step is to cast Eq. (8.4) in a less abstract format. Let us consider a semimartingale in the context of the diffusion model of this section. For example, let us consider the price process $H(t)$ of a European call option. This price process is rationally formed if $H(t) = C(t, y(t))$, for some $C \in \mathcal{C}^{1,2}\left([0,T) \times \mathbb{R}^k\right)$. By Itô's lemma,

$$dC = \bar{\mu}^C C dt + \left(\frac{\partial C}{\partial Y} J\right) dW,$$

where $\bar{\mu}^C C = \frac{\partial C}{\partial t} + \sum_{l=1}^{k} \frac{\partial C}{\partial y_l} \varphi_l(t, y) + \frac{1}{2} \sum_{l=1}^{k} \sum_{j=1}^{k} \frac{\partial^2 C}{\partial y_l \partial y_j} cov(y_l, y_j)$; $\partial C / \partial Y$ is $1 \times d$; and J is $d \times d$. Finally,

$$C(T, y) = \tilde{X} \in L^2(\Omega, \mathcal{F}, P).$$

In this context, $\bar{\mu}^C C$ and $(\partial C / \partial Y) \cdot J$ play the same roles as dF/dt and $\tilde{\gamma}$ in the previous section. In particular, the volatility identification

$$\left(\frac{\partial C}{\partial Y} J\right)(t) = \left(\pi^{\top} \sigma\right)(t),$$

corresponds to Eq. (8.3).

As an example, let $m = d = 1$, and suppose that the only state variable of the economy is the price of a share, $\varphi(s) = \mu s$, $cov(s) = \sigma^2 s^2$, and μ and σ^2 are constants, then $(\partial C / \partial Y) \cdot J = C_S \sigma S$, $\pi = C_S S$, and by Eq. (8.4),

$$\frac{\partial C}{\partial t} + \frac{\partial C}{\partial S} \mu S + \frac{1}{2} \frac{\partial^2 C}{\partial S^2} \sigma^2 S^2 = \pi(\mu - r) + rC = \frac{\partial C}{\partial S} S(\mu - r) + rC. \tag{8.5}$$

By the boundary condition, $C(T, s) = (s - K)^+$, one obtains that the solution is the celebrated Black and Scholes (1973) formula,

$$C(t, S) = S\Phi(d_1) - Ke^{-r(T-t)} \Phi(d_2) \tag{8.6}$$

where

$$d_1 = \frac{\log(\frac{S}{K}) + (r + \frac{1}{2}\sigma^2)(T - t)}{\sigma\sqrt{T - t}}, \quad d_2 = d_1 - \sigma\sqrt{T - t}; \quad \Phi(x) = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{x} e^{-\frac{1}{2}u^2} du.$$

Note that we can obtain Eq. (8.6) even without assuming that a market exists for the option during the life of the option,[2] and that the pricing function $C(t, S)$ is differentiable. As it turns out, the option price *is* differentiable, but this can be shown to be a *result*, not just an *assumption*. Indeed, let us define the function $C(t, S)$ that solves Eq. (9.27), with the boundary condition $C(T, S) = (S - K)^+$. Note, we are *not* assuming this function is the option price. Rather, we shall *show* this is the option price. Consider a self-financed portfolio of bonds and stocks, with $\pi = C_S S$. Its value satisfies,

$$dV = [C_S S(\mu - r) + rV] \, dt + C_S \sigma S dW.$$

Moreover, by Itô's lemma, $C(t, s)$ is solution to

$$dC = \left(C_t + \mu S C_S + \frac{1}{2}\sigma^2 S^2 C_{SS}\right) dt + C_S \sigma S dW.$$

By substracting these two equations,

$$dV - dC = [\underbrace{-C_t - rSC_S - \frac{1}{2}\sigma^2 S^2 C_{SS}}_{-rC} + rV]dt = r(V - C) \, dt.$$

Hence, we have that $V(\tau) - C(\tau, S(\tau)) = [V(0) - C(0, S(0))] \exp(r\tau)$, for all $\tau \in [0, T]$. Next, assume that $V(0) = C(0, S(0))$. Then, $V(\tau) = C(\tau, S(\tau))$ and $V(T) = C(T, S(T)) = (S(T) - K)^+$. That is, the portfolio $\pi = C_S S$ replicates the payoff underlying the option contract. Therefore, $V(\tau)$ equals the market price of the option. But $V(\tau) = C(\tau, S(\tau))$, and we are done.

[2]The original derivation of Black and Scholes (1973) and Merton (1973) relies on the assumption that an option market exists (see the Appendix to this chapter).

8.3.3 *The surprising cancellation, and the real meaning of "preference-free" formulae*

Due to what Heston (1993*a*) (p. 933) terms "a surprising cancellation", the constant μ doesn't show up in the final formula. Heston (1993*a*) shows that this property is not robust to modifications in the assumptions for the underlying asset price process.

8.4 Properties of models

8.4.1 *Rational price reaction to random changes in the state variables*

We now derive some general properties of option prices arising in the context of diffusion processes. The discussion in this section hinges upon the seminal contribution of Bergman, Grundy and Wiener (1996).[3] We take as primitive,

$$\frac{dS(\tau)}{S(\tau)} = \mu(S(\tau))\,d\tau + \sqrt{2\sigma(S(\tau))}dW(\tau),$$

and develop some properties of a European-style option price at time τ, denoted as $C(S(\tau), \tau, T)$, where T is time-to-expiration. Let the payoff of the option be the function $\psi(S)$, where ψ satisfies $\psi'(S) > 0$. In the absence of arbitarge, C satisfies the following partial differential equation

$$\begin{cases} 0 &= C_\tau + C_S r + C_{SS}\sigma(S) - rC & \text{for all } (\tau, S) \in [t, T) \times \mathbb{R}_{++} \\ C(S,T,T) &= \psi(S) & \text{for all } S \in \mathbb{R}_{++} \end{cases}$$

Let us differentiate the previous partial differential equation with respect to S. The result is that $H \equiv C_S$ satisfies another partial differential equation,

$$\begin{cases} 0 &= H_\tau + (r + \sigma'(S))\,H_S + H_{SS}\sigma(S) - rH & \text{for all } (\tau, S) \in [t, T) \times \mathbb{R}_{++} \\ H(S,T,T) &= \psi'(S) > 0 & \text{for all } S \in \mathbb{R}_{++} \end{cases} \tag{8.7}$$

By the maximum principle reviewed in the Appendix of Chapter 6, we have that,

$$H(S, \tau, T) > 0 \text{ for all } (\tau, S) \in [t, T] \times \mathbb{R}_{++}.$$

That is, we have that *in the scalar diffusion setting, the option price is always increasing in the underlying asset price.*

Next, let us consider how the price behave when we tilt the volatility of the underlying asset price. That is, consider two economies A and B with prices $(C^i, S^i)_{i=A,B}$. We assume that the asset price volatility is larger in the first economy than in the second economy, viz

$$\frac{dS^i(\tau)}{S^i(\tau)} = r d\tau + \sqrt{2\sigma^i(S^i(\tau))}d\hat{W}(\tau),\ i = A, B,$$

where $\hat{W}$ is Brownian motion under the risk-neutral probability, and $\sigma^A(S) > \sigma^B(S)$, for all S. It is easy to see that the price difference, $\nabla C \equiv C^A - C^B$, satisfies,

$$\begin{cases} 0 = \left[\nabla C_\tau + r\nabla C_S + \nabla C_{SS} \cdot \sigma^A(S) - r\nabla C\right] + \left(\sigma^A - \sigma^B\right) C^B_{SS},\ \text{for all } (\tau, S) \in [t, T) \times \mathbb{R}_{++} \\ \nabla C = 0,\ \text{for all } S \end{cases} \tag{8.8}$$

[3] Moreover, Eqs. (8.7), (8.8) and (8.9) below can be seen as particular cases of the general results given in Chapter 6.

By the maximum principle, again, $\nabla C > 0$ whenever $C_{SS} > 0$. Therefore, it follows that *if option prices are convex in the underlying asset price, then they are also always increasing in the volatility of the underlying asset prices.* Economically, this result follows because volatility changes are mean-preserving spread in this context. We are left to show that $C_{SS} > 0$. Let us differentiate Eq. (8.7) with respect to S. The result is that $Z \equiv H_S = C_{SS}$ satisfies the following partial differential equation,

$$\begin{cases} 0 = Z_\tau + (r + 2\sigma'(S))\, Z_S + Z_{SS}\sigma(S) - (r - \sigma''(S))Z & \text{for all } (\tau, S) \in [t, T) \times \mathbb{R}_{++} \\ H(S, T, T) = \psi''(S) & \text{for all } S \in \mathbb{R}_{++} \end{cases} \tag{8.9}$$

By the maximum principle, we have that

$$H(S, \tau, T) > 0 \text{ for all } (\tau, S) \in [t, T] \times \mathbb{R}_{++}, \text{ whenever } \psi''(S) > 0 \ \forall S \in \mathbb{R}_{++}.$$

That is, we have that *in the scalar diffusion setting, the option price is always convex in the underlying asset price if the terminal payoff is convex in the underlying asset price.* In other terms, the convexity of the terminal payoff propagates to the convexity of the pricing function. Therefore, *if the terminal payoff is convex in the underlying asset price, then the option price is always increasing in the volatility of the underlying asset price.*

8.4.2 Recoverability of the risk-neutral density from option prices

Consider the price of a European call,

$$C(S(t), t, T; K) = P(t, T) \int_0^\infty [S(T) - K]^+ \, dQ(S(T)|\, S(t)) = P(t, T) \int_K^\infty (x - K)\, q(x|\, S(t))\, dx,$$

where Q is the risk-neutral measure and $q(x^+|\, x)dx \equiv dQ(x^+|\, x)$. This is indeed a very general formula, as it does not rely on any parametric assumptions for the dynamics of the price underlying the option contract. Let us differentiate the previous formula with respect to K,

$$\frac{\partial C(S(t), t, T; K)}{\partial K} \Big/ P(t, T) = -\int_K^\infty q(x|\, S(t))\, dx,$$

where we assumed that $\lim_{x\to\infty} x \cdot q(x^+|\, S(t)) = 0$. Let us differentiate again,

$$\frac{\partial^2 C(S(t), t, T; K)}{\partial K^2} \Big/ P(t, T) = q(K|\, S(t)). \tag{8.10}$$

Eq. (8.10) provides a means to "recover" the risk-neutral density using option prices. The Arrow-Debreu state density, $\text{AD}(S^+ = u|\, S(t))$, is given by,

$$\text{AD}\left(S^+ = u \middle|\, S(t)\right) = q\left(S^+ \middle|\, S(t)\right)\Big|_{S^+=u} \Big/ P(t, T) = \left.\frac{\partial^2 C(S(t), t, T; K)}{\partial K^2}\right|_{K=u} \Big/ P(t, T)^2.$$

These results are very useful in applied work.

8.4.3 Hedges and crashes

8.5 Stochastic volatility

8.5.1 Statistical models of changing volatility

One of the most prominent advances in the history of empirical finance is the discovery that financial returns exhibit both temporal dependence in their second order moments and heavy-peaked and tailed distributions. Such a phenomenon was known at least since the seminal work of Mandelbrot (1963) and Fama (1965). However, it was only with the introduction of the autoregressive conditionally heteroscedastic (ARCH) model of Engle (1982) and Bollerslev (1986) that econometric models of changing volatility have been intensively fitted to data.

An ARCH model works as follows. Let $\{y_t\}_{t=1}^N$ be a record of observations on some asset returns. That is, $y_t = \log S_t/S_{t-1}$, where S_t is the asset price, and where we are ignoring dividend issues. The empirical evidence suggests that the dynamics of y_t are well-described by the following model:

$$y_t = a + \epsilon_t; \qquad \epsilon_t | \, F_{t-1} \sim N(0, \sigma_t^2); \qquad \sigma_t^2 = w + \alpha \epsilon_{t-1}^2 + \beta \sigma_{t-1}^2; \tag{8.11}$$

where a, w, α and β are parameters and F_t denotes the information set as of time t. This model is known as the GARCH(1,1) model (Generalized ARCH). It was introduced by Bollerslev (1986), and collapses to the ARCH(1) model introduced by Engle (1982) once we set $\beta = 0$.

ARCH models have played a prominent role in the analysis of many aspects of financial econometrics, such as the term structure of interest rates, the pricing of options, or the presence of time varying risk premia in the foreign exchange market. A classic survey is that in Bollerslev, Engle and Nelson (1994).

The quintessence of ARCH models is to make volatility dependent on the variability of past observations. An alternative formulation initiated by Taylor (1986) makes volatility driven by some unobserved components. This formulation gives rise to the *stochastic volatility* model. Consider, for example, the following stochastic volatility model,

$$\begin{aligned} y_t &= a + \epsilon_t; \qquad \epsilon_t | \, F_{t-1} \sim N(0, \sigma_t^2); \\ \log \sigma_t^2 &= w + \alpha \log \epsilon_{t-1}^2 + \beta \log \sigma_{t-1}^2 + \eta_t; \qquad \eta_t | \, F_{t-1} \sim N(0, \sigma_\eta^2) \end{aligned}$$

where a, w, α, β and σ_η^2 are parameters. The main difference between this model and the GARCH(1,1) model in Eq. (8.11) is that the volatility as of time t is not predetermined by the past forecast error, ϵ_t. Rather, this volatility depends on the realization of the stochastic volatility shock η_t at time t. This makes the stochastic volatility model considerably richer than a simple ARCH model. As for the ARCH models, SV models have also been intensively used, especially following the progress accomplished in the corresponding estimation techniques. The seminal contributions related to the estimation of this kind of models are mentioned in Mele and Fornari (2000). Early contributions that relate changes in volatility of asset returns to economic intuition include Clark (1973) and Tauchen and Pitts (1983), who assume that a stochastic process of information arrival generates a random number of intraday changes of the asset price.

8.5.2 Option pricing implications

Parallel to the empirical research into asset returns volatility, practitioners and option pricing theorists realized that the assumption of constant volatility underlying the Black and Scholes

(1973) and Merton (1973) formulae was too restrictive. The Black-Scholes model assumes that the price of the asset underlying the option contract follows a geometric Brownian motion,

$$\frac{dS(\tau)}{S(\tau)} = \mu d\tau + \sigma dW(\tau),$$

where W is a Brownian motion, and μ, σ are constants. As explained earlier, σ is the only parameter to enter the formula.

The assumption that σ is constant is inconsistent with the empirical evidence. This assumption is also inconsistent with the empirical evidence on the *cross-section* of option prices. Empirically, the "implied volatility", i.e. the value of σ that equates the Black-Scholes formula to the market price of the option, depends on the "moneyness of the option". Let us define the "moneyness of the option" as,

$$mo \equiv \frac{S(t)e^{r(T-t)}}{K},$$

where r is the short-term rate, K is the strike of the option, and T is the maturity date of the option contract. Then, the empirical evidence suggests that the "implied volatility" is $\cup$-shaped in $\frac{1}{mo}$. This phenomenon is known as the *smile effect*.

Ball and Roma (1994) (p. 602) and Renault and Touzi (1996) were the first to point out that a smile effect arises when the asset return exhibits stochastic volatility. In continuous time,[4]

$$\begin{aligned} \frac{dS(\tau)}{S(\tau)} &= \mu d\tau + \sigma(\tau) dW(\tau) \\ d\sigma(\tau)^2 &= b(S(\tau), \sigma(\tau))d\tau + a(S(\tau), \sigma(\tau))dW^\sigma(\tau) \end{aligned} \tag{8.12}$$

where W^σ is another Brownian motion, and b and a are some functions satisfying the usual regularity conditions. In other words, let us suppose that Eqs. (8.12) constitute the data generating process. Then, the fundamental theorem of asset pricing (FTAP, henceforth) tells us that there is *a* probability measure equivalent to P, Q say (the so-called risk-neutral probability measure), such that the *rational* option price $C(S(t), \sigma(t)^2, t, T)$ is given by,

$$C(S(t), \sigma(t)^2, t, T) = e^{-r(T-t)}\mathbb{E}\left[(S(T) - K)^+ \middle| S(t), \sigma(t)^2 \right],$$

where $\mathbb{E}[\cdot]$ is the expectation taken under the probability Q. Next, if we continue to assume that option prices are really given by the previous formula, then, by inverting the Black-Scholes formula produces a "constant" volatility that is $\cup$-shaped with respect to K.

The first option pricing models with stochastic volatility are developed by Hull and White (1987), Scott (1987) and Wiggins (1987). Explicit solutions have always proved hard to derive. If we exclude the approximate solution provided by Hull and White (1987) or the analytical solution provided by Heston (1993*b*),[5] we typically need to derive the option price through

[4]In an important paper, Nelson (1990) shows that under regularity conditions, the GARCH(1,1) model converges in distribution to the solution of the following stochastic differential equation:

$$d\sigma(\tau)^2 = (\omega - \varphi\sigma(\tau)^2)dt + \psi\sigma(\tau)^2 dW^\sigma(\tau),$$

where W^σ is a standard Brownian motion, and ω, φ, and ψ are parameters. See Mele and Fornari (2000) (Chapter 2) for additional results on this kind of convergence. Corradi (2000) develops a critique related to the conditions underlying these convergence results.

[5]The Heston's solution relies on the assumption that stochastic volatility is a linear mean-reverting "square-root" process. In a square root process, the instantaneous variance of the process is proportional to the level reached by that process: in model (8.12), for instance, $a(S, \sigma) = a \cdot \sigma$, where a is a constant. In this case, it is possible to show that the characteristic function is exponential-affine in the state variables s and σ. Given a closed-form solution for the characteristic function, the option price is obtained through standard Fourier methods.

some numerical methods based on Montecarlo simulation or the numerical solution to partial differential equations.

In addition to these important computational details, models with stochastic volatility lead to serious *economic* issues. Typically, the presence of stochastic volatility generates *market incompleteness*. As we pointed out earlier, market incompleteness means that we can not hedge against future contingencies. In our context, market incompleteness arises because the number of the assets available for trading (one) is less than the sources of risk (i.e. the two Brownian motions).[6] In our option pricing problem, there are no portfolios including only the underlying asset and a money market account that could replicate the value of the option at the expiration date. Precisely, let C be the *rationally* formed price at time t, i.e. $C(\tau) = C\left(S(\tau), \sigma(\tau)^2, \tau, T\right)$, where $\sigma(\tau)^2$ is driven by a Brownian motion W^σ, which is different from W. The value of the portfolio that only includes the underlying asset is only driven by the Brownian motion driving the underlying asset price, i.e. it does not include W^σ. Therefore, the value of the portfolio does not factor in all the random fluctuations that move the return volatility, $\sigma(\tau)^2$. Instead, the option price depends on this return volatility as we have assumed that the option price, $C(\tau)$, is rationally formed, i.e. $C(\tau) = C\left(S(\tau), \sigma(\tau)^2, \tau, T\right)$.

In other words, trading with only the underlying asset does not allow for a perfect replication of the option price, C. In turn, rembember, a perfect replication of C is the condition we need to obtain a unique preference-free price of the option. To summarize, the presence of stochastic volatility introduces two inextricable consequences:

- Perfect hedging strategies are impossible.
- There is an infinity of option prices that are compatible with the requirement that there are not arbitrage opportunities.

The next section aims at showing these claims in more detail.[7]
- Post 1987 crash data, smirks and asymmetries.

8.5.3 Stochastic volatility and market incompleteness

Let us suppose that the asset price is solution to Eq. (8.12). To simplify, we assume that W and W^σ are independent. Since C is *rationally* formed, $C(\tau) = C(S(\tau), \sigma(\tau)^2, \tau, T)$. By Itô's lemma,

$$dC = \left[\frac{\partial C}{\partial t} + \mu S C_S + b C_{\sigma^2} + \frac{1}{2}\sigma^2 S^2 C_{SS} + \frac{1}{2}a^2 C_{\sigma^2\sigma^2}\right] d\tau + \sigma S C_S dW + a C_{\sigma^2} dW^\sigma.$$

Next, let us consider a self-financing portfolio that includes (i) one call, (ii) $-\alpha$ shares, and (iii) $-\beta$ units of the money market account (MMA, henceforth). The value of this portfolio is $V = C - \alpha S - \beta P$, and satisfies

$$\begin{aligned} dV &= dC - \alpha dS - \beta dP \\ &= \left[\frac{\partial C}{\partial t} + \mu S\left(C_S - \alpha\right) + b C_{\sigma^2} + \frac{1}{2}\sigma^2 S^2 C_{SS} + \frac{1}{2}a^2 C_{\sigma^2\sigma^2} - r\beta P\right] d\tau + \sigma S\left(C_S - \alpha\right) dW + a C_{\sigma^2} dW^\sigma. \end{aligned}$$

[6] Naturally, markets can be "completed" by the presence of the option. However, in this case the option price is not preference free.

[7] The mere presence of stochastic volatility is not necessarily a source of market incompleteness. Mele (1998) (p. 88) considers a "circular" market with m asset prices, in which (i) the asset price no. i exhibits stochastic volatility, and (ii) this stochastic volatility is driven by the Brownian motion driving the $(i-1)$-th asset price. Therefore, in this market, each asset price is solution to the Eqs. (8.12) and yet, by the previous circular structure, markets are complete.

As is clear, only when $a = 0$, we could zero the volatility of the portfolio value. In this case, we could set $\alpha = C_S$ and $\beta P = C - \alpha S - V$, leaving

$$dV = \left(\frac{\partial C}{\partial t} + bC_{\sigma^2} + \frac{1}{2}\sigma^2 S^2 C_{SS} - rC + rSC_S + rV\right) d\tau = \left(\frac{\partial C}{\partial t} + bC_{\sigma^2} + \frac{1}{2}\sigma^2 S^2 C_{SS} + rSC_S\right) d\tau,$$

where we have used the equality $V = C$. The previous equation shows that the portfolio is locally riskless. Therefore, by the FTAP,

$$0 = \frac{\partial C}{\partial t} + bC_{\sigma^2} + \frac{1}{2}\sigma^2 S^2 C_{SS} + rSC_S - rC.$$

The previous equation generalizes the Black-Scholes equation to the case in which volatility is time-varying and *non*-stochastic, as a result of the assumption that $a = 0$. If $a \neq 0$, return volatility *is* stochastic and, hence, there are no hedging portfolios to use to derive a unique option price. However, we still have the possibility to characterize the price of the option. Indeed, consider a portfolio of (i) *two* calls with different strike prices and maturity dates (with weights 1 and γ), (ii) $-\alpha$ shares, and (iii) $-\beta$ units of the MMA. We denote the price processes of these two calls with C^1 and C^2. The value of this portfolio is $V = C^1 + \gamma C^2 - \alpha S - \beta P$, and satisfies,

$$\begin{aligned} dV &= dC^1 + \gamma dC^2 - \alpha dS - \beta dP \\ &= \left[\mathcal{L}C^1 + \gamma \mathcal{L}C^2 - \alpha\mu S - r\beta P\right] d\tau + \sigma S\left(C_S^1 + \gamma C_S^2 - \alpha\right) dW + a\left(C_{\sigma^2}^1 + \gamma C_{\sigma^2}^2\right) dW^\sigma, \end{aligned}$$

where $\mathcal{L}C^i \equiv \frac{\partial C^i}{\partial t} + \mu SC_S^i + bC_{\sigma^2}^i + \frac{1}{2}\sigma^2 S^2 C_{SS}^i + \frac{1}{2}a^2 C_{\sigma^2\sigma^2}^i$, for $i = 1, 2$. In this context, risk can be eliminated. Indeed, set

$$\gamma = -\frac{C_{\sigma^2}^1}{C_{\sigma^2}^2} \quad \text{and} \quad \alpha = C_S^1 + \gamma C_S^2.$$

The value of the portfolio is solution to,

$$dV = \left(\mathcal{L}C^1 + \gamma \mathcal{L}C^2 - \alpha\mu S + rV + \alpha rS - rC^1 - \gamma rC^2\right) d\tau.$$

Therefore, by the FTAP,

$$\begin{aligned} 0 &= \mathcal{L}C^1 + \gamma \mathcal{L}C^2 - \alpha\mu S + \alpha rS - rC^1 - \gamma rC^2 \\ &= \left[\mathcal{L}C^1 - rC^1 - C_S^1(\mu S - rS)\right] + \gamma\left[\mathcal{L}C^2 - rC^2 - C_S^2(\mu S - rS)\right] \end{aligned}$$

where the second equality follows by the definition of α, and by rearranging terms. Finally, by using the definition of γ, and by rearranging terms,

$$\frac{\mathcal{L}C^1 - rC^1 - C_S^1(\mu S - rS)}{C_{\sigma^2}^1} = \frac{\mathcal{L}C^2 - rC^2 - C_S^2(\mu S - rS)}{C_{\sigma^2}^2}.$$

These ratios agree. So they must be equal to some process $a \cdot \Lambda^\sigma$ (say) independent of both the strike prices and the maturity of the options. Therefore, we obtain that,

$$\frac{\partial C}{\partial t} + rSC_S + \left[b - a\Lambda^\sigma\right] C_{\sigma^2} + \frac{1}{2}\sigma^2 S^2 C_{SS} + \frac{1}{2}a^2 C_{\sigma^2\sigma^2} = rC. \tag{8.13}$$

The economic interpretation of Λ^σ is that of the unit risk-premium required to face the risk of stochastic fluctuations in the return volatility. The problem, the requirement of absence of arbitrage opportunities does not suffice to recover a unique choice of Λ^{σ}. In other words, by the Feynman-Kac stochastic representation of a solution to a PDE, we have that the solution to Eq. (8.13) is,

$$C(S(t),\sigma(t)^2,t,T)=e^{-r(T-t)}\mathbb{E}_{Q^{\Lambda}}\left[\left.(S(T)-K)^{+}\right| S(t),\sigma(t)^2\right], \tag{8.14}$$

where Q^{Λ} is *a* risk-neutral measure.

8.5.4 Pricing formulae

Hull and White (1987) derive the first pricing formula of the stochastic volatility literature. They assume that the return volatility is independent of the asset price, and show that,

$$C(S(t),\sigma(t)^2,t,T)=\mathbb{E}_{\tilde{V}}\left[\mathrm{BS}(S(t),t,T;\tilde{V})\right],$$

where $\mathrm{BS}(S(t),t,T;\tilde{V})$ is the Black-Scholes formula obtained by replacing the constant σ^2 with $\tilde{V}$, and

$$\tilde{V}=\frac{1}{T-t}\int_t^T\sigma(\tau)^2d\tau.$$

This formula tells us that the option price is simply the Black-Scholes formula averaged over all the possible "values" taken by the future average volatility $\tilde{V}$. A proof of this equation is given in the appendix.[8]

The most widely used formula is the Heston's (1993*b*) formula, which holds when return volatility is a square-root process.

[8]The result does not hold in the general case in which the asset price and volatility are correlated. However, Romano and Touzi (1997) prove that a similar result holds in such a more general case.

8.6 Local volatility

8.6.1 Topics & issues

- Stochastic volatility models may EXPLAIN the Smile
- Obviously, stochastic volatility models do not allow for a *perfect* hedge. Their main drawback is that they can not perfectly FIT the Smile.
- Towards the end of 1980s and the beginning of the 1990s, interest rates modelers invented models that allows a perfect fit of the initial yield curve.
- Important for interest rate derivatives.
- In 1993 and 1994, Derman & Kani, Dupire and Rubinstein come up with a technology that could be applied to options on tradables.
- Why is it important to exactly fit the structure of already existing plain vanilla options?
- Plain vanilla versus exotics. Suppose you wish to price exotic, or illiquid, options.
- The model you use to price the illiquid option must predict that the plain vanilla option prices are identical to those your company is selling! How can we trust a model that is not even able to pin down all outstanding contracts? - Arbitrage opportunities for quants and traders?

8.6.2 How does it work?

As usual in this context, we model the dynamics of asset prices directly under the risk-neutral probability. Accordingly, let $\hat{W}$ be a Brownian motion under the risk-neutral probability, and $\mathbb{E}$ the expection operator under the risk-neutral probability.

1. Start with a set of actively traded (i.e. liquid) European options. Let K and T be strikes and time-to-maturity. Let us be given a collection of prices:

$$C_{\$}(K,T) \equiv C(K,T), \quad K,T \text{ varying.}$$

2. Is it mathematically possible to conceive a diffusion process,

$$\frac{dS_t}{S_t} = rdt + \sigma(S_t,t)d\hat{W}_t,$$

such that the initial collection of European option prices is predicted without errors by the resulting model?

3. Yes. We should use the function σ_{loc}, say, that takes the form,

$$\sigma_{\text{loc}}(K,T) = \sqrt{2\frac{\dfrac{\partial C(K,T)}{\partial T} + rK\dfrac{\partial C(K,T)}{\partial K}}{K^2\dfrac{\partial^2 C(K,T)}{\partial K^2}}}. \tag{8.15}$$

We call this function σ_{loc} "local volatility".

4. Now, we can price the illiquid options through numerical methods. For example, we can use simulations. In the simulations, we use

$$\frac{dS_t}{S_t} = rdt + \sigma_{\text{loc}}(S_t, t)\, d\hat{W}_t.$$

- It turns out, empirically, that $\sigma_{\text{loc}}(x, t)$ is typically decreasing in x for fixed t, a phenomenon known as the Black-Christie-Nelson leverage effect. This fact leads some practitioners to assume from the outset that $\sigma(x, t) = x^{\alpha} f(t)$, for some function f and some constant $\alpha < 0$. This gives rise to the so-called CEV (**C**onstant **E**lasticity of **V**ariance) model.

- More recently, practitioners use models that combine "local vols" and "stoch vol", such as

$$\begin{aligned} \frac{dS_t}{S_t} &= rdt + \sigma(S_t, t) \cdot v_t \cdot d\hat{W}_t \\ dv_t &= \phi(v_t)dt + \psi(v_t)d\hat{W}_t^v \end{aligned} \tag{8.16}$$

  where $\hat{W}^v$ is another Brownian motion, and ϕ, ψ are some functions (ϕ includes a risk-premium). It is possible to show that in this specific case,

$$\tilde{\sigma}_{\text{loc}}(K, T) = \frac{\sigma_{\text{loc}}(K, T)}{\sqrt{\mathbb{E}\left(v_T^2 \middle| S_T\right)}} \tag{8.17}$$

  would be able to pin down the initial structure of European options prices. (Here $\sigma_{\text{loc}}(K, T)$ is as in Eq. (8.15).)

- In this case, we simulate

$$\begin{cases} \dfrac{dS_t}{S_t} = rdt + \hat{\sigma}_{\text{loc}}(S_t, t) \cdot v_t \cdot d\hat{W}_t \\ dv_t = \phi(v_t)dt + \psi(v_t)d\hat{W}_t^v \end{cases}$$

8.6.3 Variance swaps

- In fact, it is possible to demonstrate the following general result. Let S_t satisfy,

$$\frac{dS_t}{S_t} = rdt + \sigma_t d\hat{W}_t,$$

  where σ_t is $\mathcal{F}_t$-adapted (i.e. $\mathcal{F}_t$ can be larger than $\mathcal{F}_t^S \equiv \sigma(S_\tau : \tau \leq t)$). Then,

$$e^{-r(T-t)}\mathbb{E}\left(\sigma_T^2\right) = 2\int \frac{\frac{\partial C(K,T)}{\partial T} + rK\frac{\partial C(K,T)}{\partial K}}{K^2} dK. \tag{8.18}$$

- The previous developments can be used to address very important issues. Define the total "integrated" variance within the time interval $[T_1, T_2]$ ($T_1 > t$) to be

$$IV_{T_1,T_2} \equiv \int_{T_1}^{T_2} \sigma_u^2 du.$$

For reasons developed below, let us compute the risk-neutral expectation of such a "realized" variance. This can easily be done. If $r = 0$, then by Eq. (8.18),

$$\mathbb{E}\left(IV_{T_1,T_2}\right) = 2\int \frac{C_t\left(K,T_2\right) - C_t\left(K,T_1\right)}{K^2} dK, \tag{8.19}$$

where $C_t\left(K,T\right)$ is the price as of time t of a call option expiring at T and struck at K. A proof of Eq. (8.19) is in the appendix.

- If $r > 0$ and $T_1 = t$, $T_2 \equiv T$, we have,

$$\mathbb{E}\left(IV_{t,T}\right) = 2\left[\int_0^{F_t} \frac{P_t\left(K,T\right)}{P\left(t,T\right)} \frac{1}{K^2} dK + \int_{F_t}^{\infty} \frac{C_t\left(K,T\right)}{u\left(t,T\right)} \frac{1}{K^2} dK\right], \tag{8.20}$$

  where F_t is the forward price: $F_t = P\left(t,T\right) S_t$; $P_t\left(K,T\right)$ is the price as of time t of a put option expiring at T and struck at K and, as usual, $P\left(t,T\right)$ is the price as of time t of a pure discount bond expiring at T. A proof of Eq. (8.20) is in the appendix.

- In September 2003, the Chicago Board Option Exchange (CBOE) changed its stochastic volatility index VIX to approximate the variance swap rate of the S&P 500 index return (for 30 days). In March 2004, the CBOE launched the CBOE Future Exchange for trading futures on the new VIX. Options on VIX are also forthcoming.

  A *variance swap* is a contract that has zero value at entry (at t). At maturity T, the buyer of the swap receives,

$$\left(IV_{t,T} - SW_{t,T}\right) \times \text{notional},$$

  where $SW_{t,T}$ is the swap rate established at t and paid off at time T. Therefore, if r is deterministic,

$$SW_{t,T} = \mathbb{E}\left(IV_{t,T}\right),$$

  where $\mathbb{E}\left(IV_{t,T}\right)$ is given by Eq. (8.20). Therefore, (8.20) is used to evaluate these variance swaps.

- Finally, it is worth that we mention that the previous contracts rely on some notions of *realized* volatility as a continuous record of returns is obviously unavailable.

8.7 American options

8.8 Exotic options

8.9 Market imperfections

8.10 Appendix 1: Additional details on the Black & Scholes formula

8.10.1 The original argument

In the original approach followed by Black and Scholes (1973) and Merton (1973), it is assumed that the option is already traded. Let $dS/S = \mu d\tau + \sigma dW$. Create a self-financing portfolio of $\bar{n}_S$ units of the underlying asset and n_C units of the European call option, where $\bar{n}_S$ is an arbitrary number. Such a portfolio is worth $V = \bar{n}_S S + n_C C$ and since it is self-financing it satisfies:

$$\begin{aligned} dV &= \bar{n}_S dS + n_C dC \\ &= \bar{n}_S dS + n_C \left[C_S dS + \left(C_\tau + \frac{1}{2}\sigma^2 S^2 C_{SS} \right) d\tau \right] \\ &= (\bar{n}_S + n_C C_S)\, dS + n_C \left(C_\tau + \frac{1}{2}\sigma^2 S^2 C_{SS} \right) d\tau \end{aligned}$$

where the second line follows from Itô's lemma. Therefore, the portfolio is locally riskless whenever

$$n_C = -\bar{n}_S \frac{1}{C_S},$$

in which case V must appreciate at the r-rate

$$\frac{dV}{V} = \frac{n_C \left(C_\tau + \frac{1}{2}\sigma^2 S^2 C_{SS} \right) d\tau}{\bar{n}_S S + n_C C} = \frac{-\frac{1}{C_S} \left(C_\tau + \frac{1}{2}\sigma^2 S^2 C_{SS} \right)}{S - \frac{1}{C_S} C} d\tau = r d\tau.$$

That is,

$$\begin{cases} 0 = C_\tau + \frac{1}{2}\sigma^2 S^2 C_{SS} + rSC_S - rC, \text{ for all } (\tau, S) \in [t, T) \times \mathbb{R}_{++} \\ C(x, T, T) = (x - K)^+, \text{ for all } x \in \mathbb{R}_{++} \end{cases}$$

which is the Black-Scholes partial differential equation.

8.10.2 Some useful properties

We have

$$\frac{\partial \mathrm{BS}}{\partial S} = N(d_1).$$

The proof follows by really simple algebra. We have,

$$\mathrm{BS}(S, KP, \sigma) = SN(d_1) - KPN(d_2),$$

where $d_1 \equiv \frac{x + \frac{1}{2}\sigma^2 (T-t)}{\sigma\sqrt{T-t}}$, $d_2 = d_1 - \sigma\sqrt{T-t}$ and $x = \log\left(\frac{S}{KP}\right)$. Differentiating with respect to S yields:

$$\frac{\partial \mathrm{BS}}{\partial S} = N(d_1) + SN'(d_1)\frac{\partial d_1}{\partial S} - KPN'(d_2)\frac{\partial d_2}{\partial S} = N(d_1) + \left(SN'(d_1) - KPN'(d_2) \right) \frac{1}{S\sigma\sqrt{T-t}},$$

where

$$N'(d_2) = \frac{1}{\sqrt{2\pi}} e^{-\frac{1}{2}(d_2)^2} = \frac{1}{\sqrt{2\pi}} e^{-\frac{1}{2}\left(d_1^2 + \sigma^2 (T-t) - 2d_1\sigma\sqrt{T-t}\right)} = N'(d_1) e^{-\frac{1}{2}\sigma^2 (T-t) + d_1\sigma\sqrt{T-t}}$$

Hence,

$$\frac{\partial \mathrm{BS}}{\partial S} = N(d_1) + \left(S - KPe^{-\frac{1}{2}\sigma^2 (T-t) + d_1\sigma\sqrt{T-t}} \right) \frac{N'(d_1)}{S\sigma\sqrt{T-t}} = N(d_1) + (S - KPe^x) \frac{N'(d_1)}{S\sigma\sqrt{T-t}} = N(d_1).$$

8.11 Appendix 2: Stochastic volatility

8.11.1 Proof of the Hull and White (1987) equation

By the law of iterated expectations, (8.14) can be written as:

$$\begin{aligned} C(S(t),\sigma(t)^2,t,T) &= e^{-r(T-t)}\mathbb{E}\left[[S(T)-K]^+ \middle| S(t),\sigma(t)^2 \right] \\ &= \mathbb{E}\left\{ \mathbb{E}\left[e^{-r(T-t)} [S(T)-K]^+ \middle| S(t), \left\{\sigma(\tau)^2\right\}_{\tau\in[t,T]} \right] \middle| S(t),\sigma(t)^2 \right\} \\ &= \mathbb{E}\left[\mathrm{BS}\left(S(t),t,T;\tilde{V}\right) \middle| S(t),\sigma(t)^2 \right] \\ &= \mathbb{E}\left[\mathrm{BS}\left(S(t),t,T;\tilde{V}\right) \middle| \sigma(t)^2 \right] \\ &= \int \mathrm{BS}\left(S(t),t,T;\tilde{V}\right) \Pr\left(\tilde{V} \middle| \sigma(t)^2\right) d\tilde{V} \\ &\equiv \mathbb{E}_{\tilde{V}}\left[\mathrm{BS}\left(S(t),t,T;\tilde{V}\right) \right], \end{aligned} \tag{8.21}$$

where $\Pr(\tilde{V}\,|\,\sigma(t)^2)$ is the density of $\tilde{V}$ conditional on the current volatility value $\sigma(t)^2$.

In other terms, the price of an option on an asset with stochastic volatility is the expectation of the Black-Scholes formula over the distribution of the average (random) volatility $\tilde{V}$. To understand better this result, all we have to understand is that *conditionally on the volatility path* $\left\{\sigma(\tau)^2\right\}_{\tau\in[t,T]}$, $\log\left(\frac{S(T)}{S(t)}\right)$ is normally distributed under the risk-neutral probability measure. To see this, note that under the risk-neutral probability measure,

$$\log\left(\frac{S(T)}{S(t)}\right) = r(T-t) - \frac{1}{2}\int_t^T \sigma(\tau)^2 d\tau + \int_t^T \sigma(\tau) dW(\tau).$$

Therefore, conditionally upon the volatility path $\{\sigma(\tau)\}_{\tau\in[t,T]}$,

$$\mathbb{E}\left[\log\left(\frac{S(T)}{S(t)}\right)\right] = r(T-t) - \frac{1}{2}(T-t)\tilde{V} \quad \text{and} \quad var\left[\log\left(\frac{S(T)}{S(t)}\right)\right] = \int_t^T \sigma(\tau)^2 d\tau = (T-t)\tilde{V}.$$

This shows the claim. It also shows that the Black-Scholes formula can be applied to compute the inner expectation of the second line of (8.21). And this produces the third line of (8.21). The fourth line is trivial to obtain. Given the result of the third line, the only thing that matters in the remaining conditional distribution is the conditional probability $\Pr(\tilde{V} \mid \sigma(t)^2)$, and we are done.

8.11.2 Simple smile analytics

8.12 Appendix 3: Technical details for local volatility models

In all the proofs to follow, all expectations are taken to be expectations conditional on $\mathcal{F}_t$. However, to simplify notation, we simply write $\mathbb{E}(\cdot|\cdot) \equiv \mathbb{E}(\cdot|\cdot, \mathcal{F}_t)$.

PROOF OF EQS. (8.17) AND (8.18). We first derive Eq. (8.17), a result encompassing Eq. (8.15). By assumption,

$$\frac{dS_t}{S_t} = rdt + \sigma_t d\hat{W}_t,$$

where σ_t is some $\mathcal{F}_t$-adapted process. For example, $\sigma_t \equiv \sigma(S_t, t) \cdot v_t$, all t, where v_t is solution to the 2nd equation in (8.16). Next, by assumption we are observing a set of option prices $C(K,T)$ with a continuum of strikes K and maturities T. We have,

$$C(K,T) = e^{-r(T-t)} \mathbb{E}(S_T - K)^+, \tag{8.22}$$

and

$$\frac{\partial}{\partial K} C(K,T) = -e^{-r(T-t)} \mathbb{E}(\mathbb{I}_{S_T \geq K}). \tag{8.23}$$

For fixed K,

$$d_T (S_T - K)^+ = \left[\mathbb{I}_{S_T \geq K} r S_T + \frac{1}{2}\delta(S_T - K)\sigma_T^2 S_T^2\right] dT + \mathbb{I}_{S_T \geq K}\sigma_T S_T d\hat{W}_T,$$

where δ is the Dirac's delta. Hence, by the decomposition $(S_T - K)^+ + K\mathbb{I}_{S_T \geq K} = S_T \mathbb{I}_{S_T \geq K}$,

$$\frac{d\mathbb{E}(S_T - K)^+}{dT} = r\left[\mathbb{E}(S_T - K)^+ + K\mathbb{E}(\mathbb{I}_{S_T \geq K})\right] + \frac{1}{2}\mathbb{E}\left[\delta(S_T - K)\sigma_T^2 S_T^2\right].$$

By multiplying throughout by $e^{-r(T-t)}$, and using (8.22)-(8.23),

$$e^{-r(T-t)}\frac{d\mathbb{E}(S_T - K)^+}{dT} = r\left[C(K,T) - K\frac{\partial C(K,T)}{\partial K}\right] + \frac{1}{2}e^{-r(T-t)}\mathbb{E}\left[\delta(S_T - K)\sigma_T^2 S_T^2\right]. \tag{8.24}$$

We have,

$$\begin{aligned}
\mathbb{E}\left[\delta(S_T - K)\sigma_T^2 S_T^2\right] &= \iint \delta(S_T - K)\sigma_T^2 S_T^2 \underbrace{\phi_T(\sigma_T | S_T)\phi_T(S_T)}_{\equiv \text{ joint density of } (\sigma_T, S_T)} dS_T d\sigma_T \\
&= \int \sigma_T^2 \left[\int \delta(S_T - K) S_T^2 \phi_T(S_T)\phi_T(\sigma_T | S_T) dS_T\right] d\sigma_T \\
&= K^2 \phi_T(K) \int \sigma_T^2 \phi_T(\sigma_T | S_T = K) d\sigma_T \\
&\equiv K^2 \phi_T(K)\mathbb{E}\left[\sigma_T^2 \middle| S_T = K\right].
\end{aligned}$$

By replacing this result into Eq. (8.24), and using the famous relation

$$\frac{\partial^2 C(K,T)}{\partial K^2} = e^{-r(T-t)}\phi_T(K) \tag{8.25}$$

(which easily follows by differentiating once again Eq. (8.23)), we obtain

$$e^{-r(T-t)}\frac{d\mathbb{E}(S_T - K)^+}{dT} = r\left[C(K,T) - K\frac{\partial C(K,T)}{\partial K}\right] + \frac{1}{2}K^2\frac{\partial^2 C(K,T)}{\partial K^2}\mathbb{E}\left[\sigma_T^2 \middle| S_T = K\right]. \tag{8.26}$$

We also have,

$$\frac{\partial}{\partial T}C(K,T) = -rC(K,T) + e^{-r(T-t)}\frac{\partial \mathbb{E}(S_T-K)^+}{\partial T}.$$

Therefore, by replacing the previous equality into Eq. (8.26), and by rearranging terms,

$$\frac{\partial}{\partial T}C(K,T) = -rK\frac{\partial C(K,T)}{\partial K} + \frac{1}{2}K^2\frac{\partial^2 C(K,T)}{\partial K^2}\mathbb{E}\left[\sigma_T^2\middle| S_T = K\right].$$

This is,

$$\mathbb{E}\left[\sigma_T^2\middle| S_T=K\right] = 2\frac{\dfrac{\partial C(K,T)}{\partial T} + rK\dfrac{\partial C(K,T)}{\partial K}}{K^2\dfrac{\partial^2 C(K,T)}{\partial K^2}} \equiv \sigma_{\text{loc}}(K,T)^2. \tag{8.27}$$

For example, let $\sigma_t \equiv \sigma(S_t,t)\cdot v_t$, where v_t is solution to the 2nd equation in (8.16). Then,

$$\begin{aligned}\sigma_{\text{loc}}(K,T)^2 &= \mathbb{E}\left[\sigma_T^2\middle| S_T=K\right]\\ &= \mathbb{E}\left[\sigma(S_T,T)^2\cdot v_T^2\middle| S_T=K\right] = \sigma(K,T)^2\mathbb{E}\left[v_T^2\middle| S_T=K\right]\\ &\equiv \tilde{\sigma}_{\text{loc}}(K,T)^2\,\mathbb{E}\left(v_T^2\middle| S_T=K\right),\end{aligned}$$

which proves (8.17).

Next, we prove Eq. (8.18). We have,

$$\begin{aligned}\mathbb{E}\left(\sigma_T^2\right) &= \int \mathbb{E}\left[\sigma_T^2\middle| S_T=K\right]\phi_T(K)\,dK\\ &= 2\int\frac{\frac{\partial C(K,T)}{\partial T} + rK\frac{\partial C(K,T)}{\partial K}}{K^2\frac{\partial^2 C(K,T)}{\partial K^2}}\phi_T(K)\,dK\\ &= 2e^{r(T-t)}\int\frac{\frac{\partial C(K,T)}{\partial T} + rK\frac{\partial C(K,T)}{\partial K}}{K^2}dK\end{aligned}$$

where the 2nd line follows by Eq. (8.27), and the third line follows by Eq. (8.25). This proves Eq. (8.18). ■

Proof of Eq. (8.19). If $r=0$, Eq. (8.18) collapses to,

$$\mathbb{E}\left(\sigma_T^2\right) = 2\int\frac{\frac{\partial C(K,T)}{\partial T}}{K^2}dK.$$

Then, we have,

$$\mathbb{E}\left(IV_{T_1,T_2}\right) = \int_{T_1}^{T_2}\mathbb{E}\left(\sigma_u^2\right)du = 2\int\frac{1}{K^2}\left[\int_{T_1}^{T_2}\frac{\partial C(K,u)}{\partial T}du\right]dK = 2\int\frac{C(K,T_2)-C(K,T_1)}{K^2}dK.$$

Proof of Eq. (8.20). By the standard Taylor expansion with remainder, we have that for any function f smooth enough,

$$f(x) = f(x_0) + f'(x_0)(x-x_0) + \int_{x_0}^{x}(x-t)f''(t)\,dt.$$

By applying this formula to $\log F_T$,

$$\begin{aligned}\log F_T &= \log F_t + \frac{1}{F_t}(F_T - F_t) - \int_{F_t}^{F_T} (F_T - t)\frac{1}{t^2}dt \\ &= \log F_t + \frac{1}{F_t}(F_T - F_t) - \int_0^{F_t} (K - F_T)^+ \frac{1}{K^2}dK - \int_{F_t}^{\infty} (F_T - K)^+ \frac{1}{K^2}dK \\ &= \log F_t + \frac{1}{F_t}(F_T - F_t) - \int_0^{F_t} (K - S_T)^+ \frac{1}{K^2}dK - \int_{F_t}^{\infty} (S_T - K)^+ \frac{1}{K^2}dK,\end{aligned}$$

where the third equality follows because the forward price at T satisfies $F_T = S_T$.[9] Hence by $\mathbb{E}(F_T) = F_t$,

$$-\mathbb{E}\left(\log\frac{F_T}{F_t}\right) = \int_0^{F_t} \frac{P_t(K,T)}{u(t,T)}\frac{1}{K^2}dK + \int_{F_t}^{\infty} \frac{C_t(K,T)}{u(t,T)}\frac{1}{K^2}dK. \tag{8.28}$$

On the other hand, by Itô's lemma,

$$\mathbb{E}\left(\int_t^T \sigma_u^2 du\right) = -2\mathbb{E}\left(\log\frac{F_T}{F_t}\right). \tag{8.29}$$

By replacing this formula into Eq. (8.28) yields Eq. (8.20).[10] ■

Remark A1. Eqs. (8.28) and (8.29) reveal that variance swaps can be hedged!

Remark A2. Set for simplicity $r = 0$. In the previous proofs, it was argued that if $\frac{dC(K,T)}{dT} = \frac{d\mathbb{E}(S_T - K)^+}{dT}$, then volatility must be restricted in a way to make $\sigma^2 = 2\,\frac{\partial C(K,T)}{\partial T}\Big/ K^2 \frac{\partial^2 C(K,T)}{\partial K^2}$. The converse is also true. By Fokker-Planck,

$$\frac{1}{2}\frac{\partial^2}{\partial x^2}\left(x^2\sigma^2\phi\right) = \frac{\partial}{\partial t}\phi$$

(t, x forward). If we ignore ill-posedness issues[11] related to Eq. (8.25), $\phi = \frac{\partial^2 C}{\partial x^2}$. Replacing $\sigma^2 = 2\,\frac{\partial C(x,T)}{\partial T}\Big/ x^2\frac{\partial^2 C(x,T)}{\partial x^2}$ into the Fokker-Planck equation,

$$\frac{\partial^2}{\partial x^2}\left(\frac{\frac{\partial C(x,T)}{\partial T}}{\frac{\partial^2 C(x,T)}{\partial x^2}}\phi\right) = \frac{\partial}{\partial t}\phi,$$

which works for $\phi = \frac{\partial^2 C}{\partial x^2}$.

[9] The second equality follows because $\int_{x_0}^{x}(x-t)\frac{1}{t^2}dt = \int_0^{x_0}(t-x)^+\frac{1}{t^2}dt + \int_{x_0}^{\infty}(x-t)^+\frac{1}{t^2}dt$.

[10] Note, the previous proof is valid for the case of constant r. If r is stochastic, the formula would be different because $C_t(K,T) = P(t,T)\,E\left[P(t,T)^{-1}\exp(-\int_t^T r_s ds)\,(S_T - K)^+\right] = P(t,T)\,E_{\mathbb{Q}_F^T}(S_T - K)^+$, where $\mathbb{Q}_F^T$ is the T-forward measure introduced in the next chapter.

[11] A reference to deal with these issues is Tikhonov and Arsenin (1977).

References

Ball, C.A. and A. Roma (1994). "Stochastic Volatility Option Pricing." *Journal of Financial and Quantitative Analysis* 29: 589-607.

Bergman, Y.Z., B.D. Grundy, and Z. Wiener (1996). "General Properties of Option Prices," *Journal of Finance* 51: 1573-1610.

Black, F. and M. Scholes (1973). "The Pricing of Options and Corporate Liabilities." *Journal of Political Economy* 81: 637-659.

Bollerslev, T. (1986). "Generalized Autoregressive Conditional Heteroskedasticity." *Journal of Econometrics* 31: 307-327.

Bollerslev, T., Engle, R. and D. Nelson (1994). "ARCH Models." In: McFadden, D. and R. Engle (eds.), *Handbook of Econometrics*, Volume 4: 2959-3038. Amsterdam, North-Holland

Clark, P.K. (1973). "A Subordinated Stochastic Process Model with Fixed Variance for Speculative Prices." *Econometrica* 41: 135-156.

Corradi, V. (2000). "Reconsidering the Continuous Time Limit of the GARCH(1,1) Process." *Journal of Econometrics* 96: 145-153.

Engle, R.F. (1982). "Autoregressive Conditional Heteroskedasticity with Estimates of the Variance of United Kingdom Inflation." *Econometrica* 50: 987-1008.

Fama, E. (1965). "The Behaviour of Stock Market Prices." *Journal of Business*: 38: 34-105.

Heston, S.L. (1993*a*). "Invisible Parameters in Option Prices." *Journal of Finance* 48: 933-947.

Heston, S.L. (1993*b*). "A Closed Form Solution for Options with Stochastic Volatility with Application to Bond and Currency Options." *Review of Financial Studies* 6: 327-344.

Hull, J. and A. White (1987). "The Pricing of Options with Stochastic Volatilities." *Journal of Finance* 42: 281-300.

Mandelbrot, B. (1963). "The Variation of Certain Speculative Prices." *Journal of Business* 36: 394-419.

Mele, A. (1998). *Dynamiques non linéaires, volatilité et équilibre*. Paris: Editions Economica.

Mele, A. and F. Fornari (2000). *Stochastic volatility in financial markets. Crossing the Bridge to Continuous Time*. Boston: Kluwer Academic Publishers.

Merton, R. (1973). "Theory of Rational Option Pricing." *Bell Journal of Economics and Management Science* 4: 637-654

Nelson, D.B. (1990). "ARCH Models as Diffusion Approximations." *Journal of Econometrics* 45: 7-38.

Renault, E. (1997). "Econometric Models of Option Pricing Errors." In Kreps, D., Wallis, K. (eds.), *Advances in Economics and Econometrics*, vol. 3: 223-278. Cambridge: Cambridge University Press.

Romano, M. and N. Touzi (1997). "Contingent Claims and Market Completeness in a Stochastic Volatility Model." *Mathematical Finance* 7: 399-412.

Scott, L. (1987). "Option Pricing when the Variance Changes Randomly: Theory, Estimation, and an Application." *Journal of Financial and Quantitative Analysis* 22: 419-438.

Tauchen, G. and M. Pitts (1983). "The Price Variability-Volume Relationship on Speculative Markets." *Econometrica* 51: 485-505.

Taylor, S. (1986). *Modeling Financial Time Series*. Chichester, UK: Wiley.

Tikhonov, A. N. and V. Y. Arsenin (1977). *Solutions to Ill-Posed Problems*, Wiley, New York.

Wiggins, J. (1987). "Option Values and Stochastic Volatility. Theory and Empirical Estimates." *Journal of Financial Economics* 19: 351-372.

9
Interest rates

9.1 Prices and interest rates

9.1.1 Introduction

A pure-discount (or zero-coupon) bond is a contract that guarantees one unit of *numéraire* at some maturity date. All bonds studied in the first three sections of this chapter are pure-discount. Therefore, we will omit to qualify them as "pure-discount". We the exception of Section 9.3.6, we assume no default risk.

Let $[t, T]$ be a fixed time interval, and $P(\tau, T)$ ($\tau \in [t, T]$) be the price as of time τ of a bond maturing at $T > t$. The information structure assumed in this chapter is the Brownian information structure (except for Section 9.3.4).[1] In many models, the value of the bond price $P(\tau, T)$ is driven by some underlying multidimensional diffusion process $\{y(\tau)\}_{\tau \geq t}$. We will emphasize this fact by writing $P(y(\tau), \tau, T) \equiv P(\tau, T)$. As an example, y can be a scalar diffusion, and $r = y$ can be the short-term rate. In this particular example, bond prices are driven by short-term rate movements through the *bond pricing function* $P(r, \tau, T)$. The exact functional form of the pricing function is determined by (i) the assumptions made as regards the short-term rate dynamics and (ii) the Fundamental Theorem of Asset Pricing (henceforth, FTAP). The bond pricing function in the general multidimensional case is obtained following the same route. Models of this kind are presented in Section 9.3.

A second class of models is that in which bond prices can *not* be expressed as a function of any state variable. Rather, current bond prices are taken as primitives, and forward rates (i.e., interest rates prevailing today for borrowing in the future) are multidimensional diffusion processes. In the absence of arbitrage, there is a precise relation between bond prices and forward rates. The FTAP will then be used to restrict the dynamic behavior of future bond prices and forward rates. Models belonging to this second class are presented in Section 9.4.

The aim of this section is to develop the simplest foundations of the previously described two approaches to interest rate modeling. In the next section, we provide definitions of interest

[1] That is, we will consider a probability space $(\Omega, \mathcal{F}, \mathbb{F}, P)$, $\mathbb{F} = \{\mathcal{F}(\tau)\}_{\tau \in [t,T]}$, and take $\mathcal{F}(\tau)$ to be the information released by (technically, the filtration generated by) a d-dimensional Brownian motion. Here P is the "physical" probability measure, not the "risk-neutral" probability measure.

rates and markets. Section 9.2.3 develops the two basic representations of bond prices: one in terms of the short-term rate; and the other in terms of forward rates. Section 9.2.4 develops the foundations of the so-called forward martingale measure, which is a probability measure under which forward interest rates are martingales. The forward martingale measure is an important tool of analysis in the interest rate derivatives literature.

9.1.2 Markets and interest rate conventions

There are three main types of markets for interest rates: 1. LIBOR; 2. Treasure rate; 3. Repo rate (or repurchase agreement rate).

- *LIBOR* (London Interbank Offer Rate). Many large financial institutions trade with each other T-deposits (for $T = 1$ month to 12 months) at a given currency. The LIBOR is the rate at which these financial institutions are willing to lend money, and the LIBID (London Interbank Bid Rate) is the rate that these financial institutions are prepared to pay to borrow money. Normally, LIBID < LIBOR. The LIBOR is a fundamental point of reference. Typically, financial institutions look at the LIBOR as an opportunity cost of capital.

- *Treasury rate.* This is the rate at which a given Government can borrow at a given currency.

- *Repo rate* (or repurchase agreement rate). A Repo agreement is a contract by which one counterparty sells some assets to the other one, with the obligation to buy these assets back at some future date. The assets act as collateral. The rate at which such a transaction is made is the repo rate. One day repo agreements give rise to *overnight repos.* Longer-term agreements give rise to *term repos.*

We also have mathematical definitions of interest rates. The simplest definition is that of *simply compounded interest rates.* A simply-compounded interest rate at time τ, for the time interval $[\tau, T]$, is defined as the solution L to the following equation:

$$P(\tau, T) = \frac{1}{1 + (T - \tau) L(\tau, T)}.$$

This definition is intuitive, and is the most widely used in the market practice. As an example, LIBOR rates are computed in this way (in this case, $P(\tau, T)$ is generally interpreted as the initial amount of money to invest at time τ to obtain £ 1 at time T).

Given $L(\tau, T)$, the *short-term rate* process r is obtained as:

$$r(\tau) \equiv \lim_{T \downarrow \tau} L(\tau, T).$$

We now introduce an important piece of notation that will be used in Section 9.7. Given a non decreasing sequence of dates $\{T_i\}_{i=0,1,\cdots}$, we define:

$$L(T_i) \equiv L(T_i, T_{i+1}). \tag{9.1}$$

In other terms, $L(T_i)$ is solution to:

$$P(T_i, T_{i+1}) = \frac{1}{1 + \delta_i L(T_i)}, \tag{9.2}$$

where

$$\delta_i \equiv T_{i+1} - T_i, \quad i = 0, 1, \cdots.$$

9.1.3 Bond price representations, yield-curve and forward rates

9.1.3.1 A first representation of bond prices

Let Q be a risk-neutral probability measure. Let $\mathbb{E}[\cdot]$ denote the expectation operator taken under Q. By the FTAP, there are no arbitrage opportunities if and only if $P(\tau, T)$ satisfies:

$$P(\tau, T) = \mathbb{E}\left[e^{-\int_\tau^T r(\ell)d\ell}\right], \quad \text{all } \tau \in [t, T]. \tag{9.3}$$

A sketch of the if-part (there is no arbitrage if bond prices are as in Eq. (9.3)) is provided in Appendix 1. The proof is standard and in fact, similar to that offered in chapter 5, but it is offered again because it allows to disentangle some key issues that arise in the term-structure field.

A widely used concept is the *yield-to-maturity* $R(t, T)$, defined by,

$$P(t, T) \equiv \exp\left(-(T-t) \cdot R(t, T)\right). \tag{9.4}$$

It's a sort of "average rate" for investing from time t to time $T > t$. The function,

$$T \mapsto R(t, T)$$

is called the *yield curve*, or *term-structure of interest rates.*

9.1.3.2 Forward rates, and a second representation of bond prices

In a *forward rate agreement* (FRA, henceforth), two counterparties agree that the interest rate on a given principal in a future time-interval $[T, S]$ will be fixed at some level K. Let the principal be normalized to one. The FRA works as follows: at time T, the first counterparty receives £1 from the second counterparty; at time $S > T$, the first counterparty repays back £ $[1 + 1 \cdot (S - T) K]$ to the second counterparty. The amount K is agreed upon at time t. Therefore, the FRA makes it possible to lock-in future interest rates. We consider simply compounded interest rates because this is the standard market practice.

The amount K for which the current value of the FRA is zero is called the *simply-compounded forward rate* as of time t for the time-interval $[T, S]$, and is usually denoted as $F(t, T, S)$. A simple argument can be used to express $F(t, T, S)$ in terms of bond prices. Consider the following portfolio implemented at time t. Long one bond maturing at T and short $P(t, T)/P(t, S)$ bonds maturing at S. The current cost of this portfolio is zero because,

$$-P(t, T) + \frac{P(t, T)}{P(t, S)} P(t, S) = 0.$$

At time T, the portfolio yields £1 (originated from the bond purchased at time t). At time S, $P(t, T)/P(t, S)$ bonds maturing at S (that were shorted at t) must be purchased. But at time S, the cost of purchasing $P(t, T)/P(t, S)$ bonds maturing at S is obviously £ $P(t, T)/P(t, S)$. The portfolio, therefore, is acting as a FRA: it pays £1 at time T, and $-$£ $P(t, T)/P(t, S)$ at time S. In addition, the portfolio costs nothing at time t. Therefore, the interest rate implicitly paid in the time-interval $[T, S]$ must be equal to the forward rate $F(t, T, S)$, and we have:

$$\frac{P(t, T)}{P(t, S)} = 1 + (S - T)F(t, T, S). \tag{9.5}$$

Clearly,

$$L(T,S) = F(T,T,S).$$

Naturally, Eq. (9.5) can be derived through a simple application of the FTAP. We now show how to perform this task and at the same time, we derive the value of the FRA in the general case in which $K \neq F(t,T,S)$. Consider the following strategy. At time t, enter a FRA for the time-interval $[T,S]$ as a future lender. Come time T, honour the FRA by borrowing £1 for the time-interval $[T,S]$ at the random interest rate $L(T,S)$. The time S payoff deriving from this strategy is:

$$\pi(S) \equiv (S-T)\left[K - L(T,S)\right].$$

What is the current market value of this future, random payoff? Clearly, it's simply the value of the FRA, which we denote as $A(t,T,S;K)$. By the FTAP, then, there are no arbitrage opportunities if and only if $A(t,T,S;K) = \mathbb{E}\left[e^{-\int_t^S r(\tau)d\tau}\pi(S)\right]$, and we have:

$$\begin{aligned}
A(t,T,S;K) &= (S-T)P(t,S)K - (S-T)\mathbb{E}\left[e^{-\int_t^S r(\tau)d\tau}L(T,S)\right] \\
&= \left[1+(S-T)K\right]P(t,S) - \mathbb{E}\left[\frac{e^{-\int_t^S r(\tau)d\tau}}{P(T,S)}\right] \\
&= \left[1+(S-T)K\right]P(t,S) - P(t,T). \qquad (9.6)
\end{aligned}$$

where the second line holds by the definition of L and the third line follows by the following relation:[2]

$$P(t,T) = \mathbb{E}\left[\frac{e^{-\int_t^S r(\tau)d\tau}}{P(T,S)}\right]. \qquad (9.7)$$

Finally, by replacing (9.5) into (9.6),

$$A(t,T,S;K) = (S-T)\left[K - F(t,T,S)\right]P(t,S). \qquad (9.8)$$

As is clear, A can take on any sign, and is exactly zero when $K = F(t,T,S)$, where $F(t,T,S)$ solves Eq. (9.5).

Bond prices can be expressed in terms of these forward interest rates, namely in terms of the "instantaneous" forward rates. First, rearrange terms in Eq. (9.5) so as to obtain:

$$F(t,T,S) = -\frac{P(t,S)-P(t,T)}{(S-T)P(t,S)}.$$

The *instantaneous forward rate* $f(t,T)$ is defined as

$$f(t,T) \equiv \lim_{S\downarrow T} F(t,T,S) = -\frac{\partial \log P(t,T)}{\partial T}. \qquad (9.9)$$

[2] To show that eq. (9.7) holds, suppose that at time t, £$P(t,T)$ are invested in a bond maturing at time T. At time T, this investment will obviously pay off £1. And at time T, £1 can be further rolled over another bond maturing at time S, thus yielding £ $1/P(T,S)$ at time S. Therefore, it is always possible to invest £$P(t,T)$ at time t and obtain a "payoff" of £ $1/P(T,S)$ at time S. By the FTAP, there are no arbitrage opportunities if and only if eq. (9.7) holds true. Alternatively (and more generally), use the law of iterated expectations to obtain

$$\mathbb{E}\left[\frac{e^{-\int_t^S r(\tau)d\tau}}{P(T,S)}\right] = \mathbb{E}\left\{\mathbb{E}\left[\left.\frac{e^{-\int_t^T r(\tau)d\tau}e^{-\int_T^S r(\tau)d\tau}}{P(T,S)}\right|\mathcal{F}(T)\right]\right\} = P(t,T).$$

It can be interpreted as the marginal rate of return from committing a bond investment for an additional instant. To express bond prices in terms of f, integrate Eq. (9.9)

$$f(t,\ell) = -\frac{\partial \log P(t,\ell)}{\partial \ell}$$

with respect to maturity date ℓ, use the condition that $P(t,t) = 1$, and obtain:

$$P(t,T) = e^{-\int_t^T f(t,\ell)d\ell}. \tag{9.10}$$

9.1.3.3 More on the "marginal revenue" nature of forward rates

Comparing (9.4) with (9.10) yields:

$$R(t,T) = \frac{1}{T-t}\int_t^T f(t,\tau)d\tau. \tag{9.11}$$

By differentiating (9.11) with respect to T yields:

$$\frac{\partial R(t,T)}{\partial T} = \frac{1}{T-t}\left[f(t,T) - R(t,T)\right]. \tag{9.12}$$

Relation (9.12) very clearly reveals the "marginal revenue" nature of forward rates. Similarly as the cost function in any standard ugt micro textbook,

- If $f(t,T) < R(t,T)$, the yield-curve $R(t,T)$ is decreasing at T;
- If $f(t,T) = R(t,T)$, the yield-curve $R(t,T)$ is stationary at T;
- If $f(t,T) > R(t,T)$, the yield-curve $R(t,T)$ is increasing at T.

9.1.3.4 The expectation theory, and related issues

The expectation theory holds that *forward rates equal expected future short-term rates*, or

$$f(t,T) = E[r(T)],$$

where $E(\cdot)$ now denotes expectation under the physical measure. So by Eq. (9.11), the expectation theory implies that,

$$R(t,T) = \frac{1}{T-t}\int_t^T E[r(\tau)]\,d\tau.$$

The question whether $f(t,T)$ is higher than $E[r(T)]$ is very old. The oldest intuition we have is that only risk-adverse investors *may* induce $f(t,T)$ to be higher than the short-term rate they expect to prevail at T, viz,

$$f(t,T) \geq E(r(T)). \tag{9.13}$$

In other terms, (9.13) *never holds true if* *<u>all</u>* *investors are risk-neutral.* (9.14)

The inequality (9.13) is related to the Hicks-Keynesian *normal backwardation hypothesis*.[3] According to Hicks, firms tend to demand long-term funds while fund suppliers prefer to lend

[3] According to the *normal backwardation* (*contango*) *hypothesis*, forward prices are lower (higher) than future expected spot prices. Here the normal backwardation hypothesis is formulated with respect to interest rates.

at shorter maturity dates. The market is cleared by intermediaries who demand a *liquidity premium* to be compensated for their risky activity consisting in borrowing at short maturity dates and lending at long maturity dates. As we will see in a moment, we do not really need a liquidity risk premium to explain (9.13). Pure risk-aversion can be sufficient. In other terms, a proof of statement (9.14) can be sufficient. Here is a proof. By Jensen's inequality,

$$e^{-\int_t^T f(t,\tau)d\tau} \equiv P(t,T) = \mathbb{E}\left[e^{-\int_t^T r(\tau)d\tau}\right] \geq e^{-\int_t^T \mathbb{E}[r(\tau)]d\tau}.$$

By taking logs,

$$\int_t^T \mathbb{E}[r(\tau)]d\tau \geq \int_t^T f(t,\tau)d\tau.$$

This shows statement (9.14).

Another simple prediction on yield-curve shapes can produced as follows. By the same reasoning used to show (9.14), $e^{-(T-t)R(t,T)} \equiv P(t,T) = \mathbb{E}\left[e^{-\int_t^T r(\tau)d\tau}\right] \geq e^{-\int_t^T \mathbb{E}[r(\tau)]d\tau}$. Therefore,

$$R(t,T) \leq \frac{1}{T-t}\int_t^T \mathbb{E}[r(\tau)]d\tau.$$

As an example, suppose that the short-term rate is a martingale under the risk-neutral measure, viz. $\mathbb{E}[r(\tau)] = r(t)$. The previous relation then collapses to:

$$R(t,T) \leq r(t),$$

which means that the yield curve is not-increasing in T. Observing yield-curves that are increasing in T implies that the short-term rate is *not* a martingale under the risk-neutral measure. In some cases, this feature can be attributed to risk-aversion.

Finally, a recurrent definition. The difference

$$R(t,T) - \frac{1}{T-t}\int_t^T E\left[r(\tau)\right]d\tau$$

is usually referred to as *yield term-premium*.

9.1.4 Forward martingale measures

9.1.4.1 Definition

Let $\varphi(t,T)$ be the T-forward price of a claim $S(T)$ at T. That is, $\varphi(t,T)$ is the price agreed at t, that will be paid at T for delivery of the claim at T. Nothing has to be paid at t. By the FTAP, there are no arbitrage opportunities if and only if:

$$0 = \mathbb{E}\left[e^{-\int_t^T r(u)du}\cdot(S(T)-\varphi(t,T))\right].$$

But since $\varphi(t,T)$ is known at time t,

$$\mathbb{E}\left[e^{-\int_t^T r(u)du}\cdot S(T)\right] = \varphi(t,T)\cdot\mathbb{E}\left[e^{-\int_t^T r(u)du}\right].$$

Now use the bond pricing equation (9.3), and rearrange terms in the previous equality, to obtain

$$\varphi(t,T) = \mathbb{E}\left[\frac{e^{-\int_t^T r(u)du}}{P(t,T)}\cdot S(T)\right] = \mathbb{E}\left[\eta_T(T)\cdot S(T)\right], \tag{9.15}$$

where[4,5]

$$\eta_T(T) \equiv \frac{e^{-\int_t^T r(u)du}}{P(t,T)}.$$

Eq. (9.15) suggests that we can define a new probability Q_F^T, as follows,

$$\eta_T(T) = \frac{dQ_F^T}{dQ} \equiv \frac{e^{-\int_t^T r(u)du}}{\mathbb{E}\left[e^{-\int_t^T r(u)du}\right]}. \qquad (9.16)$$

Naturally, $\mathbb{E}[\eta_T(T)] = 1$. Moreover, if the short-term rate process $\{r(\tau)\}_{\tau\in[t,T]}$ be deterministic is deterministic, $\eta_T(T)$ equals one and Q and Q_F^T are the same.

In terms of this new probability Q_F^T, the forward price $\varphi(t,T)$ is:

$$\varphi(t,T) = \mathbb{E}\left[\eta_T(T)\cdot S(T)\right] = \int \left[\eta_T(T)\cdot S(T)\right] dQ = \int S(T)dQ_F^T = \mathbb{E}_{Q_F^T}\left[S(T)\right], \qquad (9.17)$$

where $\mathbb{E}_{Q_F^T}[\cdot]$ denotes the expectation taken under measure Q_F^T. For reasons that will be clear in a moment, Q_F^T is referred to as the *T-forward martingale measure*. The forward martingale measure is a practical tool to price interest-rate derivatives, as we shall explain in Section 9.7. It was introduced by Geman (1989) and Jamshidian (1989), and further analyzed by Geman, El Karoui and Rochet (1995). Appendix 3 provides a few mathematical details on the forward martingale measure.

9.1.4.2 Martingale properties

Forward prices

Clearly, $\varphi(T,T) = S(T)$. Therefore, (9.17) becomes:

$$\varphi(t,T) = \mathbb{E}_{Q_F^T}\left[\varphi(T,T)\right].$$

Forward rates

Forward rates exhibit an analogous property:

$$f(t,T) = \mathbb{E}_{Q_F^T}\left[r(T)\right] = \mathbb{E}_{Q_F^T}\left[f(T,T)\right]. \qquad (9.18)$$

where the last equality holds as $r(t) = f(t,t)$. The proof is also simple. We have,

$$\begin{aligned} f(t,T) &= -\frac{\partial \log P(t,T)}{\partial T} \\ &= -\frac{\partial P(t,T)}{\partial T} \Big/ P(t,T) \\ &= \mathbb{E}\left[\frac{e^{-\int_t^T r(\tau)d\tau}}{P(t,T)}\cdot r(T)\right] \\ &= \mathbb{E}\left[\eta_T(T)\cdot r(T)\right] \\ &= \mathbb{E}_{Q_F^T}\left[r(T)\right]. \end{aligned}$$

[4]As an example, suppose that S is a price process of a traded asset. By the FTAP, there are no arbitrage opportunities if and only if $\{\exp((-\int_t^\tau r(u)du)S(\tau)\}_{\tau\in[t,T]}$ is a Q-martingale. In this case, $\mathbb{E}[\exp(-\int_t^T r(u)du)S(T)] = S(t)$, and Eq. (9.15) then collapses to the well-known formula: $\varphi(t,T)u(t,T) = S(t)$. As is also well-known, entering the forward contract established at t at a later date $\tau > t$ costs. Apply the FTAP to prove that the value of a forward contract as of time $\tau \in [t,T]$ is given by $u(\tau,T)\cdot[\varphi(\tau,T) - \varphi(t,T)]$. [Hint: Notice that the final payoff is $S(T) - \varphi(t,T)$ and that the discount has to be made at time τ.]

[5]Appendix 2 relates forward prices to their certainty equivalent.

Finally, the same result is also valid for the simply-compounded forward rate:

$$F_i(\tau) = \mathbb{E}_{Q_F^{T_{i+1}}}[L(T_i)] = \mathbb{E}_{Q_F^{T_{i+1}}}[F_i(T_i)], \quad \tau \in [t, T_i]$$

where the second equality follows from Eq. (9.63). To show the previous relation, note that by definition, the simply-compounded forward rate $F(t, T, S)$ satisfies:

$$A(t, T, S; F(t, T, S)) = 0,$$

where $A(t, T, S; K)$ is the value as of time t of a FRA struck at K for the time-interval $[T, S]$. By rearranging terms in the first line of (9.6),

$$F(t, T, S)P(t, S) = \mathbb{E}\left[e^{-\int_t^S r(\tau)d\tau} L(T, S)\right].$$

By the definition of $\eta_S(S)$,

$$F(t, T, S) = \mathbb{E}_{Q_F^S}[L(T, S)].$$

Now use the definitions of $L(T_i)$ and $F_i(\tau)$ in (9.1) and (9.61) to conclude.

These relations show that it is only under the forward martingale measure that the expectation theory holds true. Consider, for instance, Eq. (9.18). We have,

$$\begin{aligned} f(t, T) &= \mathbb{E}_{Q_F^T}[r(T)] = \mathbb{E}_Q[\eta_T(T)\, r(T)] \\ &= \underbrace{\mathbb{E}[\eta_T(T)]}_{=1}\mathbb{E}[r(T)] + cov_Q[\eta_T(T), r(T)] \\ &= E[r(T)] + cov[\text{Ker}(T), r(T)] + cov_Q[\eta_T(T), r(T)], \end{aligned}$$

where Ker(T) denotes the pricing kernel in the economy. That is, forward rates in general deviate from the future expected spot rates because of risk-aversion corrections (the second term in the last equality) and because interest rates are stochastic (the third term in the last equality).

9.2 Common factors affecting the yield curve

Which systematic risks affect the entire term-structure of interest rates? How many factors are needed to explain the variation of the yield curve? The standard "duration hedging" portfolio practice relies on the idea that most of the variation of the yield curve is successfully captured by a single factor that produces parallel shifts in the yield curve. How reliable is this idea, in practice?

Litterman and Scheinkman (1991) demontrate that most of the variation (more than 95%) of the term-structure of interest rates can be attributed to the variation of three unobservable factors, which they label (i) a "level" factor, (ii) a "steepness" (or "slope") factor, and (iii) a "curvature" factor. To disentangle these three factors, the authors make an unconditional analysis based on a *fixed-factor* model. Succinctly, this methodology can be described as follows.

Suppose that k returns computed from bond prices at k different maturities are generated by a linear factor structure, with a fixed number k of factors,

$$\underset{p\times 1}{R} = \underset{p\times 1}{\bar{R}} + \underset{p\times k}{B}\,\underset{k\times 1}{F} + \underset{p\times 1}{\epsilon}, \tag{9.19}$$

where R is the vector of excess returns, F is the vector of common factors affecting the returns, $\bar{R}$ is the vector of unconditional expected returns, ϵ is a vector of idiosyncratic components of the return generating process, and B is a matrix containing the factor loadings. Each row of B contains the factor loadings for all the common factors affecting a given return, i.e. the sensitivities of a given return with respect to a change of the factors. Each comumn of B contains the *term-structure of factor loadings*, i.e. how a change of a given factor affects the term-structure of excess returns.

9.2.1 Methodological details

Estimating the model in Eq. (9.19) leads to econometric challenges, mainly because the vector of factors F is unobservable.[6] However, there exists a simpler technique, known as *principal components analysis* (PCA, henceforth), which leads to empirical results qualitatively similar to those that hold for the general model in Eq. (9.19). We discuss these empirical results in the next subsection. We now describe the main methodological issues arising within PCA.

The main idea underlying PCA is to transform the original p correlated variables R into a set of new uncorrelated variables, the *principal components*. These principal components are linear combinations of the original variables, and are arranged in order of decreased importance: the first principal component account for as much as possible of the variation in the original data, etc. Mathematically, we are looking for p linear combinations of the demeaned excess returns,

$$Y_i = C_i^\top \left(R - \bar{R}\right), \qquad i = 1, \cdots, p, \tag{9.20}$$

such that, for p vectors $C_i^\top$ of dimension $1 \times p$, (i) the new variables Y_i are uncorrelated, and (ii) their variances are arranged in decreasing order. The objective of PCA is to ascertain whether a few components of $Y = [Y_1 \cdots Y_p]^\top$ account for the bulk of variability of the original data. Let $C^\top = [C_1^\top \cdots C_p^\top]$ be a $p \times p$ matrix such that we can write Eq. (9.20) in matrix format, $Y = C^\top \left(R - \bar{R}\right)$ or, by inverting,

$$R - \bar{R} = C^{\top -1} Y. \tag{9.21}$$

Next, suppose that the vector $Y^{(k)} = [Y_1 \cdots Y_k]^\top$ accounts for most of the variability of the original data,[7] and let $C^{\top(k)}$ denote a $p \times k$ matrix extracted from the matrix $C^{\top -1}$ through the first k rows of $C^{\top -1}$. Since the components of $Y^{(k)}$ are uncorrelated and they are deemed largely responsible for the variability of the original data, it is natural to "disregard" the last $p - k$ components of Y in Eq. (9.21),

$$\underset{p\times 1}{R - \bar{R}} \approx \underset{p\times k}{C^{\top(k)}} \; \underset{k\times 1}{Y^{(k)}}.$$

[6]Suppose that in Eq. (9.19), $F \sim N(0, I)$, and that $\epsilon \sim N(0, \Psi)$, where Ψ is diagonal. Then, $R \sim N\left(\bar{R}, \Sigma\right)$, where $\Sigma = BB^\top + \Psi$. The assumptions that $F \sim N(0, I)$ and that Ψ is diagonal are necessary to identify the model, but not sufficient. Indeed, any orthogonal rotation of the factors yields a new set of factors which also satisfies Eq. (9.19). Precisely, let T be an orthonormal matrix. Then, $(BT)(BT)^\top = BTT^\top B^\top = BB^\top$. Hence, the factor loadings B and BT have the same ability to generate the matrix Σ. To obtain a unique solution, one needs to impose extra constraints on B. For example, Jöreskog (1967) develop a maximum likelihood approach in which the log-likelihood function is, $-\frac{1}{2}N\left[\log|\Sigma| + \text{Tr}\left(S\Sigma^{-1}\right)\right]$, where S is the sample covariance matrix of R, and the constraint is that $B^\top \Psi B$ be diagonal *with elements arranged in descending order*. The algorithm is: (i) for a given Ψ, maximize the log-likelihood with respect to B, under the constraint that $B^\top \Psi B$ be diagonal with elements arranged in descending order, thereby obtaining $\hat{B}$; (ii) given $\hat{B}$, maximize the log-likelihood with respect to Ψ, thereby obtaining $\hat{\Psi}$, which is fed back into step (i), etc. Knez, Litterman and Scheinkman (1994) describe this approach in their paper. Note that the identification device they describe at p. 1869 (Step 3) roughly corresponds to the requirement that $B^\top \Psi B$ be diagonal *with elements arranged in descending order*. Such a constraint is clearly related to principal component analysis.

[7]There are no rigorous criteria to say what "most of the variability" means in this context. Instead, a likelihood-ratio test is most informative in the context of the estimation of Eq. (9.19) by means of the methods explained in the previous footnote.

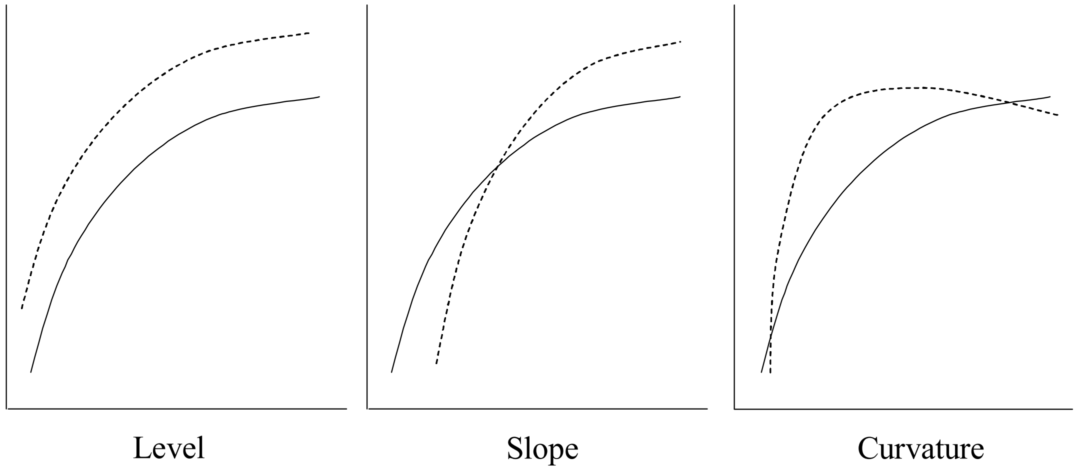


FIGURE 9.1. Changes in the term-structure of interest rates generated by changes in the "level", "slope" and "curvature" factors.

If the vector $Y^{(k)}$ really accounts for most of the movements of R, the previous approximation to Eq. (9.21) should be fairly good.

Let us make more precise what the concept of variability is in the context of PCA. Suppose that Σ has p distinct eigenvalues, ordered from the highest to the lowest, as follows: $\lambda_1 > \cdots > \lambda_p$. Then, the vector C_i in Eq. (9.20) is the eigenvector corresponding to the i-th eigenvalue. Moreover,

$$var\,(Y_i) = \lambda_i, \qquad i = 1, \cdots, p.$$

Finally, we have that

$$\mathrm{R}_{\mathrm{PCA}} = \frac{\sum_{i=1}^{k} var\,(Y_i)}{\sum_{i=1}^{p} var\,(R_i)} = \frac{\sum_{i=1}^{k} \lambda_i}{\sum_{i=1}^{p} \lambda_i}. \tag{9.22}$$

(Appendix 4 provides technical details and proofs of the previous formulae.) It is in the sense of Eq. (9.22) that in the context of PCA, we say that the first k principal components account for $\mathrm{R}_{\mathrm{PCA}}\%$ of the total variation of the data.

9.2.2 The empirical facts

The striking feature of the empirical results uncovered by Litterman and Scheinkman (1991) is that they have been confirmed to hold across a number of countries and sample periods. Moreover, the economic nature of these results is the same, independently of whether the statistical analysis relies on a rigorous factor analysis of the model in Eq. (9.19), or a more back-of-envelope computation based on PCA. Finally, the empirical results that hold for bond returns are qualitatively similar to those that hold for bond yields.

Figure 9.1 visualizes the effects that the three factors have on the movements of the term-structure of interest rates.

- The first factor is called a "level" factor as its changes lead to parallel shifts in the term-structure of interes rates. Thus, this "level" factor produces essentially the same effects on the term-structure as those underlying the "duration hedging" portfolio practice. This factor explains approximately 80% of the total variation of the yield curve.

- The second factor is called a "steepness" factor as its variations induce changes in the slope of the term-structure of interest rates. After a shock in this steepness factor, the short-end and the long-end of the yield curve move in opposite directions. The movements of this factor explain approximately 15% of the total variation of the yield curve.
- The third factor is called a "curvature" factor as its changes lead to changes in the curvature of the yield curve. That is, following a shock in the curvature factor, the middle of the yield curve and both the short-end and the long-end of the yield curve move in opposite directions. This curvature factor accounts for approximately 5% of the total variation of the yield curve.

Understanding the origins of these three factors is still a challenge to financial economists and macroeconomists. For example, macroeconomists explain that central banks affect the short-end of the yield curve, e.g. by inducing variations in federal funds rate in the US. However, the Federal Reserve decisions rest on the current macroeconomic conditions. Therefore, we should expect that the short-end of the yield-curve is related to the development of macroeconomic factors. Instead, the development of the long-end of the yield curve should largely depend on the market average expectation and risk-aversion surrounding future interest rates and economic conditions. Financial economists, then, should expect to see the long-end of the yield curve as being driven by expectations of future economic activity, and by risk-aversion. Indeed, Ang and Piazzesi (2003) demonstrate that macroeconomic factors such as inflation and real economic activity are able to explain movements at the short-end and the middle of the yield curve. However, they show that the long-end of the yield curve is driven by unobservable factors. However, it is not clear whether such unobservable factors are driven by time-varying risk-aversion or changing expectations.

The compelling lesson for practitioners is that reduced-form models with only one factor are unlikely to perform well, in practice.

9.3 Models of the short-term rate

The short-term rate represents the *velocity* at which "locally" riskless investments appreciate over the next instant. This velocity, or rate of increase, is of course *not* a traded asset. What it is traded is a bond and/or a MMA.

9.3.1 Introduction

The fundamental bond pricing equation in Eq. (9.3),

$$P(t,T)=\mathbb{E}\left[e^{-\int_t^T r(u)du}\right], \tag{9.23}$$

suggests to model the arbitrage-free bond price P by using as an input an exogenously given short-term rate process r. In the Brownian information structure considered in this chapter, r would then be the solution to a stochastic differential equation. As an example,

$$dr(\tau)=b(r(\tau),\tau)d\tau+a(r(\tau),\tau)dW(\tau),\quad \tau\in(t,T], \tag{9.24}$$

where b and a are well-behaved functions guaranteeing the existence of a strong-form solution to the previous equation.

Historically, such a modeling approach was the first to emerge. It was initiated in the seminal papers of Merton (1973)[8] and Vasicek (1977), and it is now widely used. This section illustrates the main modeling and empirical challenges related to this approach. We examine one-factor "models of the short-term rate", such as that in Eqs. (9.23)-(9.24), and also multifactor models, in which the short-term rate is a function of a number of factors, $r(\tau) = R(y(\tau))$, where R is some function and y is solution to a multivariate diffusion process.

Two fundamental issues for the model's users are that the models they deal with be (i) fast to compute, and (ii) accurate. As regards the first point, the obvious target would be to look for models with a closed form solution, such as for example, the so-called "affine" models (see Section 9.3.6). The second point is more subtle. Indeed, "perfect" accuracy can never be achieved with models such as that in Eqs. (9.23)-(9.24) - even when this model is is extended to a multifactor diffusion. After all, the model in Eqs. (9.23)-(9.24) can only be taken as it is, really - a *model of determination* of the observed term-structure of interest rates. As such the model in Eqs. (9.23)-(9.24) can never *exactly fit* the observed term structure of interest rates.

As we shall explain, the requirement to exactly fit the initial term-structure of interest rates is important when the model's user is concerned with the pricing of options or other derivatives written on the bonds. The good news is that such a perfect fit can be obtained, once we augment Eq. (9.24) with an infinite dimensional parameter calibrated to the observed term-structure. The bad news is that such a calibration device often leads to "intertemporal inconsistencies" that we will also illustrate.

The models leading to perfect accuracy are often referred to as "no-arbitrage" models. These models work by making the short-term rate *process* exactly pin down the term-structure that we observe at a given instant. As we shall illustrate, intertemporal inconsistencies arise because the parameters of the short-term rate pinning down the term structure today are generally different from the parameters of the short-term rate process which will pin down the term structure tomorrow. As is clear, this methodology goes to the opposite extreme of the initial approach, in which the short-term rate was taken as the *input* of all subsequent movements of the term-structure of interest rates. However, such an initial approach was consistent with the standard rational expectations paradigm permeating modern economic analysis. The rationale behind this approach is that economically admissible (i.e. no-arbitrage) bond prices are rationally formed. That is, they move as a consequence of random changes in the state variables. Economists try to *explain* broad phenomena with the help of a few inputs, a science reduction principle. Practitioners, instead, implement models to solve pricing problems that constantly arise in their trading rooms. Both activities are important, and the choice of the "right" model to use rests on the role that we are playing within a given institution.

9.3.2 The basic bond pricing equation

Suppose that bond prices are solutions to the following stochastic differential equation:

$$\frac{dP_i}{P_i} = \mu_{bi} d\tau + \sigma_{bi} dW, \tag{9.25}$$

where W is a standard Brownian motion in $\mathbb{R}^d$, μ_{bi} and σ_{bi} are some progressively measurable functions (σ_{bi} is vector-valued), and $P_i \equiv P(\tau, T_i)$. The exact functional form of μ_{b_i} and σ_{b_i} is *not* given, as in the BS case. Rather, it is endogenous and must be found as a part of the equilibrium.

[8]Merton's contribution in this field is in one footnote of his 1973 paper!

As shown in Appendix 1, the price system in (9.25) is arbitrage-free if and only if

$$\mu_{bi} = r + \sigma_{bi}\lambda, \tag{9.26}$$

for some $\mathbb{R}^d$-dimensional process λ satisfying some basic regularity conditions. The meaning of (9.26) can be understood by replacing it into Eq. (9.25), and obtaining:

$$\frac{dP_i}{P_i} = (r + \sigma_{bi}\lambda)\, d\tau + \sigma_{bi} dW.$$

The previous equation tells us that the growth rate of P_i is the short-term rate plus a *term-premium* equal to $\sigma_{bi}\lambda$.[9] In the bond market, there are no obvious economic arguments enabling us to sign term-premia. And empirical evidence suggests that term-premia did take both signs over the last twenty years. But term-premia would be zero in a risk-neutral world. In other terms, bond prices are solutions to:

$$\frac{dP_i}{P_i} = r d\tau + \sigma_{bi} d\tilde{W},$$

where $\tilde{W} = W + \int \lambda d\tau$ is a Q-Brownian motion and Q is the risk-neutral probability.

To derive Eq. (9.26) with the help of a specific version of theory developed in Appendix 1, we now work out the case $d = 1$. Consider two bonds, and the dynamics of the value V of a self-financing portfolio in these two bonds:

$$dV = [\pi_1(\mu_{b1} - r) + \pi_2(\mu_{b2} - r) + rV]\, d\tau + (\pi_1\sigma_{1b} + \pi_2\sigma_{b2})\, dW,$$

where π_i is wealth invested in bond maturing at T_i: $\pi_i = \theta_i P_i$. We can zero uncertainty by setting

$$\pi_1 = -\frac{\sigma_{b2}}{\sigma_{b1}}\pi_2.$$

By replacing this into the dynamics of V,

$$dV = \left[-\frac{\mu_{b1} - r}{\sigma_{b1}}\sigma_{b2} + (\mu_{b2} - r)\right]\pi_2 d\tau + rV d\tau.$$

Notice that π_2 can always be chosen so as to make the value of this portfolio appreciate at a rate strictly greater than r. It is sufficient to set:

$$\text{sign}(\pi_2) = \text{sign}\left[-\frac{\mu_{b1} - r}{\sigma_{b1}}\sigma_{b2} + (\mu_{b2} - r)\right].$$

Therefore, to rule out arbitrage opportunities, it must be the case that:

$$\frac{\mu_{b1} - r}{\sigma_{b1}} = \frac{\mu_{b2} - r}{\sigma_{b2}}.$$

The previous relation tells us that Sharpe ratios on any two bonds have to be equal to a process λ, say, and Eq. (9.26) immediately follows. Clearly, such a λ does not depend on T_1 or T_2.

[9] We call $\sigma_{bi}\lambda$ a term-premium because under the good conditions, the bond price $P(t,T)$ decreases with λ uniformly in T, by a comparison result in Mele (2003).

In models of the short-term rate such as (9.24), functions μ_{bi} and σ_{bi} in Eq. (9.25) can be determined after a simple application of Itô's lemma. If $P(r, \tau, T)$ denotes the rational bond price function (i.e., the price as of time τ of a bond maturing at T when the state at τ is r), and r is solution to (9.24), Itô's lemma then implies that:

$$dP = \left(\frac{\partial P}{\partial \tau} + bP_r + \frac{1}{2} a^2 P_{rr} \right) d\tau + aP_r dW,$$

where subscripts denote partial derivatives.

Comparing this equation with Eq. (9.25) then reveals that:

$$\begin{aligned} \mu_b P &= \frac{\partial P}{\partial \tau} + bP_r + \frac{1}{2} a^2 P_{rr}; \\ \sigma_b P &= aP_r. \end{aligned}$$

Now replace these functions into Eq. (9.26) to obtain:

$$\frac{\partial P}{\partial \tau} + bP_r + \frac{1}{2} a^2 P_{rr} = rP + \lambda a P_r, \quad \text{for all } (r, \tau) \in \mathbb{R}_{++} \times [t, T), \tag{9.27}$$

with the obvious boundary condition

$$P(r, T, T) = 1, \quad \text{all } r \in \mathbb{R}_{++}.$$

Eq. (9.27) shows that the bond price, P, depends on both the drift of the short-term rate, b, and the risk-aversion correction, λ. This circumstance occurs as the initial asset market structure is incomplete, in the following sense. In the BS case, the option is redundant, given the initial market structure. In the context we analyze here, the short-term rate r is *not* a traded asset. In other words, the initial market structure has one untraded risk (r) and zero assets - the factor generating uncertainty in the economy, r, is not traded. Therefore, the drift of the short term, b, can not equal $r \cdot r = r^2$ under the risk-neutral probability, and the bond price depends on b, a and λ.

This dependence is, perhaps, a kind of hindrance to practitioners. Instead, it can be regarded as a good piece of news to policy-makers. Indeed, starting from observations and (b, a), one may back out information about λ, which contains information about agents' attitudes towards risk. Information about agents; attitude toward risk can help central bankers to take decisions about the interest rate to set.

By specifying the drift and diffusion functions b and a, and by identifying the risk-premium λ, the partial differential equation (PDE, henceforth) (9.27) can explicitely be solved, either analytically or numerically. Choices concerning the exact functional form of b, a and λ are often dictated by either analytical or empirical reasons. In the next section, we will examine the first, famous short-term rate models in which b, a and λ have a particularly simple form. We will discuss the analytical advantages of these models, but we will also highlight the major empirical problems associated with these models. In Section 9.3.4 we provide a very succinct description of models exhibiting jump (and default) phenomena. In Section 9.3.5, we introduce multifactor models: we will explain why do we need such more complex models, and show that even in this more complex case, arbitrage-free bond prices are still solutions to PDEs such as (9.27). In Section 9.3.6, we will present a class of analytically tractable multidimensional models, known as affine models. We will discuss their historical origins, and highlight their importance as regards the econometric estimation of bond pricing models. Finally, Section 9.3.7 presents the "perfectly fitting" models, and Appendix 5 provides a few technical details about the solution of one of these models.

9.3.3 *Some famous univariate short-term rate models*

9.3.3.1 Vasicek and CIR

The article of Vasicek (1977) is considered to be the seminal contribution to the short-term rate models literature. The model proposed by Vasicek assumes that the short-term rate is solution to:

$$dr(\tau) = (\bar{\theta} - \kappa r(\tau))d\tau + \sigma dW(\tau), \quad \tau \in (t, T], \tag{9.28}$$

where $\bar{\theta}$, κ and σ are positive constants. This model generalizes the one proposed by Merton (1973) in which $\kappa \equiv 0$. The intuition behind model (9.28) is very simple. Suppose first that $\sigma = 0$. In this case, the solution is:

$$r(\tau) = \frac{\bar{\theta}}{\kappa} + e^{-\kappa(\tau-t)} \left[r(t) - \frac{\bar{\theta}}{\kappa} \right].$$

The previous equation reveals that if the current level of the short-term rate $r(t) = \bar{\theta}/\kappa$, it will be "locked-in" at $\bar{\theta}/\kappa$ forever. If instead $r(t) < \bar{\theta}/\kappa$, it will then be the case that for all $\tau > t$, $r(\tau) < \bar{\theta}/\kappa$ too, but $\left| r(\tau) - \bar{\theta}/\kappa \right|$ will eventually shrink to zero as $\tau \to \infty$. An analogous property holds when $r(t) > \bar{\theta}/\kappa$. In all cases, the "speed" of convergence of r to its "long-term" value $\bar{\theta}/\kappa$ is determined by κ: the higher is κ, the higher is the speed of convergence to $\bar{\theta}/\kappa$. In other terms, $\bar{\theta}/\kappa$ is the long-term value towards which r tends to converge, and κ determines the speed of such a convergence.

Eq. (9.28) generalizes the previous ideas to the stochastic differential case. It can be shown that a "solution" to Eq. (9.28) can be written in the following format:

$$r(\tau) = \frac{\bar{\theta}}{\kappa} + e^{-\kappa(\tau-t)} \left[r(t) - \frac{\bar{\theta}}{\kappa} \right] + \sigma e^{-\kappa\tau} \int_t^\tau e^{\kappa s} dW(s),$$

where the integral has the so-called Itô's sense meaning. The interpretation of this solution is similar to the one given above. The short-term rate tends to a sort of "central tendency" $\bar{\theta}/\kappa$. Actually, it will have the tendency to fluctuate around it. In other terms, there is always the tendency for shocks to be absorbed with a speed dictated by the value of κ. In this case, the short-term rate process r is said to exhibit a *mean-reverting* behavior. In fact, it can be shown that the expected future value of r will be given by the solution given above for the deterministic case, viz

$$E\left[r(\tau)\right] = \frac{\bar{\theta}}{\kappa} + e^{-\kappa(\tau-t)} \left[r(t) - \frac{\bar{\theta}}{\kappa} \right].$$

Of course, that is only the *expected* value, not the actual value that r will take at time τ. As a result of the presence of the Brownian motion in Eq. (9.28), r can not be *predicted*, and it is possible to show that the variance of the value taken by r at time τ is:

$$var\left[r(\tau)\right] = \frac{\sigma^2}{2\kappa} \left[1 - e^{-2\kappa(\tau-t)}\right].$$

Finally, it can be shown that r is normally distributed (with expectation and variance given by the two functions given above).

The previous properties of r are certainly instructive. Yet the main objective here is to find the price of a bond. As it turns out, the assumption that the risk premium process λ is a constant allows one to obtain a closed-form solution. Indeed, replace this constant and the functions $b(r) = \bar{\theta} - \kappa r$ and $a(r) = \sigma$ into the PDE (9.27). The result is that the bond price P is solution to the following partial differential equation:

$$0 = \frac{\partial P}{\partial \tau} + \left[(\bar{\theta} - \lambda\sigma) - \kappa r\right] P_r + \frac{1}{2}\sigma^2 P_{rr} - rP, \text{ for all } (r,\tau) \in \mathbb{R} \times [t,T), \tag{9.29}$$

with the usual boundary condition. It is now instructive to see how this kind of PDE can be solved. Guess a solution of the form:

$$P(r,\tau,T) = e^{A(\tau,T) - B(\tau,T)\cdot r}, \tag{9.30}$$

where A and B have to be found. The boundary condition is $P(r,T,T) = 1$, which implies that the two functions A and B must satisfy:

$$A(T,T) = 0 \text{ and } B(T,T) = 0. \tag{9.31}$$

Now suppose that the guess is true. By differentiating Eq. (9.30), $\frac{\partial P}{\partial \tau} = (A_1 - B_1 r)P$, $P_r = -PB$ and $P_{rr} = PB^2$, where $A_1(\tau,T) \equiv \partial A(\tau,T)/\partial\tau$ and $B_1(\tau,T) \equiv \partial B(\tau,T)/\partial\tau$. By replacing these partial derivatives into the PDE (9.29) we get:

$$0 = \left[A_1 - (\bar{\theta} - \lambda\sigma)B + \frac{1}{2}\sigma^2 B^2\right] + (\kappa B - B_1 - 1)r \text{ , for all } (r,\tau) \in \mathbb{R}_{++} \times [t,T).$$

This implies that for all $\tau \in [t,T)$,

$$\begin{aligned} 0 &= A_1 - (\bar{\theta} - \lambda\sigma)B + \frac{1}{2}\sigma^2 B^2 \\ 0 &= \kappa B - B_1 - 1 \end{aligned}$$

subject to the boundary conditions (9.31).

The solutions are

$$B(\tau,T) = \frac{1}{\kappa}\left[1 - e^{-\kappa(T-\tau)}\right]$$

and

$$A(\tau,T) = \frac{1}{2}\sigma^2 \int_\tau^T B(s,T)^2 ds - (\bar{\theta} - \lambda\sigma)\int_\tau^T B(s,T)ds.$$

By the definition of the yield curve given in Section 9.1 (see Eq. (9.4)),

$$R(\tau,T) \equiv -\frac{\ln P(r,t,T)}{T-t} = \frac{-A(t,T)}{T-t} + \frac{B(t,T)}{T-t}r.$$

It is possible to show that there a finite asymptote, i.e. $\lim_{T\to\infty} R(t,T) = \lim_{T\to\infty} \frac{-A(t,T)}{T-t} < \infty$.

The model has a number of features that can describe quite a few aspects of reality. Many textbooks show the typical shapes of the yield-curve that can be generated with the above formula (see, for example, Hull (2003, p. 540). However, this model is known to suffer from two main drawbacks. The first drawback is that the short-term rate is Gaussian and, hence, can take on negative values with positive probability. That is a counterfactual feature of the model. However, it should be stressed that on a practical standpoint, this feature is practically irrelevant. If σ is low compared to $\frac{\bar{\theta}}{\kappa}$, this probability is really very small. The second drawback is tightly connected to the first one. It refers to the fact that the short-term rate diffusion is independent of the level of the short-term rate. That is another conterfactual feature of the model. It is well-known that short-term rates changes become more and more volatile as the level of the short-term rate increases. In the empirical literature, this phenomenon is usually referred to as the *level-effect*.

The model proposed by Cox, Ingersoll and Ross (1985) (CIR, henceforth) addresses these two drawbacks at once, as it assumes that the short-term rate is solution to,

$$dr(\tau) = (\bar{\theta} - \kappa r(\tau))d\tau + \sigma\sqrt{r(\tau)}dW(\tau), \quad \tau \in (t, T].$$

The CIR model is also referred to as "square-root" process to emphasize that the diffusion function is proportional to the square-root of r. This feature makes the model address the level-effect phenomenon. Moreover, this property prevents r from taking negative values. Intuitively, when r wanders just above zero, it is pulled back to the stricly positive region at a strength of the order $dr = \bar{\theta}d\tau$.[10] The transition density of r is noncentral chi-square. The stationary density of r is a gamma distribution.

The expected value is as in Vasicek.[11] However, the variance is different, although its exact expression is really not important here.

CIR formulated a set of assumptions on the primitives of the economy (e.g., preferences) that led to a risk-premium function $\lambda = \ell\sqrt{r}$, where ℓ is a constant. By replacing this, $b(r) = \bar{\theta} - \kappa r$ and $a(r) = \sigma\sqrt{r}$ into the PDE (9.27), one gets (similarly as in the Vasicek model), that the bond price function takes the form in Eq. (9.30), but with functions A and B satisfying the following differential equations:

$$\begin{aligned} 0 &= A_1 - \bar{\theta}B \\ 0 &= -B_1 + (\kappa + \ell\sigma)B + \frac{1}{2}\sigma^2 B^2 - 1 \end{aligned}$$

subject to the boundary conditions (9.31).

In their article, CIR also showed how to compute options on bonds. They even provided hints on how to "invert the term-structure", a popular technique that we describe in detail in Section 9.3.6. For all these features, the CIR model and paper have been used in the industry for many years. And many of the more modern models are mere multidimensional extensions of the basic CIR model. (See Section 9.3.6).

9.3.3.2 Nonlinear drifts

An important issue is the analytical tractability of a given model. As demonstrated earlier, models such as Vasicek and CIR admit a closed-form solution. Among other things, this is because these models have a *linear* drift. Is evidence consistent with this linear assumption? What does empirical evidence suggest as regards mean reversion of the short-term rate?

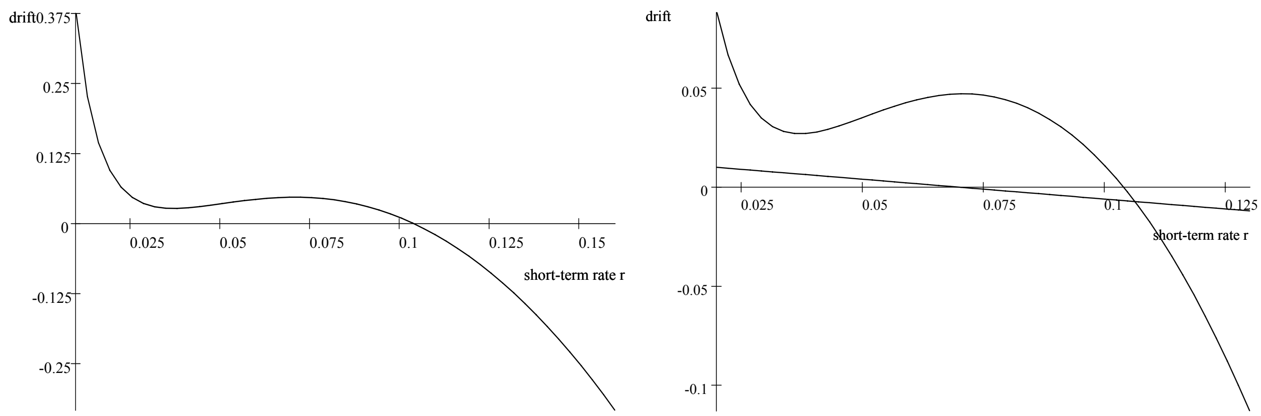
Such an empirical issue is subject to controversy. In the mid 1990s, three papers by Aït-Sahalia (1996), Conley et al. (1997) and Stanton (1997) produced evidence that mean-reverting behavior is nonlinear. As an example, Conley et al. (1997) estimated a drift function of the following form:

$$b(r) = \beta_0 + \beta_1 r + \beta_2 r^2 + \beta_3 r^{-1},$$

[10] This is only intuition. The exact condition under which the zero boundary is unattainable by r is $\bar{\theta} > \frac{1}{2}\sigma^2$. See Karlin and Taylor (1981, vol II chapter 15) for a general analysis of attainability of boundaries for scalar diffusion processes.

[11] The expected value of linear mean-reverting processes is always as in Vasicek, independently of the functional form of the diffusion coefficient. This property follows by a direct application of a general result for diffusion processes given in Chapter 6 (Appendix A).

which is reproduced in Figure 9.2 below (Panel A). Similar results were obtained in the other papers. To grasp the phenomena underlying this nonlinear drift, Figure 9.2 (Panel B) also contrasts the nonlinear shape in Panel A with a linear drift shape that can be obtained by fitting the CIR model to the same data set (US data: daily data from 1981 to 1996).



Panel A Panel B

FIGURE 9.2. Nonlinear mean reversion?

The importance of the nonlinear effects in Figure 9.2 is related to the convexity effects in Mele (2003). Mele (2003) showed that bond prices may be *concave* in the short-term rate if the risk-neutralized drift function is sufficiently convex. While the results in this Figure relate to the *physical* drift functions, the point is nevertheless important as risk-premium terms should look like very strange to completely destroy the nonlinearities of the short-term rate under the physical measure.

The main lesson is that under the "nonlinear drift dynamics", the short-term rate behaves in a way that can at least be *roughly* comparable with that it would behave under the "linear drift dynamics". However, the behavior at the extremes is dramatically different. As the short-term rate moves to the extremes, it is pulled back to the "center" in a very abrupt way. At the moment, it is not clear whether these preliminary empirical results are reliable or not. New econometric techniques are currently being developed to address this and related issues.

One possibility is that such *single factor* models of the short-term rate are simply misspecified. For example, there is strong empirical evidence that the volatility of the short-term rate is time-varying, as we shall discuss in the next section. Moreover, the term-structure implications of a single factor model are counterfactual, since we know that a single factor can not explain the entire variation of the yield curve, as we explained in Section 9.2. We now describe more realistic models driven by more than one factor.

9.3.4 Multifactor models

The empirical evidence reviewed in Section 9.2 suggests that one-factor models can not explain the entire variation of the term-structure of interest rates. Factor analysis suggests we need at least three factors. In this section, we succinctly review the advances made in the literature to address this important empirical issue.

9.3.4.1 Stochastic volatility

In the CIR model, the instantaneous short-term rate volatility is stochastic, as it depends on the level of the short-term rate, which is obviously stochastic. However, there is empirical evidence, surveyed by Mele and Fornari (2000), that suggests that the short-term rate volatility depends on some additional factors. A natural extension of the CIR model is one in which the instantaneous volatility of the short-term rate depends on (i) the level of the short-term rate, similarly as in the CIR model, and (ii) some additional random component. Such an additional random component is what we shall refer to as the "stochastic volatility" of the short-term rate. It is the term-structure counterpart to the stochastic volatility extension of the Black and Scholes (1973) model (see Chapter 8).

Fong and Vasicek (1991) write the first paper in which the volatility of the short-term rate is stochastic. They consider the following model:

$$\begin{aligned} dr(\tau) &= \kappa_r(\bar{r} - r(\tau))\, d\tau + \sqrt{v(\tau)} r(t)^\gamma\, dW_1(\tau) \\ dv(\tau) &= \kappa_v(\theta - v(\tau))\, d\tau + \xi_v \sqrt{v(\tau)} dW_2(\tau) \end{aligned} \tag{9.32}$$

in which κ_r, $\bar{r}$, κ_v, θ and ξ_v are constants, and $[W_1\ W_2]$ is a vector Brownian motion. (To obtain a closed-form solution, Fong and Vasicek set $\gamma = 0$.) The authors also make assumptions about risk aversion corrections. Namely, they assume that the unit-risk-premia for the stochastic fluctuations of the short-term rate, λ_r, and the short-term rate volatility, λ_v, are both proportional to $\sqrt{v(\tau)}$, and then they find a closed-form solution for the bond price as of time t and maturing at time T, $P(r(t), v(t), T-t)$.

Longstaff and Schwartz (1992) propose another model of the short-term rate in which the volatility of the short-term rate is stochastic. The remarkable feature of their model is that it is a general equilibrium model. Naturally, the Longstaff-Schwartz model predicts, as the Fong-Vasicek model, that the bond price is a function of both the short-term rate and its instantaneous volatility.

Note, then, the important feature of these models. The pricing function, $P(r(t), v(t), T-t)$ and, hence, the yield curve $R(r(t), v(t), T-t) \equiv -(T-t)^{-1} \log P(r(t), v(t), T-t)$, depends on the level of the short-term rate, $r(t)$, and one additional factor, the instantaneous variance of the short-term rate $v(t)$. Hence, these models predict that we now have two factors that help explain the term-structure of interest rates, $R(r(t), v(t), T-t)$.

What is the relation between the volatility of the short-term rate and the term-structure of interest rates? Does this volatility help "track" one of the factors driving the variations of the yield curve? Consider, first, the basic Vasicek (1997) model discussed in Section 3.3.3. This model is simply a one-factor model, as it assumes that the volatility of the short-term rate is constant. Yet this model can be used to develop intuition about the full stochastic volatility models, such as the Fong and Vasicek (1991) in Eqs. (9.32).

Using the solution for the Vasicek model, we find that,

$$\frac{\partial R(r(t), T-t)}{\partial \sigma} = -\frac{1}{T-t}\left[\sigma \int_t^T B(T-s)^2\, ds + \lambda \int_t^T B(T-s)\, ds\right], \tag{9.33}$$

where we remind that $B(T-s) = \frac{1}{\kappa}\left[1 - e^{-\kappa(T-s)}\right]$.

The previous expression reveals that if $\lambda \geq 0$, the term-structure of interest rates (and, hence, the bond price) is always decreasing (increasing) in the volatility of the short-term rate. This conclusion parallels a famous result in the option pricing literature, by which the option price is always increasing in the volatility of the price of the asset underlying the contract. As we explained in Chapter 8, this property arises from the convexity of the price with respect to the state variable of which we are contemplating changes in volatility.

The intriguing feature of the model arises when $\lambda < 0$, which is the empirically relevant case to consider.[12] In this case, the sign of $\frac{\partial R(t,T)}{\partial \sigma}$ is the results of a conflict between "convexity" and "slope" effects. "Convexity" effects arise through the term $\sigma \int_t^T B(T-s)^2 ds$, which are referred to as in this way because $\frac{\partial^2 P(r,T-t)}{\partial r^2} = P(r, T-t) B(T-t)^2$. "Slope" effects, instead, arise through the term $\int_t^T B(T-s) ds$, and are referred to as in this way as $\frac{\partial P(r,T-t)}{\partial r} = P(r, T-t) B(T-t)$. If λ is negative, and large in absolute value, slope effects can dominate convexity effects, and the term-structure can actually increase in the volatility parameter σ.

For intermediate values of λ, the term-structure can be both increasing and decreasing in the volatility parameter σ. Typically, at short maturity dates, the convexity effects in Eq. (9.33) are dominated by the slope effects, and the short-end of the term-structure can be increasing in σ. At longer maturity dates, convexity effects should be magnified and can sometimes dominate the slope effects. As a result, the long-end of the term-structure can be decreasing in σ.

Naturally, the previous conclusions are based on comparative statics for a simple model in which the short-term rate has constant volatility. However, these comparative statics illustrate well a general theory of bond price fluctuations in the more interesting case in which the short-term rate volatility is stochastic, as for example in model (9.32) (see Mele (2003)). To develop further intuition about the conflict between convexity and slope effects, consider the following binomial example. In the next period, the short-term rate is either $i^+ = i + d$ or $i^+ = i - d$ with equal probability, where i is the current interest rate level and $d > 0$. The price of a two-period bond is $P(i,d) = m(i,d)/(1+i)$, where $m(i,d) = E[1/(1+i^+)]$ is the expected discount factor of the next period. By Jensen's inequality, $m(i,d) > 1/(1+E[i^+]) = 1/(1+i) = m(i,0)$. Therefore, two-period bond prices increase upon activation of randomness. More generally, two-period bond prices are always increasing in the "volatility" parameter d in this example (see Figure 9.3).

This property relates to an important result derived by Jagannathan (1984, p. 429-430) in the option pricing area, and discussed in Chapter 8. Jagannathan's insight is that in a two-period economy with identical initial underlying asset prices, a terminal underlying asset price $\tilde{y}$ is a mean preserving spread of another terminal underlying asset price $\tilde{x}$ (in the Rothschild and Stiglitz (1970) sense) if and only if the price of a call option on $\tilde{y}$ is higher than the price of a call option on $\tilde{x}$. This is because if $\tilde{y}$ is a mean preserving spread of $\tilde{x}$, then $E[f(\tilde{y})] > E[f(\tilde{x})]$ for f increasing and convex.[13]

These arguments go through as we assumed that the expected short-term rate is independent of d. Consider, instead, a multiplicative setting in which either $i^+ = i(1+d)$ or $i^+ = i/(1+d)$ with equal probability. Litterman, Scheinkman and Weiss (1991) show that in such a setting, bond prices are decreasing in volatility at short maturity dates and increasing in volatility at long maturity dates. This is because expected future interest rates increase over time at

[12] In this simple model, it is more reasonable to assume $\lambda < 0$ rather than $\lambda > 0$. This is because positive risk-premia are observed more frequently than negative risk-premia, and in this model, $u_r < 0$. Together with $\lambda < 0$, $u_r < 0$ ensures that the model generates positive term-premia.

[13] To make such a connection more transparent in terms of the Rothschild and Stiglitz (1970) theory, let $\tilde{m}_d(i^+) = 1/(1+i^+)$ denote the random discount factor when $i^+ = i \mp d$. Clearly $x \mapsto -\tilde{m}_d(x)$ is increasing and concave, and so we must have: $E[-\tilde{m}_{d''}(x)] < E[-\tilde{m}_{d'}(x)] \Leftrightarrow d' < d''$, which is what demonstrated in figure 1. In Jagannathan (1984), f is increasing and convex, and so we must have: $E[f(\tilde{y})] > E[f(\tilde{x})] \Leftrightarrow \tilde{y}$ is riskier than (or a mean preserving spread of) $\tilde{x}$.

a strength positively related to d. That is, *the expected variation of the short-term rate is increasing in the volatility of the short-term rate, d*, a property that can be re-interpreted as one arising in an economy with risk-averse agents. At short maturity dates, such an effect dominates the convexity effect illustrated in Figure 9.3. At longer maturity dates, the convexity effect dominates.

This simple example illustrates the yield-curve / volatility relation in the Vasicek model summarized by Eq. (9.33). As is clear, volatility changes do not generally represent a mean preserving spread for the risk-neutral distribution in the term-structure framework considered here. The seminal contribution of Jagannathan (1984) suggests that this is generally the case in the option pricing domain. In models of the short-term rate, the short-term rate is not a traded asset. Therefore, the risk-neutral drift function of the short-term rate does in general depend on the short-term volatility. For example, in the simple and scalar Vasicek model, this property activates slope effects, as Eq. (9.33) reveals. In this case, and in the more complex stochastic volatility cases, it can be shown that if the risk-premium required to bear the interest rate risk is negative and sufficiently large in absolute value, slope effects dominate convexity effects at any finite maturity date, thus making bond prices decrease with volatility at any arbitrary maturity date.

What are the implications of these results in terms of the classical factor analysis of the term-structure reviewed in Section 9.2? Clearly, the very short-end of yield curve is not affected by movements of the volatility, as $\lim_{T\to t} R(r(t), v(t), T-t) = r(t)$, for all possible values of $v(t)$. Also, in these models, we have that $\lim_{T\to\infty} R(r(t), v(t), T-t) = \bar{R}$, where $\bar{R}$ is a constant and, hence, independent of of $v(t)$. Therefore, movements in the short-term volatility can only produce their effects on the middle of the yield curve. For example, if the risk-premium required to bear the interest rate risk is negative and sufficiently large, an upward movement in $v(t)$ can produce an effect on the yield curve qualitatively similar to that depicted in Figure 9.1 ("Curvature" panel), and would thus roughly mimic the "curvature" factor that we reviewed in Section 9.2.

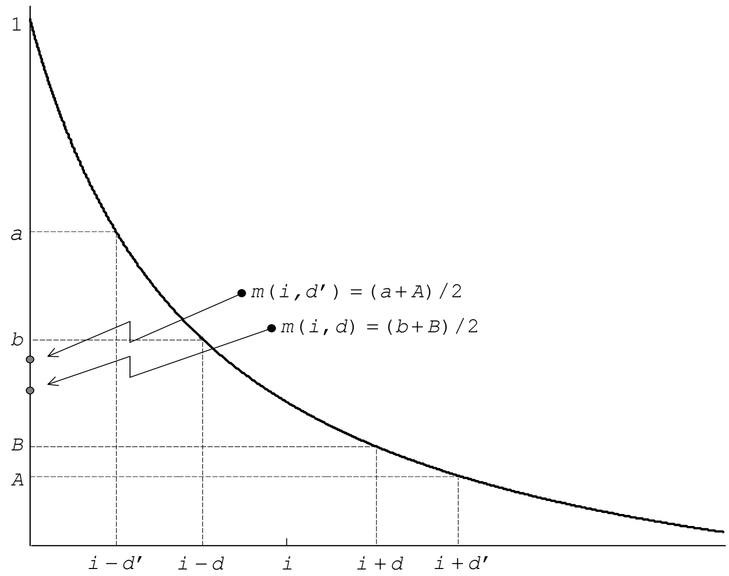


FIGURE 9.3. *A connection with the Rothschild-Stiglitz-Jagannathan theory: the simple case in which convexity of the discount factor induces bond prices to be increasing in volatility.* If the

risk-neutralized interest rate of the next period is either $i^+ = i + d$ or $i^+ = i - d$ with equal probability, the random discount factor $1/(1+i^+)$ is either B or b with equal probability. Hence $m(i,d) = E\left[1/(1+i^+)\right]$ is the midpoint of $\overline{bB}$. Similarly, if volatility is $d' > d$, $m(i,d')$ is the midpoint of $\overline{aA}$. Since $\overline{ab} > \overline{BA}$, it follows that $m(i,d') > m(i,d)$. Therefore, the two-period bond price $u(i,v) = m(i,v)/(1+i)$ satisfies: $u(i,d') > u(i,d)$ for $d' > d$.

9.3.4.2 Three-factor models

We need at least three factor to explain the entire variation in the yield-curve. A model in which the interest rate volatility is stochastic may be far from being exhaustive in this respect. A natural extension is a model in which the drift of the short-term rate contains some predictable component, $\bar{r}(\tau)$, which acts as a third factor, as in the following model:

$$\begin{aligned} dr(\tau) &= \kappa_r (\bar{r}(\tau) - r(\tau))\, d\tau + \sqrt{v(\tau)} r(t)^\gamma \, dW_1(\tau) \\ dv(\tau) &= \kappa_v (\theta - v(\tau))\, d\tau + \xi_v \sqrt{v(\tau)} dW_2(\tau) \\ d\bar{r}(\tau) &= \kappa_{\bar{r}} (\bar{\imath} - \bar{r}(\tau))\, d\tau + \xi_{\bar{r}} \sqrt{\bar{r}(\tau)} dW_3(\tau) \end{aligned} \tag{9.34}$$

where $\kappa_r, \gamma, \kappa_v, \theta, \xi_v, \kappa_{\bar{r}}, \bar{\imath}$ and $\xi_{\bar{r}}$ are constants, and $[W_1 \; W_2 \; W_3]$ is vector Brownian motion.

Balduzzi et al. (1996) develop the first model in which the drift of the short-term rate changes stochastically, as in Eqs. (9.34). Dai and Singleton (2000) consider a number of models that generalize that in Eqs. (9.34). The term-structure implications of these models can be understood very simply. First, the bond price has now the form, $P(r(t), \bar{r}(t), v(t), T-t)$ and, hence, the yield curve is, under reasonable assumptions on the risk-premia, $R(r(t), \bar{r}(t), v(t), T-t) \equiv -(T-t)^{-1} \log P(r(t), \bar{r}(t), v(t), T-t)$. Second, and intuitively, changes in the new factor $\bar{r}(t)$ should primarily affect the long-end of the yield curve. This is because empirically, the usual finding is that the short-term rate reverts relatively quickly to the long-term factor $\bar{r}(\tau)$ (i.e. κ_r is relatively high), where $\bar{r}(\tau)$ mean-reverts slowly (i.e. $\kappa_{\bar{r}}$ is relatively low). Ultimately, the slow mean-reversion of $\bar{r}(\tau)$ means that changes in $\bar{r}(\tau)$ last for the relevant part of the term-structure we are usually interested in (i.e. up to 30 years), despite the fact that $\lim_{T\to\infty} R(r(t), \bar{r}(t), v(t), T-t)$ is independent of the movements of the three factors $r(t)$, $\bar{r}(t)$ and $v(t)$.

However, it is difficult to see how to reconcile such a behavior of the long-end of the yield curve with the existence of any one of the factors discussed in Section 9.2. Indeed, a *joint* change in both the short-term rate, $r(t)$, and the "long-term" rate, $\bar{r}(t)$, should be really needed to mimic the "Level" panel of Figure 9.1 in Section 9.2. However, this interpretation is at odds with the assumption that the factors discussed in Section 9.2 are uncorrelated! Moreover, and crucially, the empirical results in Dai and Singleton reveal that if any, $r(t)$ and $\bar{r}(t)$ are negatively correlated.

Finally, to emphasize how exacerbated these puzzles are, consider the effects of changes in the short-term rate $r(t)$. We know that the long-end of the term-structure is not affected by movements of the short-term rate. Hence, the short-term rate acts as a "steepness" factor, as in Figure 9.1 of Section 9.2 ("Slope" panel). However, this interpretation is restrictive, as factor analysis reveals that the short-end and the long-end of the yield curve move in opposite directions after a change in the steepness factor. Here, instead, a change in the short-term rate only modifies the short-end (and, perhaps, the middle) of the yield curve and, hence, does not produce any variation in the long-end curve.

9.3.5 *Affine and quadratic term-structure models*

9.3.5.1 Affine

The Vasicek and CIR models predict that the bond price is exponential-affine in the short-term rate r. This property is the expression of a general phenomenon. Indeed, it is possible to show that bond prices are exponential-affine in r if, and only if, the functions b and a^2 are affine in r. Models that satisfy these conditions are known as *affine models*. More generally, these basic results extend to *multifactor* models, in which bond prices are exponential-affine in the state variables.[14] In these models, the short-term rate is a function $r(y)$ such that

$$r(y) = r_0 + r_1 \cdot y,$$

where r_0 is a constant, r_1 is a vector, and y is a multidimensional diffusion, in $\mathbb{R}^d$, and is solution to.

$$dy(\tau) = \kappa(\mu - y(t))\,dt + \Sigma V(y(\tau))\,dW(\tau),$$

where W is a d-dimensional Brownian motion, Σ is a full rank $n \times d$ matrix, and V is a full rank $d \times d$ diagonal matrix with elements,

$$V(y)_{(ii)} = \sqrt{\alpha_i + \beta_i^\top y}, \quad i = 1, \cdots, d, \tag{9.35}$$

for some scalars α_i and vectors β_i. Langetieg (1980) develops the first multifactor model of this kind, in which $\beta_i = 0$. Under the assumption that the risk-premia Λ are

$$\Lambda(y) = V(y)\,\lambda_1,$$

for some d-dimensional vector λ_1, Duffie and Kan (1996) show that the bond price is exponential-affine in the state variables y.

The clear advantage of affine models is that they considerably simplify the econometric estimation. Chapter 11 discusses the estimation details of these and related models.

9.3.5.2 Quadratic

Affine models are known to impose tight conditions on the structure of the volatility of the state variables. These restrictions arise to keep the square root in Eq. (9.35) real valued. But these constraints may hinder the actual performance of the models. There exists another class of models, known as quadratic models, that partially overcome these difficulties.

9.3.6 *Short-term rates as jump-diffusion processes*

Seminal contribution (extension of the CIR general equilibrium model to jumps-diffusion): Ahn and Gao (1988, JF). Suppose that the short-term rate is a jump-diffusion process:

$$dr(\tau) = b^J(r(\tau))d\tau + a(r(\tau))dW(\tau) + \ell(r(\tau)) \cdot \mathcal{S} \cdot dZ(\tau),$$

where the previous equation is written under the risk-neutral measure, and b^J is thus a jump-adjusted risk-neutral drift. For all $(r, \tau) \in \mathbb{R}_{++} \times [t, T)$, the bond price $P(r, \tau, T)$ is then the

[14] More generally, we say that affine models are those that make the characteristic function exponential-affine in the state variables. In the case of the multifactor interest rate models of the previous section, this condition is equivalent to the condition that bond prices are exponential affine in the state variables.

solution to,

$$0=\left(\frac{\partial}{\partial\tau}+L-r\right)P(r,\tau,T)+v^{Q}\int_{\text{supp}(\mathcal{S})}\left[P\left(r+\ell\mathcal{S},\tau,T\right)-P\left(r,\tau,T\right)\right]p\left(d\mathcal{S}\right), \tag{9.36}$$

and $P(r,T,T)=1\ \forall r\in\mathbb{R}_{++}$.

This is because, as usual, $\{\exp(-\int_t^\tau r(u)du)P(r,\tau,T)\}_{\tau\in[t,T]}$ must be a martingale under the risk-neutral measure in order to prevent arbitrage opportunities.[15] Also, we can model the presence of different quality (or "types") of jumps, and the previous formula becomes:

$$0=\left(\frac{\partial}{\partial\tau}+L-r\right)P(r,\tau,T)+\sum_{j=1}^{N}v_j^{Q}\int_{\text{supp}(\mathcal{S})}\left[P\left(r+\ell\mathcal{S},\tau,T\right)-P\left(r,\tau,T\right)\right]p^{j}\left(d\mathcal{S}\right),$$

where N is the number of jump types, but here for simplicity we just set $N=1$.

As regards the risk-neutral distribution, the important thing as usual is to identify the risk-premia. Here we simply have:

$$v^{Q}=v\cdot\lambda^{J},$$

where v is the intensity of the short-term rate jump under the *physical* distribution, and λ^J is the risk-premium demanded by agents to be compensated for the presence of jumps.[16]

Bonds subject to default-risk can be modeled through partial differential equations. This is particularly the case when default is considered as an exogeously given rare event modeled as a Poisson process. This is the so-called "reduced-form" approach. Precisely, assume that the event of default at each instant of time is a Poisson process Z with intensity v,[17] and assume that in the event of default at point τ, the holder of the bond receives a recovery payment $\bar{P}(\tau)$ which can be a deterministic function of time (e.g., a constant) or more generally, a $\sigma\left(r(s):t\leq s\leq\tau\right)$-adapted process satisfying some basic regularity conditions.

Next, let $\hat{\tau}$ be the random default time, and let's create an auxiliary state variable g with the following features:

$$g=\begin{cases}0 & \text{if } t\leq\tau<\hat{\tau}\\ 1 & \text{otherwise}\end{cases}$$

The relevant information for an investor is thus given by the following risk-neutral dynamics:

$$\begin{cases}dr(\tau)=b(r(\tau))d\tau+a(r(\tau))dW(\tau)\\ dg(\tau)=\mathcal{S}\cdot dN(\tau),\text{ where }\mathcal{S}\equiv1,\text{ with probability one}\end{cases} \tag{9.37}$$

Denote the rational bond price function as $P(r,g,\tau,T)$, $\tau\in[t,T]$. It is assumed that $\forall\tau\in[t,T]$ and $\forall v\in(0,\infty)$, $P(r,1,\tau,T)=\bar{P}(\tau)<P(r,0,\tau,T)$ a.s. As shown below, such an assumption, plus the assumption that $\bar{P}(\tau;v')\geq\bar{P}(\tau;v)\Leftrightarrow v'\geq v$, is sufficient to guarantee that default-free bond prices are higher than defaultable bond prices.

By the usual absence of arbitrage opportunities arguments, the following equation is satisfied by the pre-default bond price $P(r,0,\tau,T)=P^{\text{pre}}(r,\tau,T)$:

$$\begin{aligned}0&=\left(\frac{\partial}{\partial\tau}+L-r\right)P(r,0,\tau,T)+v(r)\cdot\left[P(r,1,\tau,T)-P(r,0,\tau,T)\right]\\ &=\left(\frac{\partial}{\partial\tau}+L-(r+v(r))\right)P(r,0,\tau,T)+v(r)\bar{P}(\tau),\quad\tau\in[t,T),\end{aligned} \tag{9.38}$$

[15] Just use $y(\tau)\equiv b(\tau)^{-1}u(r(\tau),\tau,T)$, where b solves $db(\tau)=r(\tau)b(\tau)d\tau$ (in differential form), for the connection between Eq. (9.36) and martingales.

[16] Further details on changes of measures for jump-type processes can be found in Brémaud (1981).

[17] The approach followed here is the one developed in Mele (2003).

with the usual boundary condition $P(r,0,T,T)=1$.

The solution for the pre-default bond price is:

$$\begin{aligned} P^{\text{pre}}(x,t,T) &= \mathbb{E}^*\left[\exp\left(-\int_t^T (r(\tau)+v(r(\tau)))d\tau\right)\right] \\ &\quad +\mathbb{E}^*\left[\int_t^T \exp\left(-\int_t^\tau (r(u)+v(r(u)))du\right)\cdot v(r(\tau))\bar{P}(\tau)d\tau\right], \end{aligned}$$

where $\mathbb{E}^*[\cdot]$ is the expectation operator taken with reference to only the first equation of system (9.37). This coincides with Duffie and Singleton (1999, Eq. (10) p. 696) when we define a percentage loss process l in $[0,1]$ so as to have $\bar{P}=(1-l)\cdot P$. Indeed, inserting $\bar{P}=(1-l)\cdot P$ into Eq. (9.38) gives:

$$0=\left(\frac{\partial}{\partial\tau}+L-(r+l(\tau)v(r))\right)P(r,0,\tau,T),\ \forall(r,\tau)\in\mathbb{R}_{++}\times[t,T),$$

with the usual boundary condition, the solution of which is:

$$P^{\text{pre}}(x,t,T)=\mathbb{E}^*\left[\exp\left(-\int_t^T (r(\tau)+l(\tau)\cdot v(r(\tau)))d\tau\right)\right].$$

To validate the claim that the bond price is decreasing with v, consider two economies A and B in which the corresponding default-intensities are v^A and v^B, and assume that the coefficients of L don't depend on default-intensity. The pre-default bond price function in economy i is $P^i(r,\tau,T)$, $i=A,B$, and satisfies:

$$0=\left(\frac{\partial}{\partial\tau}+L-r\right)P^i+v^i\cdot(\bar{P}^i-P^i),\quad i=A,B,$$

with the usual boundary condition. Substracting these two equations and rearranging terms reveals that the price difference $\Delta P(r,\tau,T)\equiv P^A(r,\tau,T)-P^B(r,\tau,T)$ satisfies, $\forall(r,\tau)\in\mathbb{R}_{++}\times[t,T)$,

$$0=\left(\frac{\partial}{\partial\tau}+L-(r+v^A)\right)\Delta P(r,\tau,T)+\left(v^A-v^B\right)\cdot\left[\bar{P}^B(\tau)-P^B(r,\tau,T)\right]+v^A\left[\bar{P}^A(\tau)-\bar{P}^B(\tau)\right],$$

with $\Delta P(r,T,T)=0$, $\forall r\in\mathbb{R}_{++}$. Given the previous assumptions, the proof is complete by an application of the maximum principle (see Appendix ? in chapter 6).

9.4 No-arbitrage models

9.4.1 Fitting the yield-curve, perfectly

Traders don't want to *explain* the term structure. Rather, they wish to *take* it as they observe it. The simple reason for wishing so is the increased importance of the interest rate derivatives business. Interest rate derivatives are derivatives written on term structure objects (see Section 9.7). Consider, for instance, a European option written on a bond. Traders find it unsatisfactory to have a model that only "explains" the bond price. A model's mistake on the bond price is likely to generate a huge option price mistake. How can we trust an option pricing model that is not even able to pin down the value of the underlying asset price? To illustrate these points, denote with $P(r(\tau),\tau,S)$ the rational price process of a zero coupon bond maturing at some time S. What is the price of a European option written on this bond, struck at K and expiring at $T<S$? By the FTAP, there are no arbitrage opportunities if and only if the option price C^b is:

$$C^b\left(r(t),t,T,S\right)=\mathbb{E}\left[e^{-\int_t^T r(\tau)d\tau}\cdot\left(P(r(T),T,S)-K\right)^+\right],$$

where the symbol $(a)^+$ denotes $\max(0,a)$. As an example, in affine models, P is lognormal whenever r is normally distributed. This happens precisely for the Vasicek model. Intuition developed for the Black and Scholes (1973) (BS) formula then suggests that in this case, the previous expectation is a nonlinear function of the *current* bond price $P(r(t),t,T)$. This claim can not be shown with the simple risk-neutral tools used to show the BS formula. One of the troubles is due to the presence of the $e^{-\int_t^T r(\tau)d\tau}$ term inside the brackets, which is obviously unknown at the time of evaluation t. Instead, the technology of the forward martingale measure introduced in Section 9.2.4. Precisely, let 1_{ex} be the indicator of all events s.t. the option is exercized i.e., that $P(r(T),T,S)\geq K$. We have:

$$\begin{aligned}
&C^b\left(r(t),t,T,S\right)\\
&=\mathbb{E}\left[e^{-\int_t^T r(\tau)d\tau}P(r(T),T,S)\cdot 1_{ex}\right]-K\cdot\mathbb{E}\left[e^{-\int_t^T r(\tau)d\tau}\cdot 1_{ex}\right]\\
&=P(r(t),t,S)\cdot\mathbb{E}\left[\frac{e^{-\int_t^S r(\tau)d\tau}}{P(r(t),t,S)}\cdot 1_{ex}\right]-KP(r(t),t,T)\cdot\mathbb{E}\left[\frac{e^{-\int_t^T r(\tau)d\tau}}{P(r(t),t,T)}\cdot 1_{ex}\right]\\
&=P(r(t),t,S)\cdot\mathbb{E}_{Q_F^S}\left[1_{ex}\right]-KP(r(t),t,T)\cdot\mathbb{E}_{Q_F^T}\left[1_{ex}\right]\\
&=P(r(t),t,S)\cdot Q_F^S\left[P(r(T),T,S)\geq K\right]-KP(r(t),t,T)\cdot Q_F^T\left[P(r(T),T,S)\geq K\right],
\end{aligned}\tag{9.39}$$

where the first term in the second equality has been derived by an argument nearly identical to the one produced in Section 9.1 (see footnote 2);[18] Q_F^i $(i=T,S)$ is the i-forward measure; and finally, $\mathbb{E}_{Q_F^i}[\cdot]$ is the expectation taken under the i-forward martingale measure (see Section 9.1 for more details).

In Section 9.7, we will learn how to compute the two probabilities in Eq. (9.39) in an elegant way. But as it should already be clear, the bond option price does depend on *theoretical* bond prices $P(r(t),t,T)$ and $P(r(t),t,S)$ which in turn, are generally *never* equal to the *current*, observed market prices. This is so because $P(r(t),t,T)$ is only the output of a standard rational expectations model. Evidently, this is not an important concern for economists working in Central Banks who wish to predict future term-structure movements with the help of a few, key state variables (as in the multifactor models discussed before). But practitioners concerned with bond option pricing are perfectly right in asking for a model that perfectly matches the observed term-structure they face at the time of evaluation. The aim of the models studied in

[18] By the Law of Iterated Expectations,

$$\mathbb{E}\left[e^{-\int_t^T r(\tau)d\tau}u(r(T),T,S)1_{ex}\right]=\mathbb{E}\left\{e^{-\int_t^T r(\tau)d\tau}1_{ex}\mathbb{E}\left[e^{-\int_T^S r(\tau)d\tau}\,\middle|\,F(T)\right]\right\}=\mathbb{E}\left[e^{-\int_t^S r(\tau)d\tau}1_{ex}\right].$$

this section is to exactly fit the initial term-structure, which for this reasons we call "perfectly fitting models".[19]

We do not develop a general model-building principle. Rather, we present specific models that are effectively able to deserve the "perfectly fitting" qualification. We shall focus on two celebrated models: the Ho and Lee (1986) model, and one generalization of it introduced by Hull and White (1990). In all cases, the general modeling principle consists in generating bond prices expiring at some date S that are of course random at time $T < S$, but also exactly equal to the *current* observed bond prices (at time t). Finally, these prices must be arbitrage-free. As we show, these conditions can be met by augmenting the models seen in the previous sections with a set of "infinite dimensional parameters".

A final remark. In Section 9.7, we will show that at least for the Vasicek's model, Eq. (9.39) will not explicitly depend on r because it only "depends" on $P(r(t), t, T)$ and $P(r(t), t, S)$. That is the essence of the celebrated Jamshidian's (1989) formula. One may then be right in asking why should we look for perfectly fitting models? After all, it would be sufficient to use the Jamshidian's formulae in Section 9.7, and replace $P(r(t), t, T)$ and $P(r(t), t, S)$ with the corresponding observed values $P^{\$}(t, T)$ and $P^{\$}(t, S)$ (say). In this way the model *is* perfectly fitting. Apart from being theoretically inconsistent (you would have a model predicting something generically different from prices), this way of thinking also leads to some practical drawbacks. As we will show in Section 9.7, the bond option Jamshidian's formula agrees "in notation" with that obtained with the corresponding perfectly fitting model. But as we move to more complex interest rate derivatives, results are completely different. This is the case, for example, of options on *coupon* bonds and *swaption* contracts (see Section 9.7.3 for precise details on this). Finally, it may be the case that some maturity dates are actually not traded at some point in time. As an example, it may happen that $P^{\$}(t, T)$ is not observed. (Furthermore, it may happen that one may wish to price more "exotic", or less liquid bonds or options on these bonds.) An intuitive procedure to face up to this difficulty is to "interpolate" the observed, traded maturities. And in fact, the objective of perfectly fitting models is to allow for such an "interpolation" while preserving absence of arbitrage opportunities.

9.4.2 Ho and Lee

Consider the model,

$$dr(\tau) = \theta(\tau)d\tau + \sigma d\tilde{W}(\tau), \quad \tau \geq t, \tag{9.40}$$

where $\tilde{W}$ is a Q-Brownian motion, σ is a constant, and $\theta(\tau)$ is an "infinite dimensional" parameter introduced to pin down the initial, observed term structure.[20] The time of evaluation is t. The reason we refer to $\theta(\tau)$ as "infinite dimensional" parameter is that we regard $\theta(\tau)$ as a function of calendar time $\tau \geq t$. Crucially, then, we assume that this function is known at the time of evaluation t.

Clearly, Eq. (9.40) gives rise to an affine model. Therefore, the bond price takes the following form,

$$P(r(\tau), \tau, T) = e^{A(\tau,T) - B(\tau,T)\cdot r(\tau)}, \tag{9.41}$$

[19] As noted earlier, these models are usually referred to as "no-arbitrage" models, which is quite a curious qualification. No one doubts that the models seen in the previous sections are arbitrage-free. Therefore, we prefer to qualify the short-term models that exactly fit the initial term-structure as "perfectly fitting models".

[20] The original Ho and Lee (1986) formulation was set in discrete time, and Eq. (9.40) represents its "diffusion limit".

for two functions A and B to be determined below. It is easy to show that,

$$A(\tau,T)=\int_\tau^T\theta(s)(s-T)ds+\frac{1}{6}\sigma^2(T-\tau)^3\quad;\quad B(\tau,T)=T-\tau.$$

The instantaneous forward rate $f(\tau,T)$ predicted by the model is then,

$$f(\tau,T)=-\frac{\partial\log P(r(\tau),\tau,T)}{\partial T}=-A_2(\tau,T)+B_2(\tau,T)\cdot r(\tau),\tag{9.42}$$

where $A_2(\tau,T)\equiv\partial A(\tau,T)/\partial T$ and $B_2(\tau,T)\equiv\partial B(\tau,T)/\partial T$.

On the other hand, let $f_\$(t,\tau)$ denote the instantaneous, observed forward rate. By matching $f(t,\tau)$ to $f_\$(t,\tau)$ yields:

$$f_\$(t,\tau)=f(t,\tau)=\int_t^\tau\theta(s)ds-\frac{1}{2}\sigma^2(\tau-t)^2+r(t),\tag{9.43}$$

where we have evaluated the two partial derivatives $A_2(t,\tau)$ and $B_2(t,\tau)$. Hence, since $P(t,T)=\exp\left(-\int_t^T f(t,\tau)\,d\tau\right)$, this means that this $\theta(\tau)$ guarantees an exact fit of the term-structure. By differentiating the previous equation with respect to τ, one obtains the solution for θ,[21]

$$\theta(\tau)=\frac{\partial}{\partial\tau}f_\$(t,\tau)+\sigma^2(\tau-t).\tag{9.44}$$

9.4.3 Hull and White

Consider the model

$$dr(\tau)=(\theta(\tau)-\kappa r(\tau))\,d\tau+\sigma d\tilde{W}(\tau),\tag{9.45}$$

where $\tilde{W}$ is a Q-Brownian motion, and κ,σ are constants. Clearly, this model generalizes both the Ho and Lee model (9.40) and the Vasicek model (9.28). In the original formulation of Hull and White, κ and σ were both time-varying, but the main points of this model can be learnt by working out this particular simple case.

Eq. (9.45) also gives rise to an affine model. Therefore, the solution for the bond price is given by Eq. (9.41). It is easy to show that the functions A and B are given by

$$A(\tau,T)=\frac{1}{2}\sigma^2\int_\tau^T B(s,T)^2ds-\int_\tau^T\theta(s)B(s,T)ds,\tag{9.46}$$

and

$$B(\tau,T)=\frac{1}{\kappa}\left[1-e^{-\kappa(T-\tau)}\right].\tag{9.47}$$

By reiterating the same reasoning produced to show (9.44), one shows that the solution for θ is:

$$\theta(\tau)=\frac{\partial}{\partial\tau}f_\$(t,\tau)+\kappa f_\$(t,\tau)+\frac{\sigma^2}{2\kappa}\left[1-e^{-2\kappa(\tau-t)}\right].\tag{9.48}$$

A proof of this result is in Appendix 5.

Why did we need to go for this more complex model? After all, Ho & Lee model is already able to pin down the entire yield curve. The answer is that investment banks typically prices a lot of derivatives. The yield curve is not the only thing to be exactly fit. Rather it is only the starting point. In general, the more flexible a given perfectly fitting model is, the more successful it is to price more complex derivatives.

[21] To check that θ is indeed the solution, replace Eq. (9.44) into Eq. (9.43) and verify that Eq. (9.43) holds as an identity.

9.4.4 Critiques

Two important critiques to these models:

- As we shall see in Section 9.6, closed-form solution for options on bond prices are easy to implement when the short-term rate is Gaussian. We will use the T-forward measure machinery to show this. In principle, they could also be used to price caps, floors and swaptions. But in general, no-closed form solutions are available that reproduce the standard market practice. This difficulty is overcome by a class of models known as "market models" that is built upon the modelling principles of the HJM models examined in Section 9.4.

- Intertemporal inconsistencies: θ functions have to be re-calibrated every single day. (As Eq. (9.44) demonstrates, at time t, $\theta(\tau)$ depends on the slope of $f_{\$}$ which can change every day.) This kind of problems is present in HJM-type models

- Stochastic string shocks models.

9.5 The Heath-Jarrow-Morton model

9.5.1 Motivation

The bond price representation (9.10),

$$P(\tau, T) = e^{-\int_\tau^T f(\tau,\ell)d\ell}, \quad \text{all } \tau \in [t, T], \tag{9.49}$$

is the starting point of a now popular modeling approach originally developed by Heath, Jarrow and Morton (1992) (HJM, henceforth). Given (9.49), the modeling strategy of this approach is to take as primitive *the τ-stochastic evolution of the entire structure of forward rates* (not only the short-term rate $r(t) = \lim_{\ell \downarrow t} f(t, \ell) \equiv f(t, t)$).[22] Given (9.49) and the initial, observed structure of forward rates $\{f(t, \ell)\}_{\ell \in [t,T]}$, no-arbitrage "cross-equations" relations arise to restrict the stochastic behavior of $\{f(\tau, \ell)\}_{\tau \in (t,\ell]}$ for any $\ell \in [t, T]$.

By construction, the HJM approach *allows for a perfect fit of the initial term-structure.* This point may be grasped very simply by noticing that the bond price $P(\tau, T)$ is,

$$\begin{aligned}
P(\tau, T) &= e^{-\int_\tau^T f(\tau,\ell)d\ell} \\
&= \frac{P(t,T)}{P(t,\tau)} \cdot \frac{P(t,\tau)}{P(t,T)} e^{-\int_\tau^T f(\tau,\ell)d\ell} \\
&= \frac{P(t,T)}{P(t,\tau)} \cdot e^{-\int_t^\tau f(t,\ell)d\ell + \int_t^T f(t,\ell)d\ell - \int_\tau^T f(\tau,\ell)d\ell} \\
&= \frac{P(t,T)}{P(t,\tau)} \cdot e^{\int_\tau^T f(t,\ell)d\ell - \int_\tau^T f(\tau,\ell)d\ell} \\
&= \frac{P(t,T)}{P(t,\tau)} \cdot e^{-\int_\tau^T [f(\tau,\ell) - f(t,\ell)]d\ell}.
\end{aligned}$$

The key point of the HJM methodology is to take the current forward rates structure $f(t, \ell)$ as given, and to model the future forward rate movements,

$$f(\tau, \ell) - f(t, \ell).$$

[22] One of the many checks of the internal consistency of any model consists in checking that the given model produces: $\partial P(t,T)/\partial T = -\mathbb{E}[r(T) \exp(-\int_t^T r(u)du)] = -f(t,T)P(t,T)$ and by continuity, $\lim_{T \downarrow t} \partial P(t,T)/\partial T = -r(t) = -f(t,t)$.

Therefore, the HJM methodology takes the current term-structure as given and, hence, perfectly fitted, as we we observe both $P(t,T)$ and $P(t,\tau)$. In contrast, the other approach to interest rate modeling is to *model* the *current* bond price $P(t,T)$ by means of a model for the short-term rate (see Section 9.3) and, hence, does not fit the initial term structure. As we explained in the previous section, fitting the initial term-structure is an important issue when the model's user is concerned with pricing interest-rate derivatives.

9.5.2 The model

9.5.2.1 Primitives

The primitive is still a Brownian information structure. Therefore, if we want to model future movements of $\{f(\tau,T)\}_{\tau\in[t,T]}$, we also have to accept that for every T, $\{f(\tau,T)\}_{\tau\in[t,T]}$ is $\mathcal{F}(\tau)$-adapted. Under the Brownian information structure, there thus exist functionals α and σ such that, for any T,

$$d_\tau f(\tau,T)=\alpha(\tau,T)d\tau+\sigma(\tau,T)dW(\tau),\ \tau\in(t,T], \tag{9.50}$$

where $f(t,T)$ is given. The solution is thus

$$f(\tau,T)=f(t,T)+\int_t^\tau\alpha(s,T)ds+\int_t^\tau\sigma(s,T)dW(s),\ \tau\in(t,T]. \tag{9.51}$$

In other terms, W "doesn't depend" on T. If we wish to make W also "indexed" by T in some sense, we obtain the so-called *stochastic string models* (see Section 9.7).

9.5.2.2 Arbitrage restrictions

The next step is to derive restrictions on α that are consistent with absence of arbitrage opportunities. Let $X(\tau)\equiv-\int_\tau^T f(\tau,\ell)d\ell$. We have

$$dX(\tau)=f(\tau,\tau)d\tau-\int_\tau^T(d_\tau f(\tau,\ell))\,d\ell=\left[r(\tau)-\alpha^I(\tau,T)\right]d\tau-\sigma^I(\tau,T)dW(\tau),$$

where

$$\alpha^I(\tau,T)\equiv\int_\tau^T\alpha(\tau,\ell)d\ell\quad;\quad\sigma^I(\tau,T)\equiv\int_\tau^T\sigma(\tau,\ell)d\ell.$$

By Eq. (9.49), $P=e^X$. By Itô's lemma,

$$\frac{d_\tau P(\tau,T)}{P(\tau,T)}=\left[r(\tau)-\alpha^I(\tau,T)+\frac{1}{2}\left\|\sigma^I(\tau,T)\right\|^2\right]d\tau-\sigma^I(\tau,T)dW(\tau).$$

By the FTAP, there are no arbitrage opportunties if and only if

$$\frac{d_\tau P(\tau,T)}{P(\tau,T)}=\left[r(\tau)-\alpha^I(\tau,T)+\frac{1}{2}\left\|\sigma^I(\tau,T)\right\|^2+\sigma^I(\tau,T)\lambda(\tau)\right]d\tau-\sigma^I(\tau,T)d\tilde{W}(\tau),$$

where $\tilde{W}(\tau)=W(\tau)+\int_t^\tau\lambda(s)ds$ is a Q-Brownian motion, and λ satisfies:

$$\alpha^I(\tau,T)=\frac{1}{2}\left\|\sigma^I(\tau,T)\right\|^2+\sigma^I(\tau,T)\lambda(\tau). \tag{9.52}$$

By differentiating the previous relation with respect to T gives us the arbitrage restriction that we were looking for:

$$\alpha(\tau,T)=\sigma(\tau,T)\int_\tau^T\sigma(\tau,\ell)^\top d\ell+\sigma(\tau,T)\lambda(\tau). \tag{9.53}$$

9.5.3 The dynamics of the short-term rate

By Eq. (9.51), the short-term rate satisfies:

$$r(\tau) \equiv f(\tau,\tau) = f(t,\tau) + \int_t^\tau \alpha(s,\tau)ds + \int_t^\tau \sigma(s,\tau)dW(s),\ \tau \in (t,T]. \tag{9.54}$$

Differentiating with respect to τ yields

$$dr(\tau) = \left[f_2(t,\tau) + \sigma(\tau,\tau)\lambda(\tau) + \int_t^\tau \alpha_2(s,\tau)ds + \int_t^\tau \sigma_2(s,\tau)dW(s)\right]d\tau + \sigma(\tau,\tau)dW(\tau),$$

where

$$\alpha_2(s,\tau) = \sigma_2(s,\tau)\int_s^\tau \sigma(s,\ell)^\intercal d\ell + \sigma(s,\tau)\sigma(s,\tau)^\intercal + \sigma_2(s,\tau)\lambda(s).$$

As is clear, the short-term rate is in general non-Markov. However, the short-term rate can be "risk-neutralized" and used to price exotics through simulations.

9.5.4 Embedding

At first glance, it might be guessed that HJM models are quite distinct from the models of the short-term rate introduced in Section 9.3. However, there exist "embeddability" conditions turning HJM into short-term rate models, and viceversa, a property known as "universality" of HJM models.

9.5.4.1 Markovianity

One natural question to ask is whether there are conditions under which HJM-type models predict the short-term rate to be a Markov process. The question is natural insofar as it relates to the early literature in which the entire term-structure was driven by a scalar Markov process representing the dynamics of the short-term rate. The answer to this question is in the contribution of Carverhill (1994). Another important contribution in this area is due to Ritchken and Sankarasubramanian (1995), who studied conditions under which it is possible to enlarge the original state vector in such a manner that the resulting "augmented" state vector is Markov and at the same time, includes that short-term rate as a component. The resulting model resembles a lot some of the short-term rate models surveyed in Section 9.3. In these models, the short-term rate is not Markov, yet it is part of a system that is Markov. Here we only consider the simple Markov scalar case.

Assume the forward-rate volatility structure is deterministic and takes the following form:

$$\sigma(t,T) = g_1(t)g_2(T) \quad \text{all } t,T. \tag{9.55}$$

By Eq. (9.54), r is then:

$$r(\tau) = f(t,\tau) + \int_t^\tau \alpha(s,\tau)ds + g_2(\tau)\cdot\int_t^\tau g_1(s)dW(s),\ \tau \in (t,T],$$

Also, r is solution to:

$$
\begin{aligned}
dr(\tau) &= \left[f_2(t,\tau)+\sigma(\tau,\tau)\lambda(\tau)+\int_t^\tau \alpha_2(s,\tau)ds+g_2'(\tau)\int_t^\tau g_1(s)dW(s)\right]d\tau+\sigma(\tau,\tau)dW(\tau)\\
&= \left[f_2(t,\tau)+\sigma(\tau,\tau)\lambda(\tau)+\int_t^\tau \alpha_2(s,\tau)ds+\frac{g_2'(\tau)}{g_2(\tau)}g_2(\tau)\int_t^\tau g_1(s)dW(s)\right]d\tau+\sigma(\tau,\tau)dW(\tau)\\
&= \left[f_2(t,\tau)+\sigma(\tau,\tau)\lambda(\tau)+\int_t^\tau \alpha_2(s,\tau)ds+\frac{g_2'(\tau)}{g_2(\tau)}\left(r(\tau)-f(t,\tau)-\int_t^\tau \alpha(s,\tau)ds\right)\right]d\tau\\
&\quad +\sigma(\tau,\tau)dW(\tau).
\end{aligned}
$$

Done. This is Markov. Condition (9.55) is then a condition for the HJM model to predict that the short-term rate is Markov.

Mean-reversion is ensured by the condition that $g_2' < 0$, uniformly. As an example, take $\lambda =$ constant, and:

$$g_1(t)=\sigma\cdot e^{\kappa t},\quad \sigma>0 \qquad ; \qquad g_2(t)=e^{-\kappa t},\quad \kappa>0.$$

This is the Hull-White model discussed in Section 9.3.

9.5.4.2 Short-term rate reductions

We prove everything in the Markov case. Let the short-term rate be solution to:

$$dr(\tau)=\bar{b}(\tau,r(\tau))d\tau+a(\tau,r(\tau))d\tilde{W}(\tau),$$

where $\tilde{W}$ is a Q-Brownian motion, and $\bar{b}$ is some risk-neutralized drift function. The rational bond price function is $P(r(t),t,T)=\mathbb{E}\left[e^{-\int_t^T r(\tau)d\tau}\right]$. The forward rate implied by this model is:

$$f(r(t),t,T)=-\frac{\partial}{\partial T}\log P(r(t),t,T).$$

By Itô's lemma,

$$df=\left[\frac{\partial}{\partial t}f+\bar{b}f_r+\frac{1}{2}a^2 f_{rr}\right]d\tau+af_r d\tilde{W}.$$

But for $f(r,t,T)$ to be consistent with the solution to Eq. (9.51), it must be the case that

$$
\begin{aligned}
\alpha(t,T)-\sigma(t,T)\lambda(t) &= \frac{\partial}{\partial t}f(r,t,T)+\bar{b}(t,r)f_r(r,t,T)+\frac{1}{2}a(t,r)^2 f_{rr}(r,t,T)\\
\sigma(t,T) &= a(t,r)f_r(t,r)
\end{aligned}
\tag{9.56}
$$

and

$$f(t,T)=f(r,t,T). \tag{9.57}$$

In particular, the last condition can only be satisfied if the short-term rate model under consideration is of the perfectly fitting type.

9.6 Stochastic string shocks models

The first papers are Kennedy (1994, 1997), Goldstein (2000) and Santa-Clara and Sornette (2001). Heaney and Cheng (1984) are also very useful to read.

9.6.1 Addressing stochastic singularity

Let $\sigma(\tau,T)=[\sigma_1(\tau,T),\cdots,\sigma_N(\tau,T)]$ in Eq. (9.50). For any $T_1<T_2$,

$$E[df(\tau,T_1)\,df(\tau,T_2)]=\sum_{i=1}^N\sigma_i(\tau,T_1)\,\sigma_i(\tau,T_2)\,d\tau,$$

and,

$$c(\tau,T_1,T_2)\equiv corr[df(\tau,T_1)\,df(\tau,T_2)]=\frac{\sum_{i=1}^N\sigma_i(\tau,T_1)\,\sigma_i(\tau,T_2)}{\|\sigma(\tau,T_1)\|\cdot\|\sigma(\tau,T_2)\|}. \tag{9.58}$$

By replacing this result into Eq. (9.53),

$$\begin{aligned}\alpha(\tau,T) &= \int_\tau^T\sigma(\tau,T)\cdot\sigma(\tau,\ell)^\top d\ell+\sigma(\tau,T)\lambda(\tau)\\ &= \int_\tau^T\|\sigma(\tau,\ell)\|\,\|\sigma(\tau,T)\|\,c(\tau,\ell,T)\,d\ell+\sigma(\tau,T)\lambda(\tau).\end{aligned}$$

One drawback of this model is that the correlation matrix of any $(N+M)$-dimensional vector of forward rates is degenerate for $M\geq 1$. Stochastic string models overcome this difficulty by *modeling* in an independent way the correlation structure $c(\tau,\tau_1,\tau_2)$ for all τ_1 and τ_2 rather than *implying* it from a given N-factor model (as in Eq. (9.58)). In other terms, the HJM methodology uses functions σ_i to accommodate both volatility and correlation structure of forward rates. This is unlikely to be a good model in practice. As we will now see, stochastic string models have two separate functions with which to model volatility and correlation.

The starting point is a model in which the forward rate is solution to,

$$d_\tau f(\tau,T)=\alpha(\tau,T)\,d\tau+\sigma(\tau,T)\,d_\tau Z(\tau,T),$$

where the *string* Z satisfies the following five properties:

(i) *For all* τ, $Z(\tau,T)$ *is continuous in* T;

(ii) *For all* T, $Z(\tau,T)$ *is continuous in* τ;

(iii) $Z(\tau,T)$ *is a* τ*-martingale, and hence a local martingale i.e.* $E[d_\tau Z(\tau,T)]=0$;

(iv) $var[d_\tau Z(\tau,T)]=d\tau$;

(v) $cov[d_\tau Z(\tau,T_1)\,d_\tau Z(\tau,T_2)]=\psi(T_1,T_2)$ *(say).*

Properties (iii), (iv) and (v) make Z Markovian. The functional form for ψ is crucially important to guarantee this property. Given the previous properties, we can deduce a key property of the forward rates. We have,

$$\begin{aligned}\sqrt{var[df(\tau,T)]} &= \sigma(\tau,T)\\ c(\tau,T_1,T_2) &\equiv corr[df(\tau,T_1)\,df(\tau,T_2)]=\frac{\sigma(\tau,T_1)\,\sigma(\tau,T_2)\,\psi(T_1,T_2)}{\sigma(\tau,T_1)\,\sigma(\tau,T_2)}=\psi(T_1,T_2)\end{aligned}$$

As claimed before, we now have two separate functions with which to model volatility and correlation.

9.6.2 No-arbitrage restrictions

Similarly as in the HJM-Brownian case, let $X(\tau) \equiv -\int_\tau^T f(\tau,\ell)\,d\ell$. We have,

$$dX(\tau) = f(\tau,\tau)\,d\tau - \int_\tau^T d_\tau f(\tau,\ell)\,d\ell = \left[r(\tau)\,d\tau - \alpha^I(\tau,T)\right]d\tau - \int_\tau^T \left[\sigma(\tau,\ell)\,d_\tau Z(\tau,\ell)\right]d\ell,$$

where as usual, $\alpha^I(\tau,T) \equiv \int_\tau^T \alpha(\tau,\ell)\,d\ell$. But $P(\tau,T) = \exp(X(\tau))$. Therefore,

$$\begin{aligned}\frac{dP(\tau,T)}{P(\tau,T)} &= dX(\tau) + \frac{1}{2}var\left[dX(\tau)\right] \\ &= \left[r(\tau) - \alpha^I(\tau,T) + \frac{1}{2}\int_\tau^T\int_\tau^T \sigma(\tau,\ell_1)\,\sigma(\tau,\ell_2)\,\psi(\ell_1,\ell_2)\,d\ell_1 d\ell_2\right]d\tau \\ &\quad - \int_\tau^T \left[\sigma(\tau,\ell)\,d_\tau Z(\tau,\ell)\right]d\ell.\end{aligned}$$

Next, suppose that the pricing kernel ξ satisfies:

$$\frac{d\xi(\tau)}{\xi(\tau)} = -r(\tau)\,d\tau - \int_{\mathbb{T}} \phi(\tau,T)\,d_\tau Z(\tau,T)\,dT,$$

where $\mathbb{T}$ denotes the set of all "risks" spanned by the string Z, and ϕ is the corresponding family of "unit risk-premia".

By absence of arbitrage opportunities,

$$0 = E\left[d(P\xi)\right] = E\left[P\xi\cdot\left(\text{drift}\left(\frac{dP}{P}\right) + \text{drift}\left(\frac{d\xi}{\xi}\right) + cov\left(\frac{dP}{P},\frac{d\xi}{\xi}\right)\right)\right].$$

By exploiting the dynamics of P and ξ,

$$\alpha^I(\tau,T) = \frac{1}{2}\int_\tau^T\int_\tau^T \sigma(\tau,\ell_1)\,\sigma(\tau,\ell_2)\,\psi(\ell_1,\ell_2)\,d\ell_1 d\ell_2 + cov\left(\frac{dP}{P},\frac{d\xi}{\xi}\right),$$

where

$$\begin{aligned}cov\left(\frac{dP}{P},\frac{d\xi}{\xi}\right) &= E\left[\int_{\mathbb{T}} \phi(\tau,S)\,d_\tau Z(\tau,S)\,dS \cdot \int_\tau^T \sigma(\tau,\ell)\,d_\tau Z(\tau,\ell)\,d\ell\right] \\ &= \int_\tau^T\int_{\mathbb{T}} \phi(\tau,S)\,\sigma(\tau,\ell)\,\psi(S,\ell)\,dS d\ell.\end{aligned}$$

By differentiating α^I with respect to T we obtain,

$$\alpha(\tau,T) = \int_\tau^T \sigma(\tau,\ell)\,\sigma(\tau,T)\,\psi(\ell,T)\,d\ell + \sigma(\tau,T)\int_{\mathbb{T}} \phi(\tau,S)\,\psi(S,T)\,dS. \tag{9.59}$$

A proof of Eq. (9.59) is in the Appendix.

9.7 Interest rate derivatives

9.7.1 Introduction

Options on bonds, caps and swaptions are the main interest rate derivative instruments traded in the market. The purpose of this section is to price these assets. In principle, such a pricing problem could be solved very elegantly. Let w denote the value of any of such instrument, and π be the *instantaneous* payoff process paid by it. Consider a model of the short-term rate considered in Section 9.3. To simplify, assume that $d = 1$, and that all uncertainty is subsumed by the short-term rate process in Eq. (9.24). By the FTAP, w is then the solution to the following partial differential equation:

$$0 = \frac{\partial w}{\partial \tau} + \bar{b} w_r + \frac{1}{2} a^2 w_{rr} + \pi - rw, \text{ for all } (r, \tau) \in \mathbb{R}_{++} \times [t, T) \tag{9.60}$$

subject to some appropriate boundary conditions. In the previous PDE, $\bar{b}$ is some risk-neutralized drift function of the short-term rate. The additional π term arises because to the average instantaneous increase rate of the derivative, viz $\frac{\partial w}{\partial \tau} + \bar{b} w_r + \frac{1}{2} a^2 w_{rr}$, one has to add its payoff π. The sum of these two terms must equal rw to avoid arbitrage opportunities. In many applications considered below, the payoff π can be *approximated* by a function of the short-term rate itself $\pi(r)$. However, such an approximation is at odds with standard practice. Market participants define the payoffs of interest-rate derivatives in terms of LIBOR discretely-compounded rates. The aim of this section is to present more models that are more realistic than those emananating from Eq. (9.60).

The next section introduces notation to cope expeditiously with the pricing of these interest rate derivatives. Section 9.7.3 shows how to price options within the Gaussian models of the short-term rate discussed in Section 9.3. Section 9.7.4 provides precise definitions of the remaining most important fixed-income instruments: fixed coupon bonds, floating rate bonds, interest rate swaps, caps, floors and swaptions. It also provides exact solutions based on short-term rate models. Finally, Section 9.7.5 presents the "market model", which is a HJM-style model intensively used by practitioners.

9.7.2 Notation

We introduce some pieces of notation that will prove useful to price interest rate derivatives. For a given non decreasing sequence of dates $\{T_i\}_{i=0,1,\cdots}$, we set,

$$F_i(\tau) \equiv F(\tau, T_i, T_{i+1}). \tag{9.61}$$

That is, $F_i(\tau)$ is solution to:

$$\frac{P(\tau, T_{i+1})}{P(\tau, T_i)} = \frac{1}{1 + \delta_i F_i(\tau)}. \tag{9.62}$$

An obvious but important relation is

$$F_i(T_i) = L(T_i). \tag{9.63}$$

9.7.3 European options on bonds

Let T be the expiration date of a European call option on a bond and $S > T$ be the expiration date of the bond. We consider a simple model of the short-term rate with $d = 1$, and a rational *bond pricing function* of the form $P(\tau) \equiv P(r, \tau, S)$. We also consider a rational *option price function* $C^b(\tau) \equiv C^b(r, \tau, T, S)$. By the FTAP, there are no arbitrage opportunities if and only if,

$$C^b(t) = \mathbb{E}\left[e^{-\int_t^T r(\tau)d\tau} \left(P(r(T), T, S) - K\right)^+\right], \tag{9.64}$$

where K is the strike of the option. In terms of PDEs, C^b is solution to Eq. (9.60) with $\pi \equiv 0$ and boundary condition $C^b(r, T, T, S) = (P(r, T, S) - K)^+$, where $P(r, \tau, S)$ is also the solution to Eq. (9.60) with $\pi \equiv 0$, but with boundary condition $P(r, S, S) = 1$. In terms of PDEs, the situation seems hopeless. As we show below, the problem can be considerably simplified with the help of the T-forward martingale measure introduced in Section 9.1. Indeed, we shall show that under the assumption that the short-term rate is a Gaussian process, Eq. (9.64) has a closed-form expression. We now present two models enabling this. The first one was developed in a seminal paper by Jamshidian (1989), and the second one is, simply, its perfectly fitting extension.

9.7.3.1 Jamshidian & Vasicek

Suppose that the short-term rate is solution to the Vasicek's model considered in Section 9.3 (see Eq. (9.28)):

$$dr(\tau) = (\theta - \kappa r(\tau))d\tau + \sigma d\tilde{W},$$

where $\tilde{W}$ is a Q-Brownian motion and $\theta \equiv \bar{\theta} - \sigma\lambda$. As shown in Section 9.3 (see Eq. (9.30)), the bond price takes the following form:

$$P(r(\tau), \tau, S) = e^{A(\tau,S) - B(\tau,S)r(\tau)},$$

for some function A; and for $B(t, T) = \frac{1}{\kappa}\left[1 - e^{-\kappa(T-t)}\right]$ (see formula (9.47)).

In Section 9.3 (see formula (9.39)), it was also shown that

$$\begin{aligned}
&\mathbb{E}\left[e^{-\int_t^T r(\tau)d\tau} \left(P(r(T), T, S) - K\right)^+\right] \\
&= P(r(t), t, S) \cdot Q_F^S\left[P(r(T), T, S) \geq K\right] - KP(r(t), t, T) \cdot Q_F^T\left[P(r(T), T, S) \geq K\right],
\end{aligned} \tag{9.65}$$

where Q_F^T denotes the T-forward martingale measure (see Section 9.1.4).

In Appendix 8, we show that the two probabilities in Eq. (9.65) can be evaluated by the *change of numeraire* device described in Section 9.1.4. Precisely, Appendix 8 shows that the solution for $P(r, T, S)$ can be written as:

$$\begin{aligned}
P(T, S) &= \frac{P(t, S)}{P(t, T)} e^{-\frac{1}{2}\sigma^2 \int_t^T [B(\tau,S) - B(\tau,T)]^2 d\tau + \sigma \int_t^T [B(\tau,S) - B(\tau,T)] dW^{Q_F^T}(\tau)} \quad \text{under } Q_F^T \\
P(T, S) &= \frac{P(t, S)}{P(t, T)} e^{\frac{1}{2}\sigma^2 \int_t^T [B(\tau,S) - B(\tau,T)]^2 d\tau + \sigma \int_t^T [B(\tau,S) - B(\tau,T)] dW^{Q_F^S}(\tau)} \quad \text{under } Q_F^S
\end{aligned} \tag{9.66}$$

where $W^{Q_F^T}$ is a Brownian motion under the forward measure Q_F^T, and $P(\tau, m) \equiv P(r, \tau, m)$. Therefore, simple algebra now reveals that:

$$\begin{aligned}
Q_F^S\left[P(T, S) \geq K\right] &= \Phi(d_1), \\
Q_F^T\left[P(T, S) \geq K\right] &= \Phi(d_2),
\end{aligned}$$

where

$$d_1 = \frac{\log\left[\frac{P(t,S)}{KP(t,T)}\right] + \frac{1}{2}v^2}{v}; \quad d_2 = d_1 - v; \quad v^2 = \sigma^2 \int_t^T \left[B(\tau,S) - B(\tau,T)\right]^2 d\tau.$$

Finally, by simple but tedious computations,

$$v^2 = \sigma^2 \frac{1 - e^{-2\kappa(T-t)}}{2\kappa} B(T,S)^2.$$

9.7.3.2 Perfectly fitting extension

We now consider the perfectly fitting extension of the previous results. Namely, we consider model (9.45) in Section 9.3, viz

$$dr(\tau) = (\theta(\tau) - \kappa r(\tau))d\tau + \sigma d\tilde{W}(\tau),$$

where $\theta(\tau)$ is now the infinite dimensional parameter that is used to "invert the term-structure".

The solution to Eq. (9.64) is the same as in the previous section. However, in Section 9.7.3 we shall argue that the advantage of using such a perfectly fitting extension arises as soon as one is concerned with the evaluation of more complex options on fixed coupon bonds.

9.7.4 Related pricing problems

9.7.4.1 Fixed coupon bonds

Given a set of dates $\{T_i\}_{i=0}^n$, a fixed coupon bond pays off a fixed *coupon* c_i at T_i, $i = 1, \cdots, n$ and one unit of numéraire at time T_n. Ideally, one generic coupon at time T_i pays off for the time-interval $T_i - T_{i-1}$. It is assumed that the various coupons are known at time $t < T_0$. By the FTAP, the value of a fixed coupon bond is

$$P_{\text{fcp}}(t, T_n) = P(t, T_n) + \sum_{i=1}^n c_i P(t, T_i).$$

9.7.4.2 Floating rate bonds

A floating rate bond works as a fixed coupon bond, with the important exception that the coupon payments are defined as:

$$c_i = \delta_{i-1} L(T_{i-1}) = \frac{1}{P(T_{i-1}, T_i)} - 1, \tag{9.67}$$

where $\delta_i \equiv T_{i+1} - T_i$, and where the second equality is the definition of the simply-compounded LIBOR rates introduced in Section 9.1 (see Eq. (9.2)). By the FTAP, the price p_{frb} as of time t of a floating rate bond is:

$$\begin{aligned}
p_{\text{frb}}(t) &= P(t, T_n) + \sum_{i=1}^n \mathbb{E}\left[e^{-\int_t^{T_i} r(\tau)d\tau} \delta_{i-1} L(T_{i-1})\right] \\
&= P(t, T_n) + \sum_{i=1}^n \mathbb{E}\left[\frac{e^{-\int_t^{T_i} r(\tau)d\tau}}{P(T_{i-1}, T_i)}\right] - \sum_{i=1}^n P(t, T_i) \\
&= P(t, T_n) + \sum_{i=1}^n P(t, T_{i-1}) - \sum_{i=1}^n P(t, T_i) \\
&= P(t, T_0).
\end{aligned}$$

where the second line follows from Eq. (9.67) and the third line from Eq. (9.7) given in Section 9.1.

The same result can be obtained by assuming an economy in which the floating rates continuously pay off the instantaneous short-term rate r. Let $T_0 = t$ for simplicity. In this case, p_{frb} is solution to the PDE (9.60), with $\pi(r) = r$, and boundary condition $p_{\text{frb}}(T) = 1$. As it can verified, $p_{\text{frb}} = 1$ all r and τ is indeed solution to the PDE (9.60).

9.7.4.3 Options on fixed coupon bonds

The payoff of an option maturing at T_0 on a fixed coupon bond paying off at dates $T_1, \cdots, T_n$ is given by:

$$[P_{\text{fcp}}(T_0, T_n) - K]^+ = \left[P(T_0, T_n) + \sum_{i=1}^{n} c_i P(T_0, T_i) - K \right]^+ . \tag{9.68}$$

At first glance, the expectation of the payoff in Eq. (9.68) seems very difficult to evaluate. Indeed, even if we end up with a model that predicts bond prices at time T_0, $P(T_0, T_i)$, to be lognormal, we know that the sum of lognormal is not lognormal. However, there is an elegant way to solve this problem. Suppose we wish to model the bond price $P(t, T)$ through any one of the models of the short-term rate reviewed in Section 9.3. In this case, the pricing function is obviously $P(t, T) = P(r, t, T)$. Assume, further, that

$$\text{For all } t, T, \ \frac{\partial P(r, t, T)}{\partial r} < 0, \tag{9.69}$$

and that

$$\text{For all } t, T, \ \lim_{r \to 0} P(r, t, T) > K \ \text{ and } \ \lim_{r \to \infty} P(r, t, T) = 0. \tag{9.70}$$

Under conditions (G3) and (9.70), there is one and only one value of r, say r^*, that solves the following equation:

$$P(r^*, T_0, T_n) + \sum_{i=1}^{n} c_i P(r^*, T_0, T_i) = K. \tag{9.71}$$

The payoff in Eq. (9.68) can then be written as:

$$\left[\sum_{i=1}^{n} \bar{c}_i P(r(T_0), T_0, T_i) - K \right]^+ = \left\{ \sum_{i=1}^{n} \bar{c}_i \left[P(r(T_0), T_0, T_i) - P(r^*, T_0, T_i) \right] \right\}^+ ,$$

where $\bar{c}_i = c_i$, $i = 1, \cdots, n-1$, and $\bar{c}_n = 1 + c_n$.

Next notice that by condition (G3), the terms $P(r(T_0), T_0, T_i) - P(r^*, T_0, T_i)$ share all the same sign for all i.[23] Therefore, the payoff in Eq. (9.68) is,

$$\left[\sum_{i=1}^{n} \bar{c}_i P(r(T_0), T_0, T_i) - K \right]^+ = \sum_{i=1}^{n} \bar{c}_i \left[P(r(T_0), T_0, T_i) - P(r^*, T_0, T_i) \right]^+ . \tag{9.72}$$

Next, note that every i-th term in Eq. (9.72) can be evaluated as an option on a pure discount bond with strike price equal to $P(r^*, T_0, T_i)$. Typically, the threshold r^* must be found with some numerical method. The device to *reduce* the problem of an option on a fixed coupon bond

[23] Suppose that $P(r(T_0), T_0, T_1) > P(r^*, T_0, T_1)$. By Eq. (G3), $r(T_0) < r^*$. Hence $P(r(T_0), T_0, T_2) > P(r^*, T_0, T_2)$, etc.

to a problem involving the sum of options on zero coupon bonds was invented by Jamshidian (1989).

Is this reduction likely to work, in practice? In other words, are the conditions in Eqs. (G3) and (9.70) likely to be satisfied, in practice? Do they hold, in general? In the Vasicek's model that Jamshidian considered in his paper, the bond price is $P(r,t,T) = e^{A(t,T)-B(t,T)r}$ and Eq. (G3) holds true as $B > 0$.[24] In fact, it is possible to show that the condition in Eq. (G3) holds for all one-factor stationary, Markov models of the short-term rate. However, the condition in Eq. (G3) is not a general property of bond prices in multi-factor models (see Mele (2003)).

We conclude this section and illustrate the importance of perfectly fitting models. Suppose that in Eq. (9.71), the critical value r^* is computed by means of the Vasicek's model. This assumption is attractive because in this case, the payoff in Eq. (9.72) could be evaluated by means of the Jamshidian's formula introduced in Section 9.7.2. However, this way to proceed does not ensure that the yield curve is perfectly fitted. The natural alternative is to use the corresponding perfectly fitting extension. However, such a perfectly fitting extension gives rise to a zero-coupon bond option price that is perfectly equal to the one that can be obtained through the Jamshidian's formula. However, things differ as far as options on zero coupon bonds are concerned. Indeed, by using the perfectly fitting model (9.45), one obtains bond prices such that the solution r^* in Eq. (9.71) is radically different from the one obtained when bond prices are obtained with the simple Vasicek's model.

9.7.4.4 Interest rate swaps

An interest rate swap is an exchange of interest rate payments. Typically, one counterparty exchanges a fixed against a floating interest rate payment. For example, the counterparty receiving a floating interest rate payment has "good" (or only) access to markets for "variable" interest rates, but wishes to pay fixed interest rates. And viceversa. The counterparty receiving a floating interest rate payment and paying a fixed interest rate K has a payoff equal to,

$$\delta_{i-1}\left[L(T_{i-1}) - K\right]$$

at time T_i, $i = 1, \cdots, n$. By the FTAP, the value p_{irs} as of time t of an interest rate swap is:

$$p_{\text{irs}}(t) = \sum_{i=1}^{n} \mathbb{E}\left[e^{-\int_t^{T_i} r(\tau)d\tau}\delta_{i-1}\left(L(T_{i-1}) - K\right)\right] = -\sum_{i=1}^{n} A(t, T_{i-1}, T_i; K),$$

where A is the value of a forward-rate agreement and is, by Eq. (9.8) in Section 9.1,

$$A\left(t, T_{i-1}, T_i; K\right) = \delta_{i-1}\left[K - F\left(t, T_{i-1}, T_i\right)\right] P\left(t, T_i\right).$$

The *forward swap rate* R_{swap} is defined as the value of K such that $p_{\text{irs}}(t) = 0$. Simple computations yield:

$$R_{\text{swap}}(t) = \frac{\sum_{i=1}^{n} \delta_{i-1} F(t, T_{i-1}, T_i) P(t, T_i)}{\sum_{i=1}^{n} \delta_{i-1} P(t, T_i)} = \frac{P(t, T_0) - P(t, T_n)}{\sum_{i=1}^{n} \delta_{i-1} P(t, T_i)}, \tag{9.73}$$

where the last equality is due to the definition of $F(t, T_{i-1}, T_i)$ given in Section 9.1 (see Eq. (9.62)).

[24] In this context, the second part of the condition in Eq. (9.70) also holds true. Moreover, the first part of condition (9.70) should also hold, in practice.

Finally, note that this case could have also been solved by casting it in the format of the PDE (9.60). It suffices to consider continuous time swap exchanges, to set $p_{\text{irs}}(T) \equiv 0$ as a boundary condition, and to set $\pi(r) = r - k$, where k plays the same role as K above. It is easy to see that if the bond price $P(\tau)$ is solution to (9.60) with its usual boundary condition, the following function

$$p_{\text{irs}}(\tau) = 1 - P(\tau) - k \int_{\tau}^{T} P(s) ds$$

does satisfy (9.60) with $\pi(r) = r - k$.

9.7.4.5 Caps & floors

A cap works as an interest rate swap, with the important exception that the exchange of interest rates payments takes place only if actual interest rates are higher than K. A cap protects against upward movements of the interest rates. Therefore, we have that the payoff as of time T_i is

$$\delta_{i-1} \left[L\left(T_{i-1}\right) - K \right]^{+}, \quad i = 1, \cdots, n.$$

Such a payoff is usually referred to as the *caplet*.

Floors are defined in a similar way, with a single *floorlet* paying off,

$$\delta_{i-1} \left[K - L(T_{i-1}) \right]^{+}$$

at time T_i, $i = 1, \cdots, n$.

We will only focus on caps. By the FTAP, the value p_{cap} of a cap as of time t is:

$$\begin{aligned} p_{\text{cap}}(t) &= \sum_{i=1}^{n} \mathbb{E}\left[e^{-\int_t^{T_i} r(\tau) d\tau} \delta_{i-1} \left(L(T_{i-1}) - K \right)^{+} \right] \\ &= \sum_{i=1}^{n} \mathbb{E}\left[e^{-\int_t^{T_i} r(\tau) d\tau} \delta_{i-1} \left(F(T_{i-1}, T_{i-1}, T_i) - K \right)^{+} \right]. \end{aligned} \tag{9.74}$$

Models of the short-term rate can be used to give an explicit solution to this pricing problem. First, we use the standard definition of simply compounded rates given in Section 9.1 (see formula (9.2)), viz $\delta_{i-1} L(T_{i-1}) = \frac{1}{P(T_{i-1}, T_i)} - 1$, and rewrite the caplet payoff as follows:

$$\left[\delta_{i-1} L(T_{i-1}) - \delta_{i-1} K \right]^{+} = \frac{1}{P(T_{i-1}, T_i)} \left[1 - (1 + \delta_{i-1} K) P(T_{i-1}, T_i) \right]^{+}.$$

We have,

$$\begin{aligned} p_{\text{cap}}(t) &= \sum_{i=1}^{n} \mathbb{E}\left\{ \frac{e^{-\int_t^{T_i} r(\tau) d\tau}}{P(T_{i-1}, T_i)} \left[1 - (1 + \delta_{i-1} K) P(T_{i-1}, T_i) \right]^{+} \right\} \\ &= \sum_{i=1}^{n} \mathbb{E}\left\{ e^{-\int_t^{T_{i-1}} r(\tau) d\tau} \left[1 - (1 + \delta_{i-1} K) P(T_{i-1}, T_i) \right]^{+} \right\}, \end{aligned}$$

where the last equality follows by a simple computation.[25] If bond prices are as in Jamshidian or in Hull and White, the previous pricing problem has an analytical solution.

Finally, note that the pricing problems associated with caps and floors could also have been approximated by casting them in the format of the PDE (9.60), with $\pi(r) = (r-k)^+$ (caps) and $\pi(r) = (k-r)^+$ (floors), and where k plays the same role played by K above.

9.7.4.6 Swaptions

Swaptions are options to enter a swap contract on a future date. Let the maturity date of this option be T_0. Then, at time T_0, the payoff of the swaption is the maximum between zero and the value of an interest rate swap at T_0 $p_{\text{irs}}(T_0)$, viz

$$(p_{\text{irs}}(T_0))^+ = \left[-\sum_{i=1}^n A(T_0, T_{i-1}, T_i; K)\right]^+ = \left[\sum_{i=1}^n \delta_{i-1}\left(F(T_0,T_{i-1},T_i) - K\right) P(T_0,T_i)\right]^+. \quad (9.75)$$

By the FTAP, the value p_{swaption} as of time t of a swaption is:

$$\begin{aligned} p_{\text{swaption}}(t) &= \mathbb{E}\left\{e^{-\int_t^{T_0} r(\tau)d\tau}\left[\sum_{i=1}^n \delta_{i-1}\left(F(T_0,T_{i-1},T_i) - K\right) P(T_0,T_i)\right]^+\right\} \\ &= \mathbb{E}\left\{e^{-\int_t^{T_0} r(\tau)d\tau}\left[1 - P(T_0,T_n) - \sum_{i=1}^n c_i P(T_0,T_i)\right]^+\right\}, \end{aligned} \quad (9.76)$$

where $c_i \equiv \delta_{i-1}K$, and where we used the relation $\delta_{i-1}F(T_0,T_{i-1},T_i) = \frac{P(T_0,T_{i-1})}{P(T_0,T_i)} - 1$.

Eq. (9.76) is the expression of the price of a *put* option *on a fixed coupon bond* struck at one. Therefore, we see that this pricing problem can be dealt with the same technology used to deal with the short-term rate models. In particular, all remarks made there as regards the difference between simple short-term rate models and perfectly fitting short-term rate models also apply here.

9.7.5 Market models

9.7.5.1 Models and market practice reality

As demonstrated in the previous sections, models of the short-term rate can be used to obtain closed-form solutions of virtually every important product of the interest rates derivatives business. The typical examples are the Vasicek's model and its perfectly fitting extension. Yet market practitioners has been evaluating caps through Black's (1976) formula for years. The

[25] By the law of iterated expectations,

$$\begin{aligned} \mathbb{E}\left[\frac{e^{-\int_t^{T_i} r(\tau)d\tau}}{P(T_{i-1},T_i)}\left[1-(1+\delta_{i-1}K)P(T_{i-1},T_i)\right]^+\right] &= \mathbb{E}\left[\mathbb{E}\left(\frac{e^{-\int_t^{T_i} r(\tau)d\tau}}{P(T_{i-1},T_i)}\left[1-(1+\delta_{i-1}K)P(T_{i-1},T_i)\right]^+\middle|\mathcal{F}(T_i)\right)\right] \\ &= \mathbb{E}\left[\mathbb{E}\left(e^{-\int_t^{T_i} r(\tau)d\tau}e^{\int_{T_{i-1}}^{T_i} r(\tau)d\tau}\left[1-(1+\delta_{i-1}K)P(T_{i-1},T_i)\right]^+\middle|\mathcal{F}(T_i)\right)\right] \\ &= \mathbb{E}\left[\mathbb{E}\left(e^{-\int_t^{T_{i-1}} r(\tau)d\tau}\left[1-(1+\delta_{i-1}K)P(T_{i-1},T_i)\right]^+\middle|\mathcal{F}(T_i)\right)\right] \\ &= \mathbb{E}\left[e^{-\int_t^{T_{i-1}} r(\tau)d\tau}\left[1-(1+\delta_{i-1}K)P(T_{i-1},T_i)\right]^+\right] \end{aligned}$$

assumption of this market practice is that the simply-compounded forward rate is lognormally distributed. As it turns out, the analytically tractable (Gaussian) short-term rate models are *not* consistent with this assumption. Clearly, the (Gaussian) Vasicek's model does not predict that the simply-compounded forward rates are Geometric Brownian motions.[26]

Can a non-Markovian HJM model address this problem? Yes. However, a practical difficulty arising with the HJM approach is that *instantaneous* forward rates are not observed. Does this compromise the practical appeal of the HJM methodology to the pricing of caps and floors - which constitute an important portion of the interest rates business? No. Brace, Gatarek and Musiela (1997), Jamshidian (1997) and Miltersen, Sandmann and Sondermann (1997) observed that general HJM framework can be "forced" to address some of the previous difficulties.

The key feature of the models identified by these authors is the emphasis on the dynamics of the *simply-compounded* forward rates. An additional, and technical, assumption is that these simply-compounded forward rates are lognormal under the risk-neutral probability Q. That is, given a non-decreasing sequence of reset times $\{T_i\}_{i=0,1,\cdots}$, each simply-compounded rate, F_i, is solution to the following stochastic differential equation:[27]

$$\frac{dF_i(\tau)}{F_i(\tau)} = m_i(\tau)d\tau + \gamma_i(\tau)d\tilde{W}(\tau), \quad \tau \in [t, T_i], \quad i = 0, \cdots, n-1, \tag{9.77}$$

where $F_i(\tau) \equiv F(\tau, T_i, T_{i+1})$, and m_i and γ_i are some deterministic functions of time (γ_i is vector valued). On a mathematical point of view, that assumption that F_i follows Eq. (9.77) is innocous.[28]

As we shall show, this simple framework can be used to use the simple Black's (1976) formula to price caps and floors. However, we need to emphasize that there is nothing wrong with the short-term rate models analyzed in previous sections. The real advance of the so-called market model is to give a rigorous foundation to the standard market practice to price caps and floors by means of the Black's (1976) formula.

9.7.5.2 Simply-compounded forward rate dynamics

By the definition of simply-compounded forward rates (9.62),

$$\ln\left[\frac{P(\tau, T_i)}{P(\tau, T_{i+1})}\right] = \ln\left[1 + \delta_i F_i(\tau)\right]. \tag{9.78}$$

The logic to be followed here is exactly the same as in the HJM representation of Section 9.4. The objective is to express bond prices volatility in terms of forward rates volatility. To achieve this task, we first assume that bond prices are driven by Brownian motions and expand the l.h.s. of Eq. (9.78) (step 1). Then, we expand the r.h.s. of Eq. (9.78) (step 2). Finally, we identify the two diffusion terms derived from the previous two steps (step 3).

Step 1: Let $P_i \equiv P(\tau, T_i)$, and assume that under the risk-neutral measure Q, P_i is solution to:

$$\frac{dP_i}{P_i} = rd\tau + \sigma_{bi}d\tilde{W}.$$

[26]Indeed, $1 + \delta_i F_i(\tau) = \frac{P(\tau,T_i)}{P(\tau,T_{i+1})} = \exp\left[\Delta A_i(\tau) - \Delta B_i(\tau) r(\tau)\right]$, where $\Delta A_i(\tau) = A(\tau, T_i) - A(\tau, T_{i+1})$, and $\Delta B_i(\tau) = B(\tau, T_i) - B(\tau, T_{i+1})$. Hence, $F_i(\tau)$ is not a Geometric Brownian motion, despite the fact that the short-term rate r is Gaussian and, hence, the bond price is log-normal. Black '76 can not be applied in this context.

[27]Brace, Gatarek and Musiela (1997) derived their model by specifying the dynamics of the spot simply-compounded Libor interest rates. Since $F_i(T_i) = L(T_i)$ (see Eq. (9.63)), the two derivations are essentially the same.

[28]It is well-known that lognormal *instantaneous* forward rates create mathematical problems to the money market account (see, for example, Sandmann and Sondermann (1997) for a succinct overview on how this problem is easily handled with *simply-compounded* forward rates).

In terms of the HJM framework in Section 9.4,

$$\sigma_{bi}(\tau) = -\sigma^I(\tau, T_i) = -\int_\tau^{T_i} \sigma(\tau, \ell) d\ell, \tag{9.79}$$

where $\sigma(\tau, \ell)$ is the instantaneous volatility of the instantaneous ℓ-forward rate as of time τ. By Itô's lemma,

$$d\ln\left[\frac{P(\tau, T_i)}{P(\tau, T_{i+1})}\right] = -\frac{1}{2}\left[\|\sigma_{bi}\|^2 - \|\sigma_{b,i+1}\|^2\right] d\tau + (\sigma_{bi} - \sigma_{b,i+1})\, d\tilde{W}. \tag{9.80}$$

Step 2: Applying Itô's lemma to $\ln[1 + \delta_i F_i(\tau)]$ and using Eq. (9.77) yields:

$$\begin{aligned} d\ln[1 + \delta_i F_i(\tau)] &= \frac{\delta_i}{1 + \delta_i F_i} dF_i - \frac{1}{2}\frac{\delta_i^2}{(1 + \delta_i F_i)^2}(dF_i)^2 \\ &= \left[\frac{\delta_i m_i F_i}{1 + \delta_i F_i} - \frac{1}{2}\frac{\delta_i^2 F_i^2 \|\gamma_i\|^2}{(1 + \delta_i F_i)^2}\right] d\tau + \frac{\delta_i F_i}{1 + \delta_i F_i}\gamma_i d\tilde{W}. \end{aligned} \tag{9.81}$$

Step 3: By Eq. (9.78), the diffusion terms in Eqs. (9.80) and (9.81) have to be the same. Therefore,

$$\sigma_{bi}(\tau) - \sigma_{b,i+1}(\tau) = \frac{\delta_i F_i(\tau)}{1 + \delta_i F_i(\tau)}\gamma_i(\tau), \quad \tau \in [t, T_i].$$

By summing over i, we get the following no-arbitrage restriction for bond price volatility:

$$\sigma_{bi}(\tau) - \sigma_{b,0}(\tau) = -\sum_{j=0}^{i-1} \frac{\delta_j F_j(\tau)}{1 + \delta_j F_j(\tau)}\gamma_j(\tau). \tag{9.82}$$

It should be clear that the relation obtained before is only a specific restriction on the general HJM framework. Indeed, assume that the instantaneous forward rates are as in Eq. (9.50) of Section 9.4. As we demonstrated in Section 9.4, then, bond prices volatility is given by Eq. (9.79). But if we also assume that simply-compounded rates are solution to Eq. (9.77), then bond prices volatility is also equal to Eq. (9.82). Comparing Eq. (9.79) with Eq. (9.82) produces,

$$\int_{T_0}^{T_i} \sigma(\tau, \ell) d\ell = \sum_{j=0}^{i-1} \frac{\delta_j F_j(\tau)}{1 + \delta_j F_j(\tau)}\gamma_j(\tau).$$

The practical interest to restrict the forward-rate volatility dynamics in this way lies in the possibility to obtain closed-form solutions for some of the interest rates derivatives surveyed in Section 9.7.3.

9.7.5.3 Pricing formulae

Caps & Floors

We provide the solution for caps only. We have:

$$\begin{aligned} p_{\text{cap}}(t) &= \sum_{i=1}^n \mathbb{E}\left[e^{-\int_t^{T_i} r(\tau)d\tau}\delta_{i-1}\left(F(T_{i-1}, T_{i-1}, T_i) - K\right)^+\right] \\ &= \sum_{i=1}^n \delta_{i-1} P(t, T_i) \cdot \mathbb{E}_{Q_F^{T_i}}\left[F(T_{i-1}, T_{i-1}, T_i) - K\right]^+, \end{aligned} \tag{9.83}$$

where $\mathbb{E}_{Q_F^{T_i}}[\cdot]$ denotes, as usual, the expectation taken under the T_i-forward martingale measure $Q_F^{T_i}$; the first equality is Eq. (9.74); and the second equality has been obtained through the usual change of measure technique introduced Section 9.1.4.

The key point is that

$$F_{i-1}(\tau) \equiv F_{i-1}(\tau, T_{i-1}, T_i),\ \tau \in [t, T_{i-1}],\ \textit{is a martingale under measure } Q_F^{T_i}.$$

A proof of this statement is in Section 9.1. By Eq. (9.77), this means that $F_{i-1}(\tau)$ is solution to

$$\frac{dF_{i-1}(\tau)}{F_{i-1}(\tau)} = \gamma_{i-1}(\tau) dW^{Q_F^{T_i}}(\tau),\quad \tau \in [t, T_{i-1}],\quad i = 1, \cdots, n,$$

under measure $Q_F^{T_i}$.

The pricing problem in Eq. (9.83) is then reduced to the standard Black's (1976) one. The result is

$$\mathbb{E}_{Q_F^{T_i}}[F(T_{i-1}, T_{i-1}, T_i) - K]^+ = F_{i-1}(t)\Phi(d_{1,i-1}) - K\Phi(d_{2,i-1}), \tag{9.84}$$

where

$$d_{1,i-1} = \frac{\log\left[\frac{F_{i-1}(t)}{K}\right] + \frac{1}{2}s^2}{s};\quad d_{2,i-1} = d_{1,i-1} - s;\quad s^2 = \int_t^{T_{i-1}} \gamma_{i-1}(\tau)^2 d\tau.$$

For sake of completeness, Appendix 8 provides the derivation of the Black's formula in the case of deterministic time-varying volatility γ.

Swaptions

By (9.75), and the equation defining the forward swap rate R_{swap} (see Eq. (9.73)), the payoff of the swaption can be written as:

$$\left[\sum_{i=1}^n \delta_{i-1}\left(F(T_0, T_{i-1}, T_i) - K\right) P(T_0, T_i)\right]^+ = \sum_{i=1}^n \delta_{i-1} P(T_0, T_i) \left(R_{\text{swap}}(T_0) - K\right)^+.$$

Therefore, by the FTAP, and a change of measure,

$$\begin{aligned} p_{\text{swaption}}(t) &= \mathbb{E}\left[e^{-\int_t^{T_0} r(\tau)d\tau} \cdot \sum_{i=1}^n \delta_{i-1} P(T_0, T_i) \left(R_{\text{swap}}(T_0) - K\right)^+\right] \\ &= P(t, T_0) \cdot \mathbb{E}_{Q_F^{T_0}}\left[\sum_{i=1}^n \delta_{i-1} P(T_0, T_i) \left(R_{\text{swap}}(T_0) - K\right)^+\right]. \end{aligned}$$

This can be dealt with through the so-called *forward swap measure.* Define the forward swap measure Q^{swap} by:

$$\frac{dQ^{\text{swap}}}{dQ_F^{T_0}} = \frac{\sum_{i=1}^n \delta_{i-1} P(T_0, T_i)}{\mathbb{E}_{Q_F^{T_0}}\left[\sum_{i=1}^n \delta_{i-1} P(T_0, T_i)\right]} = P(t, T_0) \frac{\sum_{i=1}^n \delta_{i-1} P(T_0, T_i)}{\sum_{i=1}^n \delta_{i-1} P(t, T_i)},$$

where the last equality follows from the following elementary facts: $P(t, T_i) = \mathbb{E}\left[e^{-\int_t^{T_0} r(\tau)d\tau} P(T_0, T_i)\right] = P(t, T_0)\mathbb{E}_{Q_F^{T_0}}[P(T_0, T_i)]$ (The first equality is the FTAP, the second is a change of measure.)

Therefore,

$$\begin{aligned} p_{\text{swaption}}(t) &= P(t,T_0)\cdot \mathbb{E}_{Q_F^{T_0}}\left[\sum_{i=1}^{n}\delta_{i-1}P(T_0,T_i)\left(R_{\text{swap}}(T_0)-K\right)^+\right] \\ &= \mathbb{E}_{Q^{\text{swap}}}\left[\sum_{i=1}^{n}\delta_{i-1}P(t,T_i)\left(R_{\text{swap}}(T_0)-K\right)^+\right]. \end{aligned}$$

Furthermore, the forward swap rate R_{swap} is a Q^{swap}-martingale.[29] And naturally, it is positive. Therefore, it must satisfy:

$$\frac{dR_{\text{swap}}(\tau)}{R_{\text{swap}}(\tau)} = \gamma^{\text{swap}}(\tau)dW^{\text{swap}}(\tau), \quad \tau\in[t,T_0],$$

where W^{swap} is a Q^{swap}-Brownian motion, and $\gamma^{\text{swap}}(\tau)$ is adapted.

If $\gamma^{\text{swap}}(\tau)$ is deterministic, use Black's (1976) to price the swaption in closed-form.

Inconsistencies

The trouble is that if F is solution to Eq. (9.77), γ^{swap} can not be deterministic. And as you may easily conjecture, if you assume that forward swap rates are lognormal, then you don't end up with Eq. (9.77). Therefore, you may use Black 76 to price either caps or swaptions, not both. I believe this limits considerably the importance of market models. A couple of tricks that seem to work in practice. The best known trick is based on a suggestion by Rebonato (1998) to replace the true pricing problem with an approximating pricing problem in which γ^{swap} is deterministic. That works in practice, but in a world with stochastic volatility, I expect that trick to generate unstable things in periods experiencing highly volatile volatility. See, also, Rebonato (1999) for another very well written essay on related issues. The next section suggests to use numerical approximation based on Montecarlo techniques.

9.7.5.4 Numerical approximations

Suppose forward rates are lognormal. Then you price caps with Black 76. As regards swaptions, you may wish to implement Montecarlo integration as follows.

By a change of measure,

$$\begin{aligned} p_{\text{swaption}}(t) &= \mathbb{E}\left\{e^{-\int_t^{T_0} r(\tau)d\tau}\cdot\left[\sum_{i=1}^{n}\delta_{i-1}\left(F(T_0,T_{i-1},T_i)-K\right)P(T_0,T_i)\right]^+\right\} \\ &= P(t,T_0)\mathbb{E}_{Q_F^{T_0}}\left[\sum_{i=1}^{n}\delta_{i-1}\left(F(T_0,T_{i-1},T_i)-K\right)P(T_0,T_i)\right]^+, \end{aligned}$$

where $F(T_0,T_{i-1},T_i)$, $i=1,\cdots,n$, can be simulated under measure $Q_F^{T_0}$.

Details are as follows. We know that

$$\frac{dF_{i-1}(\tau)}{F_{i-1}(\tau)} = \gamma_{i-1}(\tau)dW^{Q_F^{T_i}}(\tau). \tag{9.85}$$

[29] By Eq. (9.73), and one change of measure,

$$\mathbb{E}_{Q^{\text{swap}}}\left[R_{\text{swap}}(\tau)\right] = \mathbb{E}_{Q^{\text{swap}}}\left[\frac{P(\tau,T_0)-P(\tau,T_n)}{\sum_{i=1}^n\delta_{i-1}P(\tau,T_i)}\right] = \frac{P(t,\tau)\mathbb{E}_{Q_F^{\tau}}\left[P(\tau,T_0)-P(\tau,T_n)\right]}{\sum_{i=1}^n\delta_{i-1}P(t,T_i)} = \frac{P(t,T_0)-P(t,T_n)}{\sum_{i=1}^n\delta_{i-1}P(t,T_i)} = R_{\text{swap}}(t).$$

By results in Appendix 3, we also know that:

$$
\begin{aligned}
dW^{Q_F^{T_i}}(\tau) &= dW^{Q_F^{T_0}}(\tau) - [\sigma_{bi}(\tau) - \sigma_{b0}(\tau)]\, d\tau \\
&= dW^{Q_F^{T_0}}(\tau) + \sum_{j=0}^{i-1} \frac{\delta_j F_j(\tau)}{1 + \delta_j F_j(\tau)} \gamma_j(\tau) d\tau,
\end{aligned}
$$

where the second line follows from Eq. (9.82) in the main text. Replacing this into Eq. (9.85) leaves:

$$
\frac{dF_{i-1}(\tau)}{F_{i-1}(\tau)} = \gamma_{i-1}(\tau) \sum_{j=0}^{i-1} \frac{\delta_j F_j(\tau)}{1 + \delta_j F_j(\tau)} \gamma_j(\tau) d\tau + \gamma_{i-1}(\tau) dW^{Q_F^{T_0}}(\tau), \quad i = 1, \cdots, n.
$$

These can easily be simulated with the methods described in any standard textbook such as Kloeden and Platen (1992).

9.8 Appendix 1: Rederiving the FTAP for bond prices: the diffusion case

Suppose there exist m pure discount bond prices $\{\{P_i \equiv P(\tau, T_i)\}_{i=1}^m\}_{\tau\in[t,T]}$ satisfying:

$$\frac{dP_i}{P_i} = \mu_{bi} \cdot d\tau + \sigma_{bi} \cdot dW, \quad i = 1, \cdots, m, \tag{9.86}$$

where W is a Brownian motion in $\mathbb{R}^d$, and μ_{bi} and σ_{bi} are progressively $\mathcal{F}(\tau)$-measurable functions guaranteeing the existence of a strong solution to the previous system (σ_{bi} is vector-valued). The value process V of a self-financing portfolio in these m bonds and a money market technology satisfies:

$$dV = [\pi^\intercal(\mu_b - \mathbf{1}_m r) + rV]\, d\tau + \pi^\intercal \sigma_b dW,$$

where π is some portfolio, $\mathbf{1}_m$ is a m-dimensional vector of ones, and

$$\begin{aligned} \mu_b &= (\mu_{b1}, \cdots, \mu_{b2})^\intercal \\ \sigma_b &= (\sigma_{b1}, \cdots, \sigma_{b2})^\intercal \end{aligned}$$

Next, suppose that there exists a portfolio $\underline{\pi}$ such that $\underline{\pi}^\intercal \sigma_b = 0$. This is an arbitrage opportunity if there exist events for which at some time, $\mu_b - \mathbf{1}_m r \neq 0$ (use $\underline{\pi}$ when $\mu_b - \mathbf{1}_m r > 0$, and $-\underline{\pi}$ when $\mu_m - \mathbf{1}_d r < 0$: the drift of V will then be appreciating at a deterministic rate that is strictly greater than r). Therefore, arbitrage opportunities are ruled out if:

$$\pi^\intercal(\mu_b - \mathbf{1}_m r) = 0 \text{ whenever } \pi^\intercal \sigma_b = 0.$$

In other terms, arbitrage opportunities are ruled out when every vector in the null space of σ_b is orthogonal to $\mu_b - \mathbf{1}_m r$, or when there exists a λ taking values in $\mathbb{R}^d$ satisfying some basic integrability conditions,[30] and such that

$$\mu_b - \mathbf{1}_m r = \sigma_b \lambda$$

or,

$$\mu_{bi} - r = \sigma_{bi} \lambda, \quad i = 1, \cdots, m. \tag{9.87}$$

In this case,

$$\frac{dP_i}{P_i} = (r + \sigma_{bi}\lambda) \cdot d\tau + \sigma_{bi} \cdot dW, \quad i = 1, \cdots, m.$$

Now define $\tilde{W} = W + \int \lambda d\tau$, $\frac{dQ}{dP} = \exp(-\int_t^T \lambda^\intercal dW - \frac{1}{2}\int_t^T \|\lambda\|^2 d\tau)$. The Q-martingale property of the "normalized" bond price processes now easily follows by Girsanov's theorem. Indeed, define for a generic i, $P(\tau, T) \equiv P(\tau, T_i) \equiv P_i$, and:

$$g(\tau) \equiv e^{-\int_t^\tau r(u)du} \cdot P(\tau, T), \quad \tau \in [t, T].$$

By Girsanov's theorem, and an application of Itô's lemma,

$$\frac{dg}{g} = \sigma_{bi} \cdot d\tilde{W}, \qquad \text{under measure } Q.$$

Therefore,

$$g(\tau) = \mathbb{E}[g(T)], \quad \text{all } \tau \in [t, T].$$

That is:

$$g(\tau) \equiv e^{-\int_t^\tau r(u)du} \cdot P(\tau, T) = \mathbb{E}[g(T)] = \mathbb{E}[e^{-\int_t^T r(u)du} \cdot \underbrace{P(T, T)}_{=1}] = \mathbb{E}\left[e^{-\int_t^T r(u)du}\right],$$

[30] Clearly, λ doesn't depend on the institutional characteristics of the bonds, such as their maturity.

or

$$P(\tau,T) = e^{\int_t^\tau r(u)du} \cdot \mathbb{E}\left[e^{-\int_t^T r(u)du}\right] = \mathbb{E}\left[e^{-\int_\tau^T r(u)du}\right], \quad \text{all } \tau \in [t,T],$$

which is (9.3).

Notice that no assumption has been formulated as regards m. The previous result holds for all integers m, be them less or greater than d. To fix ideas, suppose that there are no other traded assets in the economy. Then, if $m < d$, there exists an infinite number of risk-neutral measures Q. If $m = d$, there exists one and only one risk-neutral measure Q. If $m > d$, there exists one and only one risk-neutral measure but then, the various bond prices have to satisfy some basic no-arbitrage restrictions. As an example, take $m = 2$ and $d = 1$. Relation (9.87) then becomes[31]

$$\frac{\mu_{b1} - r}{\sigma_{b1}} = \lambda = \frac{\mu_{b2} - r}{\sigma_{b2}}.$$

Relation (9.87) will be used several times in this chapter.

- In Section 9.3, it is assumed that the primitive of the economy is the short-term rate, solution of a multidimensional diffusion process, and μ_{bi} and σ_{bi} will be derived via Itô's lemma.
- In Section 9.4, μ_{bi} and σ_{bi} are restricted through a model describing the evolution of the forward rates.

[31] In other terms, the Sharpe ratio of any two bonds must be identical. This reasoning can be extended to the multidimensional case. An interesting empirical issue is to use these ideas to test for the number of factors in the fixed-income market.

9.9 Appendix 2: Certainty equivalent interpretation of forward prices

Multiply both sides of the bond pricing equation (9.3) by the amount $S(T)$:

$$P(t,T) \cdot S(T) = \mathbb{E}\left[e^{-\int_t^T r(\tau)d\tau}\right] \cdot S(T).$$

Suppose momentarily that $S(T)$ is known at T. In this case, we have:

$$P(t,T) \cdot S(T) = \mathbb{E}\left[e^{-\int_t^T r(\tau)d\tau} \cdot S(T)\right].$$

But in the applications we have in mind, $S(T)$ is random. Define then its certainty equivalent by the number $\overline{S(T)}$ that solves:

$$P(t,T) \cdot \overline{S(T)} = \mathbb{E}\left[e^{-\int_t^T r(\tau)d\tau} \cdot S(T)\right],$$

or

$$\overline{S(T)} = \mathbb{E}\left[\eta_T(T) \cdot S(T)\right], \tag{9.88}$$

where $\eta_T(T)$ has been defined in (9.16).

Comparing (9.88) with (9.15) reveals that forward prices can be interpreted in terms of the previously defined certainty equivalent.

9.10 Appendix 3: Additional results on T-forward martingale measures

Eq. (9.16) defines $\eta_T(T)$ as:

$$\eta_T(T) = \frac{e^{-\int_t^T r(\tau)d\tau} \cdot 1}{\mathbb{E}\left[e^{-\int_t^T r(\tau)d\tau}\right]}$$

More generally, we can define a *density process* as:

$$\eta_T(\tau) \equiv \frac{e^{-\int_t^\tau r(u)du} \cdot P(\tau, T)}{\mathbb{E}\left[e^{-\int_t^T r(\tau)d\tau}\right]}, \quad \tau \in [t, T].$$

By the FTAP, $\{\exp(-\int_t^\tau r(u)du) \cdot P(\tau,T)\}_{\tau\in[t,T]}$ is a Q-martingale (see Appendix1 to this chapter). Therefore, $\mathbb{E}[\frac{dQ_F^T}{dQ}\Big| \mathcal{F}_\tau] = \mathbb{E}[\eta_T(T)|\, \mathcal{F}_\tau] = \eta_T(\tau)$ all $\tau \in [t,T]$, and in particular, $\eta_T(t) = 1$. We now show that this works. And at the same time, we show this by deriving a representation of $\eta_T(\tau)$ that can be used to find "forward premia".

We begin with the dynamic representation (9.86) given for a generic bond price # i, $P(\tau,T) \equiv P(\tau,T_i) \equiv P_i$:

$$\frac{dP}{P} = \mu \cdot d\tau + \sigma \cdot dW,$$

where we have defined $\mu \equiv \mu_{bi}$ and $\sigma \equiv \sigma_{bi}$.

Under the risk-neutral measure Q,

$$\frac{dP}{P} = r \cdot d\tau + \sigma \cdot d\tilde{W},$$

where $\tilde{W} = W + \int \lambda$ is a Q-Brownian motion.

By Itô's lemma,

$$\frac{d\eta_T(\tau)}{\eta_T(\tau)} = -\left[-\sigma(\tau,T)\right] \cdot d\tilde{W}(\tau), \quad \eta_T(t) = 1.$$

The solution is:

$$\eta_T(\tau) = \exp\left[-\frac{1}{2}\int_t^\tau \|\sigma(u,T)\|^2 du - \int_t^\tau (-\sigma(u,T)) \cdot d\tilde{W}(u)\right].$$

Under the usual integrability conditions, we can now use the Girsanov's theorem and conclude that

$$W^{Q_F^T}(\tau) \equiv \tilde{W}(\tau) + \int_t^\tau (-\sigma(u,T)^\intercal)\, du \tag{9.89}$$

is a Brownian motion under the T-forward martingale measure Q_F^T.

Finally, note that for all integers i and non decreasing sequences of dates $\{T_i\}_{i=0,1,\cdots,}$,

$$W^{Q_F^{T_i}}(\tau) = \tilde{W}(\tau) + \int_t^\tau (-\sigma(u,T_i)^\intercal)\, du, \quad i = 0, 1, \cdots.$$

Therefore,

$$W^{Q_F^{T_i}}(\tau) = W^{Q_F^{T_{i-1}}}(\tau) - \int_t^\tau \left[\sigma(u,T_i)^\intercal - \sigma(u,T_{i-1})^\intercal\right] du, \quad i = 1, 2, \cdots, \tag{9.90}$$

is a Brownian motion under the T_i-forward martingale measure $Q_F^{T_i}$. Eqs. (9.90) and (9.89) are used in Section 9.7 on interest rate derivatives.

9.11 Appendix 4: Principal components analysis

Principal component analysis transforms the original data into a set of uncorrelated variables, the principal components, with variances arranged in descending order. Consider the following program,

$$\max_{C_1} [var(Y_1)] \text{ s.t. } C_1^\top C_1 = 1,$$

where $var(Y_1) = C_1^\top \Sigma C_1$, and the constraint is an identification constraint. The first order conditions lead to,

$$(\Sigma - \lambda I) C_1 = 0,$$

where λ is a Lagrange multiplier. The previous condition tells us that λ must be one eigenvalue of the matrix Σ, and that C_1 must be the corresponding eigenvector. Moreover, we have $var(Y_1) = C_1^\top \Sigma C_1 = \lambda$ which is clearly maximized by the largest eigenvalue. Suppose that the eigenvalues of Σ are distinct, and let us arrange them in descending order, i.e. $\lambda_1 > \cdots > \lambda_p$. Then,

$$var(Y_1) = \lambda_1.$$

Therefore, the first principal component is $Y_1 = C_1^\top (R - \bar{R})$, where C_1 is the eigenvector corresponding to the largest eigenvalue, λ_1.

Next, consider the second principal component. The program is, now,

$$\max_{C_2} [var(Y_2)] \text{ s.t. } C_2^\top C_2 = 1 \text{ and } C_2^\top C_1 = 0,$$

where $var(Y_2) = C_2^\top \Sigma C_2$. The first constraint, $C_2^\top C_2 = 1$, is the usual identification constraint. The second constraint, $C_2^\top C_1 = 0$, is needed to ensure that Y_1 and Y_2 are orthogonal, i.e. $E(Y_1 Y_2) = 0$. The first order conditions for this problem are,

$$0 = \Sigma C_2 - \lambda C_2 - \nu C_1$$

where λ is the Lagrange multiplier associated with the first constraint, and ν is the Lagrange multiplier associated with the second constraint. By premultiplying the first order conditions by $C_1^\top$,

$$0 = C_1^\top \Sigma C_2 - \nu,$$

where we have used the two constraints $C_1^\top C_2 = 0$ and $C_1^\top C_1 = 1$. But, $C_1^\top \Sigma C_2 = C_1^\top \lambda C_2 = 0$ and, hence, $\nu = 0$. So the first order conditions can be rewritten as,

$$(\Sigma - \lambda I) C_2 = 0.$$

The solution is now λ_2, and C_2 is the eigenvector corresponding to λ_2. (Indeed, this time we can not choose λ_1 as this choice would imply that $Y_2 = C_1^\top (R - \bar{R})$, implying that $E(Y_1 Y_2) \neq 0$.) It follows that $var(Y_2) = \lambda_2$.

In general, we have,

$$var(Y_i) = \lambda_i, \qquad i = 1, \cdots, p.$$

Let Λ be the diagonal matrix with the eigenvalues λ_i on the diagonal. By the spectral decomposition of Σ, $\Sigma = C \Lambda C^\top$, and by the orthonormality of C, $C^\top C = I$, we have that $C^\top \Sigma C = \Lambda$ and, hence,

$$\sum_{i=1}^p var(R_i) = \text{Tr}(\Sigma) = \text{Tr}\left(\Sigma C C^\top\right) = \text{Tr}\left(C^\top \Sigma C\right) = \text{Tr}(\Lambda).$$

Hence, Eq. (9.22) follows.

9.12 Appendix 6: On some analytics of the Hull and White model

As in the Ho and Lee model, the instantaneous forward rate $f(\tau, T)$ predicted by the Hull and White model is as in (9.42), where functions A_2 and B_2 can be easily computed from Eqs. (9.46) and (9.47) as:

$$\begin{aligned} A_2(\tau, T) &= \sigma^2 \int_\tau^T B(s, T) B_2(s, T) ds - \int_\tau^T \theta(s) B_2(s, T) ds \\ B_2(\tau, T) &= e^{-\kappa(T-\tau)} \end{aligned}$$

Therefore, the instantaneous forward rate $f(\tau, T)$ predicted by the Hull and White model is obtained by replacing the previous equations in Eq. (9.42). The result is then equated to the observed forward rate $f_\$(t, \tau)$ so as to obtain:

$$f_\$(t, \tau) = -\frac{\sigma^2}{2\kappa^2}\left[1 - e^{-\kappa(\tau - t)}\right]^2 + \int_t^\tau \theta(s) e^{-\kappa(\tau - s)} ds + e^{-\kappa(\tau - t)} r(t).$$

By differentiating the previous equation with respect to τ, and rearranging terms,

$$\begin{aligned} \theta(\tau) &= \frac{\partial}{\partial \tau} f_\$(t, \tau) + \frac{\sigma^2}{\kappa}\left(1 - e^{-\kappa(\tau - t)}\right) e^{-\kappa(\tau - t)} + \kappa \left[\int_t^\tau \theta(s) e^{-\kappa(\tau - s)} ds + e^{-\kappa(\tau - t)} r(t)\right] \\ &= \frac{\partial}{\partial \tau} f_\$(t, \tau) + \frac{\sigma^2}{\kappa}\left(1 - e^{-\kappa(\tau - t)}\right) e^{-\kappa(\tau - t)} + \kappa \left[f_\$(t, \tau) + \frac{\sigma^2}{2\kappa^2}\left(1 - e^{-\kappa(\tau - t)}\right)^2\right], \end{aligned}$$

which reduces to Eq. (9.48) after using simple algebra.

9.13 Appendix 6: Exercises

A. Expectation theory

Exercise # 1

Suppose that $\sigma(\cdot,\cdot)=\sigma$ and that $\lambda(\cdot)=\lambda$, where σ,λ are constants. Derive the dynamics of r and compare them with f to deduce something substantive about the expectation theory.

Solution. We have:

$$r(\tau)=f(t,\tau)+\int_t^\tau \alpha(s,\tau)ds+\sigma\left(W(\tau)-W(t)\right),$$

where

$$\alpha(\tau,T)=\sigma(\tau,T)\int_\tau^T \sigma(\tau,\ell)d\ell+\sigma(\tau,T)\lambda(\tau)=\sigma^2(T-\tau)+\sigma\lambda.$$

Hence,

$$\int_t^\tau \alpha(s,\tau)ds=\frac{1}{2}\sigma^2\left(\tau-t\right)^2+\sigma\lambda(\tau-t).$$

Finally,

$$r(\tau)=f(t,\tau)+\frac{1}{2}\sigma^2\left(\tau-t\right)^2+\sigma\lambda(\tau-t)+\sigma\left(W(\tau)-W(t)\right),$$

and since $E\left(W(\tau)|\,\mathcal{F}(t)\right)=W(t)$,

$$E\left[r(\tau)|\,\mathcal{F}(t)\right]=f(t,\tau)+\frac{1}{2}\sigma^2\left(\tau-t\right)^2+\sigma\lambda(\tau-t).$$

Even with $\lambda<0$, this model is *not* able to *always* generate $E[r(\tau)|\,\mathcal{F}(t)]<f(t,\tau)$. As shown in the following exercise, this is due to the "nonstationary nature" of the volatility function.

Exercise # 2 [Vasicek]

Suppose instead, that $\sigma(t,T)=\sigma\cdot\exp(-\gamma(T-t))$ and $\lambda(\cdot)=\lambda$, where σ,γ and λ are constants. Redo the same as in exercise 1.

Solution. We have

$$r(\tau)=f(t,\tau)+\int_t^\tau \alpha(s,\tau)ds+\sigma\int_t^\tau e^{-\gamma(\tau-s)}\cdot dW(s),$$

where

$$\alpha(s,\tau)=\sigma^2e^{-\gamma(\tau-s)}\int_s^\tau e^{-\gamma(\ell-s)}d\ell+\sigma\lambda e^{-\gamma(\tau-s)}=\frac{\sigma^2}{\gamma}\left[e^{-\gamma(\tau-s)}-e^{-2\gamma(\tau-s)}\right]+\sigma\lambda e^{-\gamma(\tau-s)}.$$

Finally,

$$E\left[r(\tau)|\,\mathcal{F}(t)\right]=f(t,\tau)+\int_t^\tau \alpha(s,\tau)ds=f(t,\tau)+\frac{\sigma}{\gamma}\left(1-e^{-\gamma(\tau-t)}\right)\left[\frac{\sigma}{2\gamma}\left(1-e^{-\gamma(\tau-t)}\right)+\lambda\right].$$

Therefore, it is sufficient to observe risk-aversion to the extent that

$$-\lambda>\frac{\sigma}{2\gamma}$$

to make

$$E\left[r(\tau)|\,\mathcal{F}(t)\right]<f(t,\tau)\ \text{ for any } \tau.$$

$\lambda<0$ is thus necessary, not sufficient. Notice that when $\lambda=0$, $E\left(r(\tau)|\,\mathcal{F}(t)\right)>f(t,\tau)$, always.

B. Embedding

Exercise # 3 [Ho and Lee ctd.]

Embed the Ho and Lee model in Section 9.5.2 in the HJM format.
Solution. The Ho and Lee model is:

$$dr(\tau) = \theta(\tau)d\tau + \sigma d\tilde{W}(\tau),$$

where $\tilde{W}$ is a Q-Brownian motion. By Eq. (9.??) in Section 9.3,

$$f(r, t, T) = -A_2(t, T) + B_2(t, T)r,$$

where $A_2(t, T) = \int_t^T \theta(s)ds - \frac{1}{2}\sigma^2(T-t)^2$ and $B_2(t, T) = 1$. Therefore, by Eqs. (9.56),

$$\begin{aligned} \sigma(t, T) &= B_2(t, T) \cdot \sigma = \sigma; \\ \alpha(t, T) - \sigma(t, T)\lambda(t) &= -A_{12}(t, T) + B_{12}(t, T)r + B_2(t, T)\theta(t) = \sigma^2(T-t). \end{aligned}$$

Exercise # 4 [Vasicek ctd.]

Embed the Vasicek model in Section 9.3 in the HJM format.
Solution. The Vasicek model is:

$$dr(\tau) = (\theta - \kappa r(\tau))d\tau + \sigma d\tilde{W}(\tau),$$

where $\tilde{W}$ is a Q-Brownian motion. Results from Section 9.3 imply that:

$$f(r, t, T) = -A_2(t, T) + B_2(t, T)r,$$

where $-A_2(t, T) = -\sigma^2 \int_t^T B(s, T)B_2(s, T)ds + \theta \int_t^T B_2(s, T)ds$, $B_2(t, T) = e^{-\kappa(T-t)}$ and $B(t, T) = \frac{1}{\kappa}\left[1 - e^{-\kappa(T-t)}\right]$.
By Eqs. (9.56),

$$\begin{aligned} \sigma(t, T) &= \sigma \cdot B_2(t, T) = \sigma \cdot e^{-\kappa(T-t)}; \\ \alpha(t, T) - \sigma(t, T)\lambda(t) &= -A_{12}(t, T) + B_{12}(t, T)r + (\theta - \kappa r)B_2(t, T) = \frac{\sigma^2}{\kappa}\left[1 - e^{-\kappa(T-t)}\right]e^{-\kappa(T-t)}. \end{aligned}$$

Naturally, this model can never be embedded in a HJM model because it is not of the perfectly fitting type. In practice, condition (9.57) can never hold in the simple Vasicek model. However, the model is embeddable once θ is turned into an infinite dimensional parameter *à la* Hull and White (see Section 9.3).

9.14 Appendix 7: Additional results on string models

Here we prove Eq. (9.59). We have, $\alpha^{I}(\tau,T)=\frac{1}{2}\int_{\tau}^{T}g(\tau,T,\ell_{2})\,d\ell_{2}+cov(\frac{dP}{P},\frac{d\xi}{\xi})$, where

$$g(\tau,T,\ell_{2})\equiv\int_{\tau}^{T}\sigma(\tau,\ell_{1})\,\sigma(\tau,\ell_{2})\,\psi(\ell_{1},\ell_{2})\,d\ell_{1}.$$

Differentiation of the *cov* term is straight forward. Moreover,

$$\begin{aligned}\frac{\partial}{\partial T}\int_{\tau}^{T}g(\tau,T,\ell_{2})\,d\ell_{2} &= g(\tau,T,T)+\int_{\tau}^{T}\frac{\partial g(\tau,T,\ell_{2})}{\partial T}d\ell_{2}\\ &= \sigma(\tau,T)\left[\int_{\tau}^{T}\sigma(\tau,x)\left[\psi(x,T)+\psi(T,x)\right]dx\right]\\ &= 2\sigma(\tau,T)\left[\int_{\tau}^{T}\sigma(\tau,x)\,\psi(x,T)\,dx\right].\end{aligned}$$

9.15 Appendix 8: Change of numeraire techniques

A. Jamshidian (1989)

To handle the expectation in Eq. (9.65), consider the following *change-of-numeraire* result. Let

$$\frac{dA}{A} = \mu_A d\tau + \sigma_A dW,$$

and consider a similar process B with coefficients μ_B and σ_B. We have:

$$\frac{d(A/B)}{A/B} = \left(\mu_A - \mu_B + \sigma_B^2 - \sigma_A\sigma_B\right) d\tau + (\sigma_A - \sigma_B)\, dW. \tag{9.91}$$

Next, let us apply this change-of-numeraire result to the process $y(\tau, S) \equiv \frac{P(\tau,S)}{P(\tau,T)}$ under Q_F^S and under Q_F^T. The goal is to obtain the solution as of time T of $y(\tau, S)$ viz

$$y(T, S) \equiv \frac{P(T,S)}{P(T,T)} = P(T, S) \text{ under } Q_F^S \text{ and under } Q_F^T.$$

This will enable us to evaluate Eq. (9.65).

By Itô's lemma, the PDE (9.29) and the fact that $P_r = -BP$,

$$\frac{dP(\tau, x)}{P(\tau, x)} = r d\tau - \sigma B(\tau, x) d\tilde{W}(\tau), \quad x \geq T.$$

By applying Eq. (9.91) to $y(\tau, S)$,

$$\frac{dy(\tau, S)}{y(\tau, S)} = \sigma^2 \left[B(\tau, T)^2 - B(\tau, T)B(\tau, S)\right] d\tau - \sigma \left[B(\tau, S) - B(\tau, T)\right] d\tilde{W}(\tau). \tag{9.92}$$

All we need to do now is to change measure with the tools of Appendix 3 [see formula (9.76)]: we have that

$$dW^{Q_F^x}(\tau) = d\tilde{W}(\tau) + \sigma B(\tau, x) d\tau$$

is a Brownian motion under the x-forward martingale measure. Replace then $W^{Q_F^x}$ into Eq. (9.92), then integrate, and obtain:

$$\begin{aligned}
\frac{y(T,S)}{y(t,S)} &= P(T,S)\frac{P(t,T)}{P(t,S)} = e^{-\frac{1}{2}\sigma^2 \int_t^T [B(\tau,S)-B(\tau,T)]^2 d\tau + \sigma \int_t^T [B(\tau,S)-B(\tau,T)] dW^{Q_F^T}(\tau)}, \\
\frac{y(T,S)}{y(t,S)} &= P(T,S)\frac{P(t,T)}{P(t,S)} = e^{\frac{1}{2}\sigma^2 \int_t^T [B(\tau,S)-B(\tau,T)]^2 d\tau + \sigma \int_t^T [B(\tau,S)-B(\tau,T)] dW^{Q_F^S}(\tau)},
\end{aligned}$$

Rearranging terms gives Eqs. (9.66) in the main text.

B. Black (1976)

To prove Eq. (9.84) is equivalent to evaluate the following expectation:

$$\mathbb{E}\left[x(T) - K\right]^+,$$

where

$$x(T) = x(t) e^{-\frac{1}{2}\int_t^T \gamma(\tau)^2 d\tau + \int_t^T \gamma(\tau) d\tilde{W}(\tau)}. \tag{9.93}$$

Let 1_{ex} be the indicator of all events s.t. $x(T) \geq K$. We have

$$\begin{aligned}
\mathbb{E}\left[x(T) - K\right]^{+} &= \mathbb{E}\left[x(T) \cdot 1_{ex}\right] - K \cdot \mathbb{E}\left[1_{ex}\right] \\
&= x(t) \cdot \mathbb{E}\left[\frac{x(T)}{x(t)} \cdot 1_{ex}\right] - K \cdot \mathbb{E}\left[1_{ex}\right] \\
&= x(t) \cdot \mathbb{E}_{Q^x}\left[1_{ex}\right] - K \cdot \mathbb{E}\left[1_{ex}\right] \\
&= x(t) \cdot Q^x\left(x(T) \geq K\right) - K \cdot Q\left(x(T) \geq K\right).
\end{aligned}$$

where the probability measure Q^x is defined as:

$$\frac{dQ^x}{dQ} = \frac{x(T)}{x(t)} = e^{-\frac{1}{2}\int_t^T \gamma(\tau)^2 d\tau + \int_t^T \gamma(\tau) d\tilde{W}(\tau)},$$

a Q-martingale starting at one. Under Q^x,

$$dW^x(\tau) = d\tilde{W}(\tau) - \gamma d\tau$$

is a Brownian motion, and x in (9.93) can be written as:

$$x(T) = x(t) e^{\frac{1}{2}\int_t^T \gamma(\tau)^2 d\tau + \int_t^T \gamma(\tau) d\tilde{W}(\tau)}.$$

It is straightforward that $Q\left(x(T) \geq K\right) = \Phi(d_2)$ and $Q^x\left(x(T) \geq K\right) = \Phi(d_1)$, where

$$d_{2/1} = \frac{\ln\left[\frac{x(t)}{K}\right] \mp \frac{1}{2}\int_t^T \gamma(\tau)^2 d\tau}{\sqrt{\int_t^T \gamma(\tau)^2 d\tau}}.$$

Applying this to $\mathbb{E}_{Q_F^{T_i}}\left[F_{i-1}(T_{i-1}) - K\right]^{+}$ gives the formulae of the text.

References

Aït-Sahalia, Y. (1996). "Testing Continuous-Time Models of the Spot Interest Rate." *Review of Financial Studies* 9: 385-426.

Ang, A. and M. Piazzesi (2003). "A No-Arbitrage Vector Autoregression of Term Structure Dynamics with Macroeconomic and Latent Variables." *Journal of Monetary Economics* 50: 745-787.

Balduzzi, P., S. R. Das, S. Foresi and R. K. Sundaram (1996). "A Simple Approach to Three Factor Affine Term Structure Models." *Journal of Fixed Income* 6: 43-53.

Black, F. (1976). "The Pricing of Commodity Contracts." *Journal of Financial Economics* 3: 167-179.

Black, F. and M. Scholes (1973). "The Pricing of Options and Corporate Liabilities." *Journal of Political Economy* 81: 637-659.

Brace, A., D. Gatarek and M. Musiela (1997). "The Market Model of Interest Rate Dynamics." *Mathematical Finance* 7: 127-155.

Brémaud, P. (1981). *Point Processes and Queues*: *Martingale Dynamics*, Berlin: Springer Verlag.

Carverhill, A. (1994). "When is the Short-Rate Markovian?" *Mathematical Finance* 4: 305-312.

Conley, T. G., L. P. Hansen, E. G. J. Luttmer and J. A. Scheinkman (1997). "Short-Term Interest Rates as Subordinated Diffusions." *Review of Financial Studies* 10: 525-577.

Cox, J. C., J. E. Ingersoll and S. A. Ross (1985). "A Theory of the Term Structure of Interest Rates." *Econometrica* 53: 385-407.

Dai, Q. and K.J. Singleton (2000). "Specification Analysis of Affine Term Structure Models." *Journal of Finance* 55: 1943-1978.

Duffie, D. and R. Kan (1996). "A Yield-Factor Model of Interest Rates." *Mathematical Finance* 6: 379-406.

Duffie, D. and K.J. Singleton (1999). "Modeling Term Structures of Defaultable Bonds." *Review of Financial Studies* 12: 687–720.

Fong, H. G. and O. A. Vasicek (1991). "Fixed Income Volatility Management." *The Journal of Portfolio Management* (Summer): 41-46.

Geman, H. (1989). "The Importance of the Forward Neutral Probability in a Stochastic Approach to Interest Rates." Unpublished working paper, ESSEC.

Geman H., N. El Karoui and J.C. Rochet (1995). "Changes of Numeraire, Changes of Probability Measures and Pricing of Options." *Journal of Applied Probability* 32: 443-458.

Goldstein, R.S. (2000). "The Term Structure of Interest Rates as a Random Field." *Review of Financial Studies* 13: 365-384.

Heaney, W. J. and P. L. Cheng (1984). "Continuous Maturity Diversification of Default-Free Bond Portfolios and a Generalization of Efficient Diversification." *Journal of Finance* 39: 1101-1117.

Heath, D., R. Jarrow and A. Morton (1992). "Bond Pricing and the Term-Structure of Interest Rates: a New Methodology for Contingent Claim Valuation." *Econometrica* 60: 77-105.

Ho, T.S.Y. and S.-B. Lee (1986). "Term Structure Movements and the Pricing of Interest Rate Contingent Claims." *Journal of Finance* 41: 1011-1029.

Hull, J.C. (2003). *Options, Futures, and Other Derivatives.* Prentice Hall. 5th edition (International Edition).

Hull, J.C. and A. White (1990). "Pricing Interest Rate Derivative Securities." *Review of Financial Studies* 3: 573-592.

Jagannathan, R. (1984). "Call Options and the Risk of Underlying Securities." *Journal of Financial Economics* 13: 425-434.

Jamshidian, F. (1989). "An Exact Bond Option Pricing Formula." *Journal of Finance* 44: 205-209.

Jamshidian, F. (1997). "Libor and Swap Market Models and Measures." *Finance and Stochastics* 1: 293-330.

Jöreskog, K.G. (1967). "Some Contributions to Maximum Likelihood Factor Analysis." *Psychometrica* 32: 443-482.

Karlin, S. and H. M. Taylor (1981). *A Second Course in Stochastic Processes.* San Diego: Academic Press.

Kennedy, D.P. (1994). "The Term Structure of Interest Rates as a Gaussian Random Field." *Mathematical Finance* 4: 247-258.

Kennedy, D.P. (1997). "Characterizing Gaussian Models of the Term Structure of Interest Rates." *Mathematical Finance* 7: 107–118.

Kloeden, P. and E. Platen (1992). *Numeric Solutions of Stochastic Differential Equations.* Berlin: Springer Verlag.

Knez, P.J., R. Litterman and J. Scheinkman (1994). "Explorations into Factors Explaining Money Market Returns." *Journal of Finance* 49: 1861-1882.

Langetieg, T. (1980). "A Multivariate Model of the Term Structure of Interest Rates." *Journal of Finance* 35: 71-97.

Litterman, R. and J. Scheinkman (1991). "Common Factors Affecting Bond Returns." *Journal of Fixed Income* 1: 54-61.

Litterman, R., J. Scheinkman, and L. Weiss (1991). "Volatility and the Yield Curve." *Journal of Fixed Income* 1: 49-53.

Longstaff, F.A. and E.S. Schwartz (1992). "Interest Rate Volatility and the Term Structure: A Two-Factor General Equilibrium Model." *Journal of Finance* 47: 1259-1282.

Mele, A. (2003). "Fundamental Properties of Bond Prices in Models of the Short-Term Rate." *Review of Financial Studies* 16: 679-716.

Mele, A. and F. Fornari (2000). *Stochastic Volatility in Financial Markets: Crossing the Bridge to Continuous Time*. Boston: Kluwer Academic Publishers.

Merton, R.C. (1973). "Theory of Rational Option Pricing." *Bell Journal of Economics and Management Science* 4: 141-183.

Miltersen, K., K. Sandmann and D. Sondermann (1997). "Closed Form Solutions for Term Structure Derivatives with Lognormal Interest Rate." *Journal of Finance* 52: 409-430.

Rebonato, R. (1998). *Interest Rate Option Models*. 2nd edition. Wiley.

Rebonato, R. (1999). *Volatility and Correlation*. Wiley.

Ritchken, P. and L. Sankarasubramanian (1995). "Volatility Structure of Forward Rates and the Dynamics of the Term Structure." *Mathematical Finance* 5: 55-72.

Rothschild, M. and J.E. Stiglitz (1970). "Increasing Risk: I. A Definition." *Journal of Economic Theory* 2: 225-243.

Sandmann, K. and D. Sondermann (1997). "A Note on the Stability of Lognormal Interest Rate Models and the Pricing of Eurodollar Futures." *Mathematical Finance* 7: 119-125.

Santa-Clara, P. and D. Sornette (2001). "The Dynamics of the Forward Interest Rate Curve with Stochastic String Shocks." *Review of Financial Studies* 14: 149-185.

Stanton, R. (1997). "A Nonparametric Model of Term Structure Dynamics and the Market Price of Interest Rate Risk." *Journal of Finance* 52: 1973-2002.

Vasicek, O. (1977). "An Equilibrium Characterization of the Term Structure." *Journal of Financial Economics* 5: 177-188.

Part IV

Taking models to data

10
Statistical inference for dynamic asset pricing models

10.1 Introduction

The next parts of the lectures develop applications and models emanating from the theory. This chapter surveys econometric methods for estimating and testing these models. We start from the very foundational issues on identification, specification and testing; and present the details of classical estimation and testing methodologies such as the Method of Moments, in which the number of moment conditions equals the dimension of the parameter vector (Pearson (1894)); Maximum Likelihood (ML) (Fisher (1912); see also Gauss (1816)); the Generalized Method of Moments (GMM), in which the number of moment conditions exceeds the dimension of the parameter vector, and thus leads to minimum chi-squared (Hansen (1982); see also Neyman and Pearson (1928)); and finally the recent developments based on simulations, which aim at implementing ML and GMM estimation for models that are analytically very complex - but that can be simulated. The present version of the chapter emphasizes the asymptotic theory (that is what happens when the sample size is large). I plan to include more applications in subsequent versions of this chapter as well as in the chapters that follow.

10.2 Stochastic processes and econometric representation

10.2.1 Generalities

Heuristically, a n-dimensional stochastic process $y = \{y_t\}_{t\in\mathcal{T}}$ is a sequence of random variables indexed by time; that is, y is defined through the probability space $(\Omega, \mathcal{F}, P)$, where Ω is the set of states, and $\mathcal{F}$ is a tribe, or a σ-algebra on Ω, i.e. a set of parts of Ω.[1] If $x(\cdot,\cdot) : \Omega \times \mathcal{T} \mapsto \mathbb{R}^n$, we say that the sequence of random variables $x(\cdot, t)_{t\in\mathcal{T}}$ is the stochastic process. For a given $\omega \in \Omega$, $x(\omega, \cdot)$ is a *realization* of the process. (For example, Ω can be the Cartesian product $\mathbb{R}^n \times \mathbb{R}^n \times \cdots$.)

Clearly, the econometrician only observes data, which he assumes to be generated by a given random process, called the *data generating process* (DGP). We assume this process can be

[1] We say that if $\Omega = \mathbb{R}^d$, the algebra generated by the open sets of $\mathbb{R}^d$ is the Borel tribe on $\mathbb{R}^d$ - denoted as $\mathcal{B}(\mathbb{R}^d)$.

represented in terms of a probability distribution. And since the probability distribution is unknown, the aim of the econometrician is to use all of available data (or observations) to get useful insights into the nature of this process. The task of econometric methods is to address this information search process.

Given the previous assumptions on the DGP, we can say that the DGP is characterized by a conditional law - let's say the law of y_t given the set of past values $\underline{y_{t-1}} = \{y_{t-1}, y_{t-2}, \cdots\}$, and some additional strongly exogenous variable z, with $\underline{z_t} = \{z_t, z_{t-1}, z_{t-2}, \cdots\}$. Thus the DGP is a description of data based on their conditional density (the true law), say

$$\text{DGP} : \ell_0(y_t| \, x_t),$$

where $x_t = (\underline{y_{t-1}}, \ \underline{z_t})$.

We are now ready to introduce three useful definitions. First, we say that a *parametric model* is a set of conditional laws for y_t indexed by a parameter vector $\theta \in \Theta \subseteq \mathbb{R}^p$, viz

$$(M) = \{\ell\,(y_t| \, x_t; \theta)\,,\ \theta \in \Theta \subseteq \mathbb{R}^p\}\,.$$

Second, we say that the model (M) is *well-specified* if,

$$\exists \theta_0 \in \Theta : \ell\,(y_t| \, x_t; \theta_0) = \ell_0\,(y_t| \, x_t)\,.$$

Third, we say that the model (M) is *identifiable* if θ_0 is unique.

The main concern in this appendix is to provide details on *parameter estimation* of the true parameter θ_0.

10.2.2 Mathematical restrictions on the DGP

The previous definition of the DGP is very rich. In practice, it is extremely difficult to cope with DGPs defined at such a level of generality. In this appendix, we only describe estimation methods applying to DGPs satisfying some restrictions. There are two fundamental restrictions usually imposed on the DGP.

- Restrictions related to the *heterogeneity* of the stochastic process - which pave the way to the concept of *stationarity.*
- Restrictions related to the *memory* of the stochastic process - which pave the way to the concept of *ergodicity.*

10.2.2.1 Stationarity

Stationary processes can be thought of as describing phenomena that are able to approach a well defined long run equilibrium in some statistical sense. "Statistical sense" means that these phenomena are subject to random fluctuations; yet as time unfolds, the probability generating them settles down. In other terms, in the "long-run", the probability distribution generating the random fluctuations doesn't change. Some economists even define a long-run equilibrium as a will defined "invariant" (or stationary) probability distribution.

We have two notions of stationarity.

- Strong (or strict) stationarity. Definition: Homogeneity in law.
- Weak stationarity, or stationarity of order p. Definition: Homogeneity in moments.

Even if the DGP is stationary, the number of parameters to estimate increases with the sample size. Intuitively, the dimension of the parameter space is high because of auto-covariance effects. For example, consider two stochastic processes; one, for which $cov(y_t, y_{t+\tau}) = \tau^2$; and another, for which $cov(y_t, y_{t+\tau}) = \exp(-|\tau|)$. In both cases, the DGP is stationary. Yet in the first case, the dependence increases with τ; and in the second case, the dependence decreases with τ. This is an issue related to the memory of the process. As is clear, a stationary stochastic process may have "long memory". "Ergodicity" further restricts DGPs so as to make this memory play a somewhat more limited role.

10.2.2.2 Ergodicity

We shall consider situations in which the dependence between y_{t_1} and y_{t_2} decreases with $|t_2 - t_1|$. Let's introduce some concepts and notation. Two events A and B are independent when $P(A \cap B) = P(A)P(B)$. A stochastic process is *asymptotically independent* if, given

$$\beta(\tau) \geq \left| F(y_{t_1}, \cdots, y_{t_n}, y_{t_1+\tau}, \cdots, y_{t_n+\tau}) - F(y_{t_1}, \cdots, y_{t_n}) F(y_{t_1+\tau}, \cdots, y_{t_n+\tau}) \right|,$$

one also has that $\lim_{\tau \to \infty} \beta(\tau) \to 0$. A stochastic process is *p-dependent* if $\beta(\tau) \neq 0 \ \forall p < \tau$. A stochastic process is *asymptotically uncorrelated* if there exists $\rho(\tau)$ ($\tau \geq 1$) such that for all t, $\rho(\tau) \geq \left| \frac{cov(y_t - y_{t+\tau})}{var(y_t) \cdot var(y_{t+\tau})} \right|$, and that $0 \leq \rho(\tau) \leq 1$ with $\sum_{\tau=0}^{\infty} \rho(\tau) < \infty$. For example, $\rho(\tau) = \tau^{-(1+\delta)}$ ($\delta > 0$), in which case $\rho(\tau) \downarrow 0$ as $\tau \uparrow \infty$.

Let $\mathbb{B}_1^t$ denote the σ-algebra generated by $\{y_1, \cdots, y_t\}$ and $A \in \mathbb{B}_{-\infty}^t$, $B \in \mathbb{B}_{t+\tau}^{\infty}$, and define

$$\begin{aligned} \alpha(\tau) &= \sup\nolimits_\tau \left| P(A \cap B) - P(A)P(B) \right| \\ \varphi(\tau) &= \sup\nolimits_\tau \left| P(B \mid A) - P(B) \right|, \ P(A) > 0 \end{aligned}$$

We say that 1) y is *strongly mixing*, or *α-mixing* if $\lim_{\tau \to \infty} \alpha(\tau) \to 0$; 2) y is *uniformly mixing* if $\lim_{\tau \to \infty} \varphi(\tau) \to 0$. Clearly, a uniformly mixing process is also strongly mixing.

A second order stationary process is *ergodic* if $\lim_{T \to \infty} \sum_{\tau=1}^{T} cov(y_t, y_{t+\tau}) = 0$. If a second order stationary process is strongly mixing, it is also ergodic.

10.2.3 Parameter estimators: basic definitions

Consider an estimator of the parameter vector θ of the model,

$$(M) = \{\ell(y_t | x_t; \theta), \ \theta \in \Theta \subset \mathbb{R}^p\}.$$

Naturally, any estimator must necessarily be some function(al) of the observations $\hat{\theta}_T \equiv t_T(y)$. Of a given estimator $\hat{\theta}_T$, we say that it is

- *Correct* (or *unbiased*), if $E(\hat{\theta}_T) = \theta_0$. The difference $E(\hat{\theta}_T) - \theta_0$ is the distortion, or bias.

- *Weakly consistent* if $p\lim \hat{\theta}_T = \theta_0$. And *strongly consistent* if $\hat{\theta}_T \overset{a.s.}{\to} \theta_0$.

Finally, an estimator $\hat{\theta}_T^{(1)}$ is more *efficient* than another estimator $\hat{\theta}_T^{(2)}$ if, for any vector of constants c, we have that $c^\top \cdot var(\hat{\theta}_T^{(1)}) \cdot c < c^\top \cdot var(\hat{\theta}_T^{(2)}) \cdot c$.

10.2.4 Basic properties of density functions

We have T observations $y_T^1 = \{y_1, \cdots, y_T\}$. Suppose these observations are the realization of a T-dimensional random variable with joint density,

$$f(\tilde{y}_1, \cdots, \tilde{y}_T; \theta) = f\left(\tilde{y}_1^T; \theta\right).$$

We have put the tilde on the y_is to emphasize that we view them as random variables.[2] But to ease notation, from now on we write y_i instead of $\tilde{y}_i$. By construction, $\int f(y|\,\theta)\,dy \equiv \int \cdots \int f\left(y_1^T\middle|\,\theta\right) dy_1^T = 1$ or,[3]

$$\forall \theta \in \Theta, \quad \int f(y; \theta)\,dy = 1.$$

Now suppose that the support of y doesn't depend on θ. Under standard regularity conditions,

$$\nabla_\theta \int f(y;\theta)\,dy = \int \nabla_\theta f(y;\theta)\,dy = \mathbf{0}_p,$$

where $\mathbf{0}_p$ is a column vector of zeros in $\mathbb{R}^p$. Moreover, for all $\theta \in \Theta$,

$$\mathbf{0}_p = \int \nabla_\theta f(y;\theta)\,dy = \int [\nabla_\theta \log f(y;\theta)]\, f(y;\theta)\,dy = E_\theta[\nabla_\theta \log f(y;\theta)].$$

Finally we have,

$$\begin{aligned}\mathbf{0}_{p\times p} &= \nabla_\theta \int [\nabla_\theta \log f(y;\theta)]\, f(y;\theta)\,dy \\ &= \int [\nabla_{\theta\theta} \log f(y;\theta)]\, f(y;\theta)\,dy + \int |\nabla_\theta \log f(y;\theta)|_2\, f(y;\theta)\,dy,\end{aligned}$$

where $|x|_2$ denotes outer product, i.e. $|x|_2 = x \cdot x^\top$. But since $E_\theta[\nabla_\theta \log f(y;\theta)] = \mathbf{0}_p$,

$$E_\theta[\nabla_{\theta\theta} \log f(y;\theta)] = -E_\theta |\nabla_\theta \log f(y;\theta)|_2 = -var_\theta[\nabla_\theta \log f(y;\theta)] \equiv -\mathcal{J}(\theta),\ \forall\theta \in \Theta.$$

The matrix $\mathcal{J}$ is known as the Fisher's information matrix.

10.2.5 The Cramer-Rao lower bound

Let $t(y)$ some unbiased estimator of θ. Let $p = 1$. We have,

$$E[t(y)] = \int t(y) f(y;\theta)\,dy.$$

Under regularity conditions,[4]

$$\nabla_\theta E[t(y)] = \int t(y) \nabla_\theta f(y;\theta)\,dy = \int t(y)\,[\nabla_\theta \log f(y;\theta)]\, f(y;\theta)\,dy = cov(t(y), \nabla_\theta \log f(y;\theta)).$$

---

[2] We thus follow a classical perspective. A Bayesian statistician would view the sample as *given*. We do not consider Bayesian methods in this appendix.

[3] In this subsection, all statement will be referring to all possible parameters acting as the "true" parameters, i.e. $\forall\theta \in \Theta$. It also turns out that in correspondence of the ML estimator $\hat{\theta}_T$ we also have $0 = \frac{\partial}{\partial\theta} \log L(\hat{\theta}_T \mid x)$, but this has nothing to do with the results of the present section.

[4] The last equality is correct because $E(xy) = cov(x, y)$ if $E(y) = 0$.

By the basic inequality, $cov\,(x,y)^2 \leq var\,(x)\,var\,(y)$,

$$[cov\,(t(y), \nabla_\theta \log f\,(y;\theta))]^2 \leq var\,[t(y)] \cdot var\,[\nabla_\theta \log f\,(y;\theta)]\,.$$

Therefore,

$$[\nabla_\theta E\,(t(y))]^2 \leq var\,[t(y)] \cdot var\,[\nabla_\theta \log f\,(y;\theta)] = -var\,[t(y)] \cdot E\,[\nabla_{\theta\theta} \log f\,(y;\theta)]\,.$$

But if $t(y)$ is unbiased, or $E\,[t(y)] = \theta$,

$$var\,[t(y)] \geq [-E\,(\nabla_\theta \log f\,(y;\theta))]^{-1} \equiv \mathcal{J}\,(\theta)^{-1}\,.$$

This is the celebrated Cramer-Rao bound. The same results holds with a change in notation in the multidimensional case. (For a proof, see, for example, Amemiya (1985, p. 14-17).)

10.3 The likelihood function

10.3.1 Basic motivation and definitions

The density $f(y_1^T \,|\, \theta)$ is a function such that $\mathbb{R}^{nT} \times \Theta \mapsto \mathbb{R}_+$: it maps every possible sample and values of θ on to positive numbers. Clearly, we may trace the joint density of the entire sample through the thought experiment in which we modify the sample y_1^T. We ask, "Which value of θ makes the sample we observed the most likely to occur?" We introduce the "likelihood function"

$$L(\theta |\, y_1^T) \equiv f(y_1^T;\theta).$$

It is the function $\theta \mapsto f(y;\theta)$ for y_1^T given ($\bar{y}$, say):

$$L(\theta |\, \bar{y}) \equiv f(\bar{y};\theta).$$

We maximize $L(\theta |\, y_1^T)$ with respect to θ. Once again, we are looking for the value of θ which maximizes the probability to observe what we effectively observed. We will see that if the model is not misspecified, the Cramer-Rao lower bound is attained by the ML estimator. To grasp the intuition of this result in the i.i.d. case, note that

$$\frac{1}{T}\sum_{t=1}^{T} \nabla_{\theta\theta}\ell_t(\theta_0) \xrightarrow{p} E_{\theta_0}\left[\nabla_{\theta\theta}\ell_t(\theta_0)\right],$$

where

$$\ell_t(\theta) \equiv \ell(\theta |\, y_t) = \log f(y_t;\theta),$$

a function of y_t only, so a standard law of large numbers would suffice to conclude. In the dependent case, $\ell_t(\theta) \equiv \ell(\theta |\, y_t |\, x_t) = \log f(\, y_t |\, x_t |\, \theta)$.[5]

[5]This is not a *fixed* function. Rather, it varies with t. In this case, we need additional mathematical tools reviewed in section A.6.3.

10.3.2 Preliminary results on probability factorizations

We have,

$$P\left(\bigcap_{i=1}^{n} A_i\right) = \prod_{i=1}^{n} P\left(A_i \left| \bigcap_{j=1}^{i-1} A_j \right.\right).$$

Indeed, $P(A_1 \bigcap A_2) = P(A_1) \cdot P(A_2 | A_1)$. Let $E \equiv A_1 \bigcap A_2$. We still have,

$$P(A_3 | A_1 \bigcap A_2) = P(A_3 | E) = \frac{P(A_3 \bigcap E)}{P(E)} = \frac{P(A_3 \bigcap A_1 \bigcap A_2)}{P(A_1 \bigcap A_2)}.$$

That is,

$$P\left(\bigcap_{i=1}^{3} A_i\right) = P(A_1 \bigcap A_2) \cdot P(A_3 | A_1 \bigcap A_2) = P(A_1) \cdot P(A_2 | A_1) \cdot P(A_3 | A_1 \bigcap A_2),$$

etc. ‖

10.3.3 Asymptotic properties of the MLE

10.3.3.1 Limiting problem

By definition, the MLE satisfies,

$$\hat{\theta}_T = \arg\max_{\theta \in \Theta} L_T(\theta) = \arg\max_{\theta \in \Theta} (\log L_T(\theta)),$$

where

$$\log L_T(\theta) \equiv \log \prod_{t=1}^{T} f\left(y_t \left| y_1^{t-1} \right.; \theta\right) = \sum_{t=1}^{T} \log f\left(y_t \left| y_1^{t-1} \right.; \theta\right) \equiv \sum_{t=1}^{T} \log f(y_t; \theta) \equiv \sum_{t=1}^{T} \ell_t(\theta),$$

and $\ell_t(\theta)$ is the "log-likelihood" of a single observation: $\ell_t(\theta) \equiv \ell(\theta | y_t) \equiv \log f(y_t | \theta)$. Clearly, we also have that:

$$\hat{\theta}_T = \arg\max_{\theta \in \Theta} (\log L_T(\theta)) = \arg\max_{\theta \in \Theta} \left(\frac{1}{T} \log L_T(\theta)\right),$$

where $\frac{1}{T} \log L_T(\theta) \equiv \frac{1}{T} \sum_{t=1}^{T} \ell_t(\theta)$. The MLE satisfies the following first order conditions,

$$\mathbf{0}_p = \nabla_\theta \log L_T(\theta)|_{\theta = \hat{\theta}_T}.$$

Consider a Taylor expansion of the first order conditions around θ_0,

$$\mathbf{0}_p = \nabla_\theta \log L_T(\theta)|_{\theta = \hat{\theta}_T} \equiv \nabla_\theta \log L_T(\hat{\theta}_T) \overset{d}{=} \nabla_\theta \log L_T(\theta_0) + \nabla_{\theta\theta} \log L_T(\theta_0)(\hat{\theta}_T - \theta_0),$$

where the notation $x_T \overset{d}{=} y_T$ means that the difference $x_T - y_T = o_p(1)$. Here θ_0 is solution to the limiting problem,

$$\theta_0 = \arg\max_{\theta \in \Theta} \left[\lim_{T \to \infty} \left(\frac{1}{T} \log L_T(\theta)\right)\right] = \arg\max_{\theta \in \Theta} [E(\ell(\theta))],$$

where L satisfies all of the regularity conditions needed to ensure that,

$$\theta_0 : E[\nabla_\theta \ell(\theta_0)] = \mathbf{0}_p.$$

To show that this is indeed the solution, suppose θ_0 is identified; that is, $\theta \neq \theta_0$ and $\theta, \theta_0 \in \Theta \Leftarrow f(y|\theta) \neq f(y|\theta_0)$. Let $E_\theta[\log f(y|\theta)] < \infty$, $\forall \theta \in \Theta$. We have, $\theta_0 = \arg\max_{\theta \in \Theta} E_{\theta_0}[\log f(y|\theta)]$ and is a singleton. The proof is very simple. We have, $Q(\theta_0) - Q(\theta) = E_{\theta_0}[-\log\left(\frac{f(y|\theta)}{f(y|\theta_0)}\right)] > -\log E_{\theta_0}\left(\frac{f(y|\theta)}{f(y|\theta_0)}\right) = -\log \int \frac{f(y|\theta)}{f(y|\theta_0)} f(y|\theta_0) dy = -\log \int f(y|\theta) dy = 0.$

10.3.3.2 Consistency

Under regularity conditions, $\hat{\theta}_T \xrightarrow{p} \theta_0$ and even $\hat{\theta}_T \xrightarrow{a.s.} \theta_0$ if the model is well-specified. One example of conditions required to obtain weak consistency is that the following UWLLN ("Uniform weak law of large numbers") holds,

$$\lim_{T\to\infty} P\left[\sup_{\theta\in\Theta} \left|\ell_T(\theta) - E(\ell(\theta))\right|\right] \to 0.$$

Asymptotic normality: sketch

Consider again the previous asymptotic expansion:

$$\mathbf{0}_p \stackrel{d}{=} \nabla_\theta \log L_T(\theta_0) + \nabla_{\theta\theta} \log L_T(\theta_0)(\hat{\theta}_T - \theta_0).$$

We have,

$$\begin{aligned}\sqrt{T}(\hat{\theta}_T - \theta_0) &\stackrel{d}{=} -\sqrt{T}\left[\frac{T}{T}\nabla_{\theta\theta}\log L_T(\theta_0)\right]^{-1} \nabla_\theta \log L_T(\theta_0)\\
&= -\left[\frac{1}{T}\nabla_{\theta\theta}\log L_T(\theta_0)\right]^{-1}\frac{1}{\sqrt{T}}\nabla_\theta \log L_T(\theta_0)\\
&= -\left[\frac{1}{T}\sum_{t=1}^T \nabla_{\theta\theta}\ell_t(\theta_0)\right]^{-1}\frac{1}{\sqrt{T}}\sum_{t=1}^T \nabla_\theta \ell_t(\theta_0).\end{aligned}$$

From now on, we consider the i.i.d. case only. By the law of large numbers (weak law no. 1),

$$\frac{1}{T}\sum_{t=1}^T \nabla_{\theta\theta}\ell_t(\theta_0) \xrightarrow{p} E_{\theta_0}\left[\nabla_{\theta\theta}\ell_t(\theta_0)\right] = -\mathcal{J}(\theta_0).$$

Therefore, asymptotically,

$$\sqrt{T}(\hat{\theta}_T - \theta_0) \stackrel{d}{=} \mathcal{J}(\theta_0)^{-1}\frac{1}{\sqrt{T}}\sum_{t=1}^T \nabla_\theta \ell_t(\theta_0).$$

We also have,

$$\frac{1}{\sqrt{T}}\sum_{t=1}^T \nabla_\theta \ell_t(\theta_0) \xrightarrow{d} N\left(0, \mathcal{J}(\theta_0)\right).$$

Indeed,

- The convergence to a normal follows by the central limit theorem: $\frac{\sqrt{T}(\bar{y}_T-\mu)}{\sigma} \xrightarrow{d} N(0,1)$, $\bar{y}_T \equiv \frac{1}{T}\sum_{t=1}^T y_t$. Indeed, here we have that: $\frac{\sqrt{T}(\bar{y}_T-\mu)}{\sigma} = \sqrt{T}\frac{\frac{1}{T}\sum_{t=1}^T y_t - \mu}{\sigma} = \sqrt{T}\frac{\frac{1}{T}\sum_{t=1}^T (y_t-\mu)}{\sigma} = \frac{1}{\sqrt{T}}\frac{\sum_{t=1}^T (y_t-\mu)}{\sigma}$, $y_t = \nabla_\theta \ell_t(\theta_0)$.
- Moreover, $\forall t$, $E\left[\nabla_\theta \ell_t(\theta_0)\right] = 0$ by the first order conditions of the limiting problem.
- Finally, $\forall t$, $var\left[\nabla_\theta \ell_t(\theta_0)\right] = \mathcal{J}(\theta_0)$, which was shown to hold in subsection A.2.1.

By the Slutzky's theorem in section ???,

$$\sqrt{T}(\hat{\theta}_T - \theta_0) \xrightarrow{d} N\left(0, \mathcal{J}(\theta_0)^{-1}\right).$$

Hence, the ML estimator attains the Cramer-Rao bound found in subsection A.1.9.

10.3.3.3 Theory

We take for granted the convergence of $\hat{\theta}_T$ to θ_0 and an additional couple of mild regularity conditions. Suppose that

$$\hat{\theta}_T \overset{a.s.}{\to} \theta_0,$$

and that $H(y,\theta) \equiv \nabla_{\theta\theta} \log L(\theta|\, y)$ exists, is continuous in θ uniformly in y and that we can differentiate twice under the integral operator $\int L(\theta|\, y)dy = 1$. We have

$$s_T(\omega,\theta) = \frac{1}{T}\sum_{t=1}^{T} \nabla_\theta \log L(\theta|\, y).$$

Consider the c-parametrized curves $\theta(c) = c\Box(\theta_0 - \hat{\theta}_T) + \hat{\theta}_T$ where, for all $c \in (0,1)^p$ and $\theta \in \Theta$, $c\Box\theta$ denotes a vector in Θ where the ith element is $c^{(i)}\theta^{(i)}$. By the intermediate value theorem, there exists then a c^* in $(0,1)^p$ such that for all $\omega \in \Omega$ we have:

$$s_T(\omega, \hat{\theta}_T(\omega)) = s_T(\omega,\theta_0) + H_T(\omega,\theta^*) \cdot (\hat{\theta}_T(\omega) - \theta_0),$$

where $\theta^* \equiv \theta(c^*)$ and:

$$H_T(\omega,\theta) = \frac{1}{T}\sum_{t=1}^{T} H(\theta|\, y_t).$$

The first order conditions tell us that:

$$s_T(\omega, \hat{\theta}_T(\omega)) = 0.$$

Hence,

$$0 = s_T(\omega,\theta_0) + H_T(\omega,\theta_T^*(\omega)) \cdot (\hat{\theta}_T(\omega) - \theta_0).$$

On the other hand,

$$|H_T(\omega,\theta_T^*(\omega)) - H_T(\omega,\theta_0)| \le \frac{1}{T}\sum_{t=1}^{T} |H(\omega,\theta_T^*(\omega)) - H_T(\omega,\theta_0)| \le \sup_y |H(\omega,\theta_T^*(\omega)) - H_T(\omega,\theta_0)|\,.$$

Since $\hat{\theta}_T \overset{a.s.}{\to} \theta_0$, we also have that $\theta_T^* \overset{a.s.}{\to} \theta_0$. Since H is continuous in θ uniformly in y, the previous inequality implies that,

$$H_T(\omega,\theta_T^*(\omega)) \overset{a.s.}{\to} -\mathcal{J}(\theta_0)\,.$$

This is so because by the law of large numbers,

$$H_T(\omega,\theta_0) = \frac{1}{T}\sum_{t=1}^{T} H(\theta_0|\, y_t) \overset{p}{\to} E\left[H(\theta_0|\, y_t)\right] = -\mathcal{J}(\theta_0)\,.$$

Therefore, as $T \to \infty$,

$$\sqrt{T}\left(\hat{\theta}_T(\omega) - \theta_0\right) = -H_T^{-1}(\omega,\theta_0) \cdot s_T(\omega,\theta_0)\sqrt{T} = \mathcal{J}^{-1} \cdot \sqrt{T} s_T(\omega,\theta_0).$$

By the central limit theorem $s_T(\omega,\theta_0) = \frac{1}{T}\sum_{t=1}^{T} s(\theta_0, y_t)$ is such that

$$\sqrt{T} \cdot s_T(\omega,\theta_0) \overset{d}{\to} N(0, var(s(\theta_0,y_t)))\,,$$

where

$$var(s(\theta_0,y_t)) = \mathcal{J}.$$

since $E(s_T) = 0$. The result follows by the Slutzky's theorem and the symmetry of $\mathcal{J}$.

Finally we should show the existence of a sequence $\hat{\theta}_T$ converging a.s. to θ_0. Proofs on such a convergence can be found in Amemiya (1985), or in Newey and McFadden (1994).

10.4 M-estimators

Consider a function g of the (unknown) parameters θ. A *M-estimator* of the function $g(\theta)$ is the solution to,

$$\max_{g\in G}\sum_{t=1}^{T}\Psi\left(x_t,y_t;g\right),$$

where Ψ is a given real function, y is the endogenous variable, and x is as in section A.1. We assume that a solution exists, that it is interior and that it is unique. We denote the M-estimator with $\hat{g}_T\left(x_1^T,y_1^T\right)$. Naturally, the M-estimator satisfies the following first order conditions,

$$0=\frac{1}{T}\sum_{t=1}^{T}\nabla_g\Psi\left(y_t,x_t;\hat{g}_T\left(x_1^T,y_1^T\right)\right).$$

To simplify the presentation, we assume that (x,y) are independent and that they have the same law. By the law of large numbers,

$$\frac{1}{T}\sum_{t=1}^{T}\Psi\left(y_t,x_t;g\right)\overset{p}{\to}\iint\Psi\left(y,x;g\right)dF(x,y)=\iint\Psi\left(y,x;g\right)dF\left(y|\,x\right)dZ\left(x\right)\equiv E_xE_0\left[\Psi\left(y,x;g\right)\right]$$

where E_0 is the expectation operator taken with respect to the true conditional law of y given x and E_x is the expectation operator taken with respect to the true marginal law of x. The limit problem is,

$$\max_{g\in G}E_xE_0\left[\Psi\left(y,x;g\right)\right].$$

Under standard regularity conditions,[6] there exists a sequence of M-estimators $\hat{g}_T(x,y)$ converging a.s. to $g_\infty=g_\infty\left(\theta_0\right)$. Under some additional regularity conditions, the M-estimator is asymptotic normal:

THEOREM: *Let* $\mathcal{I}\equiv E_xE_0\left(\nabla_g\Psi\left(y,x;g_\infty\left(\theta_0\right)\right)\left[\nabla_g\Psi\left(y,x;g_\infty\left(\theta_0\right)\right)\right]^\top\right)$ *and assume that the matrix* $\mathcal{J}\equiv E_xE_0\left[-\nabla_{gg}\Psi\left(y,x;g\right)\right]$ *exists and has an inverse. We have,*

$$\sqrt{T}\left(\hat{g}_T-g_\infty\left(\theta_0\right)\right)\overset{d}{\to}N\left(0,\mathcal{J}^{-1}\mathcal{I}\mathcal{J}^{-1}\right).$$

PROOF (sketchy). The M-estimator satisfies the following first order conditions,

$$\begin{aligned}0&=\frac{1}{\sqrt{T}}\sum_{t=1}^{T}\nabla_g\Psi\left(y_t,x_t;\hat{g}_T\right)\\&\overset{d}{=}\frac{1}{\sqrt{T}}\sum_{t=1}^{T}\nabla_g\Psi\left(y_t,x_t;g_\infty\right)+\sqrt{T}\left[\frac{1}{T}\sum_{t=1}^{T}\nabla_{gg}\Psi\left(y_t,x_t;g_\infty\right)\right]\cdot\left(\hat{g}_T-g_\infty\right).\end{aligned}$$

[6] G is compact; Ψ is continuous with respect to g and integrable with respect to the true law, for each g; $\frac{1}{T}\sum_{t=1}^{T}\Psi\left(y_t,x_t;g\right)\overset{a.s.}{\to}E_xE_0\left[\Psi\left(y,x;g\right)\right]$ uniformly on G; the limit problem has a unique solution $g_\infty=g_\infty\left(\theta_0\right)$.

By rearranging terms,

$$\begin{aligned}\sqrt{T}\left(\hat{g}_T-g_\infty\right) &\overset{d}{=} \left[-\frac{1}{T}\sum_{t=1}^{T}\nabla_{gg}\Psi\left(y_t,x_t;g_\infty\right)\right]^{-1}\left[\frac{1}{\sqrt{T}}\sum_{t=1}^{T}\nabla_{g}\Psi\left(y_t,x_t;g_\infty\right)\right]\\ &\overset{d}{=} \left[E_xE_0\left(-\nabla_{gg}\Psi\left(y,x;g\right)\right)\right]^{-1}\cdot\frac{1}{\sqrt{T}}\sum_{t=1}^{T}\nabla_{g}\Psi\left(y_t,x_t;g_\infty\right)\\ &\overset{d}{=} \mathcal{J}^{-1}\cdot\frac{1}{\sqrt{T}}\sum_{t=1}^{T}\nabla_{g}\Psi\left(y_t,x_t;g_\infty\right).\end{aligned}$$

It is easily seen that $E_xE_0\left[\nabla_g\Psi\left(y,x;g_\infty\right)\right]=0$ (by the limiting problem). Since $var\left(\nabla_g\Psi\right)=E\left(\nabla_g\Psi\cdot\left[\nabla_g\Psi\right]^\top\right)=\mathcal{I}$, then,

$$\frac{1}{\sqrt{T}}\sum_{t=1}^{T}\nabla_{g}\Psi\left(y_t,x_t;g_\infty\right)\overset{d}{\to}N\left(0,\mathcal{I}\right).$$

The result follows by the Slutzky's theorem and symmetry of $\mathcal{J}$.

One simple example of M-estimator is the Nonlinear Least Square estimator,

$$\hat{\theta}_T=\arg\min_{\theta\in\Theta}\sum_{t=1}^{T}\left[y_t-m\left(x_t;\theta\right)\right]^2,$$

for some function m. In this case, $\Psi\left(x,y;\theta\right)=\left[y-m\left(x;\theta\right)\right]^2$. Section B.4 and B.5 below review additional estimators which can be seen as particular cases of these M-estimators.

10.5 Pseudo (or quasi) maximum likelihood

The maximum likelihood estimator is in fact a M-estimator. Just set $\Psi=\log L$ (the log-likelihood function). Moreover, assume that the model is well-specified. In this case, $\mathcal{J}=\mathcal{I}$ - which confirms we are back to the MLE described in section A.2.

But in general, it can well be the case that the true DGP is a density $\ell_0\left(y_t|\,x_t\right)$ which doesn't belong to the family of laws spanned by our model,

$$\ell_0\left(y_t|\,x_t\right)\notin\left(M\right)=\left\{f\left(y_t|\,x_t;\theta\right),\quad\theta\in\Theta\right\}.$$

We say that in this case, our model (M) is *misspecified.*

Suppose we insist in maximizing $\Psi=\log L$, where $L=\sum_t f\left(y_t|\,x_t;\theta\right)$, a misspecified density. In this case,

$$\sqrt{T}(\hat{\theta}_T-\theta_0^*)\overset{d}{\to}N\left(0,\mathcal{J}^{-1}\mathcal{I}\mathcal{J}^{-1}\right),$$

where θ_0^* is the "pseudo-true" value,[7] and

$$\mathcal{J}=-E_xE_0\left[\nabla_{\theta\theta}\log f(y_t|\,\underline{y_{t-1}};\theta_0^*)\right];\quad\mathcal{I}=E_xE_0\left(\nabla_\theta\log f(y_t|\,\underline{y_{t-1}};\theta_0^*)\cdot\left[\nabla_\theta\log f(y_t|\,\underline{y_{t-1}};\theta_0^*)\right]^\top\right).$$

[7] That is, θ_0^* is clearly solution to some (misspecified) limiting problem. But it is possible to show that it has a nice interpretation in terms of entropy distance minimizer.

In the presence of specification errors, $\mathcal{J} \neq \mathcal{I}$. But by comparing the two (estimated) matrixes may lead to detect specification errors. These tests are very standard. Finally, please note that in this general case, the variance-covariance matrix $\mathcal{J}^{-1}\mathcal{I}\mathcal{J}^{-1}$ depends on the unknown law of (y_t, x_t). To assess the precision of the estimates of $\hat{g}_T$, one needs to estimate such a variance-covariance matrix. A common practice is to use the following a.s. consistent estimators,

$$\hat{J} = -\frac{1}{T}\sum_{t=1}^{T} \nabla_{gg}\Psi(y_t, x_t; \hat{g}_T), \quad \text{and} \quad \hat{\mathcal{I}} = -\frac{1}{T}\sum_{t=1}^{T} \left(\nabla_g \Psi(y_t, x_t; \hat{g}_T) \left[\nabla_g \Psi(y_t, x_t; \hat{g}_T)^\top\right]\right).$$

10.6 GMM

Economic theory often places restrictions on models that have the following format,

$$E\left[h\left(y_t; \theta_0\right)\right] = \mathbf{0}_q, \tag{A1}$$

where $h : \mathbb{R}^n \times \Theta \mapsto \mathbb{R}^q$, θ_0 is the true parameter vector, y_t is the n-dimensional vector of the observable variables and $\Theta \subseteq \mathbb{R}^p$. The MLE is often unfeasible here. Moreover, the MLE requires specifying a density function. This is not the case here. Hansen (1982) proposed the following **G**eneralized **M**ethod of **M**oments (GMM) estimation procedure. Consider the sample counterpart to the population in eq. (A1),

$$\bar{h}\left(y_1^T; \theta\right) = \frac{1}{T}\sum_{t=1}^{T} h\left(y_t; \theta\right),$$

where we have rewritten h as a function of the parameter vector $\theta \in \Theta$. The basic idea of GMM is to find a θ which makes $\bar{h}\left(y_1^\top; \theta\right)$ as close as possible to zero. Precisely, we have,

Definition (GMM estimator): The GMM estimator *is the sequence* $\hat{\theta}_T$ *satisfying,*

$$\hat{\theta}_T = \arg\min_{\theta \in \Theta \subseteq \mathbb{R}^p} \underbrace{\bar{h}\left(y_1^T; \theta\right)^\top}_{1\times q} \cdot \underbrace{W_T}_{q\times q} \cdot \underbrace{\bar{h}\left(y_1^T; \theta\right)}_{q\times 1},$$

where $\{W_T\}$ *is a sequence of weighting matrices the elements of which may depend on the observations.*

The simplest analytical situation arises in the so-called *just-identified* case $p = q$. In this case, the GMM estimator is simply:[8]

$$\hat{\theta}_T : \bar{h}(y_1^T; \hat{\theta}_T) = \mathbf{0}_q.$$

But in general, one has $p \leq q$. If $p < q$, we say that the GMM estimator imposes *overidentifying* restrictions.

We analyze the i.i.d. case only. Under mild regularity conditions, there exists an optimal matrix W_T (i.e. a matrix which minimizes the asymptotic variance of the GMM estimator) which satisfies asymptotically,

$$W = \left[\lim_{T\to\infty} T \cdot E\left(\bar{h}(y_1^T; \hat{\theta}_T) \cdot \bar{h}(y_1^T; \hat{\theta}_T)^\top\right)\right]^{-1} \equiv \Sigma_0^{-1}.$$

[8]This is also easy to see by inspection of formula (A3) below.

An estimator of Σ_0 can be:

$$\Sigma_T = \frac{1}{T}\sum_{t=1}^{T}\left[h(y_t;\hat{\theta}_T)\cdot h(y_t;\hat{\theta}_T)^\top\right],$$

but because $\hat{\theta}_T$ depends on the weighting matrix Σ_T and vice versa, one needs to implement an iterative procedure. The more one iterates, the less the final results will depend on the initial weighting matrix $\Sigma_T^{(0)}$. For example, one can start with $\Sigma_T^{(0)} = \mathbf{I}_q$.

We have the following asymptotic theory:

Theorem: *Suppose to be given a sequence of GMM estimators $\hat{\theta}_T$ such that*:

$$\hat{\theta}_T \xrightarrow{p} \theta_0.$$

We have,

$$\sqrt{T}(\hat{\theta}_T - \theta_0) \xrightarrow{d} N\left(\mathbf{0}_p, \left[E\left(h_\theta\right)\Sigma_0^{-1}E\left(h_\theta\right)^\top\right]^{-1}\right), \quad \textit{where } E\left(h_\theta\right) \equiv E\left[\nabla_\theta h(y;\theta_0)\right].$$

Proof (Sketch): The assumption that $\hat{\theta}_T \xrightarrow{p} \theta_0$ is easy to check under mild regularity conditions - such as stochastic equicontinuity of the criterion. Moreover, the GMM satisfies,

$$\mathbf{0}_p = \underset{p\times q}{\nabla_\theta \bar{h}(y_1^T;\hat{\theta}_T)}\underset{q\times q}{\Sigma_T^{-1}}\underset{q\times 1}{\bar{h}(y_1^T;\hat{\theta}_T)}.$$

Eq. (A3) confirms that if $p = q$ the GMM satisfies $\hat{\theta}_T : \bar{h}(y_1^T;\hat{\theta}_T) = 0$. This is so because if $p = q$, then $\nabla_\theta h \Sigma_T^{-1}$ is full-rank, and (A3) can only be satisfied by $\bar{h} = 0$. In the general case $q > p$, we have,

$$\sqrt{T}\underset{q\times 1}{\bar{h}(y_1^T;\hat{\theta}_T)} = \sqrt{T}\,\underset{q\times 1}{\bar{h}\left(y_1^T;\theta_0\right)} + \underset{q\times p}{\left[\nabla_\theta \bar{h}\left(y_1^T;\theta_0\right)\right]^\top}\sqrt{T}(\hat{\theta}_T - \theta_0) + o_p(1).$$

By premultiplying both sides of the previous equality by $\nabla_\theta \bar{h}\left(y_1^T;\hat{\theta}_T\right)\Sigma_T^{-1}$,

$$\begin{aligned}&\sqrt{T}\nabla_\theta \bar{h}(y_1^T;\hat{\theta}_T)\Sigma_T^{-1}\cdot \bar{h}(y_1^T;\hat{\theta}_T)\\ =\ &\sqrt{T}\nabla_\theta \bar{h}(y_1^T;\hat{\theta}_T)\Sigma_T^{-1}\cdot \bar{h}\left(y_1^T;\theta_0\right) + \nabla_\theta \bar{h}(y_1^T;\hat{\theta}_T)\Sigma_T^{-1}\cdot\left[\nabla_\theta \bar{h}\left(y_1^T;\theta_0\right)\right]^\top\sqrt{T}(\hat{\theta}_T - \theta_0) + o_p(1).\end{aligned}$$

The l.h.s. of this equality is zero by the first order conditions in eq. (A3). By rearranging terms,

$$\begin{aligned}\sqrt{T}(\hat{\theta}_T - \theta_0) &\overset{d}{=} -\left(\nabla_\theta \bar{h}\left(y_1^T;\theta_0\right)\Sigma_T^{-1}\left[\nabla_\theta \bar{h}\left(y_1^T;\theta_0\right)\right]^\top\right)^{-1}\nabla_\theta \bar{h}(y_1^T;\hat{\theta}_T)\Sigma_T^{-1}\cdot\sqrt{T}\bar{h}\left(y_1^T;\theta_0\right)\\ &= -\left(\frac{1}{T}\sum_{t=1}^{T}\nabla_\theta h(y_t;\hat{\theta}_T)\Sigma_T^{-1}\frac{1}{T}\sum_{t=1}^{T}\left[\nabla_\theta h(y_t;\hat{\theta}_T)\right]^\top\right)^{-1}\frac{1}{T}\sum_{t=1}^{T}\nabla_\theta h(y_t;\hat{\theta}_T)\Sigma_T^{-1}\\ &\qquad\times\sqrt{T}\bar{h}\left(y_1^T;\theta_0\right)\\ &\overset{d}{=} -\left(E\left(h_\theta\right)\Sigma_0^{-1}E\left(h_\theta\right)^\top\right)^{-1}E\left(h_\theta\right)\Sigma_0^{-1}\cdot\frac{1}{\sqrt{T}}\sum_{t=1}^{T}h\left(y_t;\theta_0\right).\end{aligned}$$

Next, the term $\frac{1}{\sqrt{T}}\sum_{t=1}^{T} h(y_t;\theta_0)$ satisfies a central limit theorem:

$$\frac{1}{\sqrt{T}}\sum_{t=1}^{T} h(y_t;\theta_0) \xrightarrow{d} N(E(h), var(h)),$$

where $E(h)=0$ (by eq. (A1)), and $var(h)=E\left(h\cdot h^\top\right)=\Sigma_0$. We then have:

$$\frac{1}{\sqrt{T}}\sum_{t=1}^{T} h(y_t;\theta_0) \xrightarrow{d} N(0,\Sigma_0).$$

Therefore, $\sqrt{T}(\hat{\theta}_T-\theta_0)$ is asymptotic normal with expectation $\mathbf{0}_p$, and variance,

$$\left[E(h_\theta)\Sigma_0^{-1}E(h_\theta)^\top\right]^{-1}E(h_\theta)\Sigma_0^{-1}\Sigma_0\Sigma_0^{-1}E(h_\theta)^\top\left[E(h_\theta)\Sigma_0^{-1}E(h_\theta)^\top\right]^{\top-1}=\left[E(h_\theta)\Sigma_0^{-1}E(h_\theta)^\top\right]^{-1}.$$

$\|$

A global specification test is the celebrated test of "overidentifying restrictions". First, consider the behavior of the statistic,

$$\sqrt{T}\bar{h}\left(y_1^T;\theta_0\right)^\top\Sigma_0^{-1}\sqrt{T}\bar{h}\left(y_1^T;\theta_0\right)^\top \xrightarrow{d} \chi^2(q).$$

One might be led to think that the same result applies when θ_0 is replaced with $\hat{\theta}_T$ (which is a consistent estimator of θ_0). Wrong. Consider,

$$\mathcal{C}_T=\sqrt{T}\bar{h}(y_1^T;\hat{\theta}_T)^\top\Sigma_T^{-1}\cdot\sqrt{T}\bar{h}(y_1^T;\hat{\theta}_T).$$

We have,

$$\begin{aligned}
\sqrt{T}\bar{h}(y_1^T;\hat{\theta}_T) &\stackrel{d}{=} \sqrt{T}\bar{h}\left(y_1^T;\theta_0\right)+\nabla_\theta\bar{h}\left(y_1^T;\theta_0\right)\sqrt{T}(\hat{\theta}_T-\theta_0)\\
&\stackrel{d}{=} \sqrt{T}\bar{h}\left(y_1^T;\theta_0\right)-\left[\nabla_\theta\bar{h}\left(y_1^T;\theta_0\right)\right]^\top\left[E(h_\theta)\Sigma_0^{-1}E(h_\theta)^\top\right]^{-1}E(h_\theta)\Sigma_0^{-1}\cdot\sqrt{T}\bar{h}\left(y_1^T;\theta_0\right)\\
&\stackrel{d}{=} \sqrt{T}\bar{h}\left(y_1^T;\theta_0\right)-E(h_\theta)^\top\left[E(h_\theta)\Sigma_0^{-1}E(h_\theta)^\top\right]^{-1}E(h_\theta)\Sigma_0^{-1}\cdot\sqrt{T}\bar{h}\left(y_1^T;\theta_0\right)\\
&= \underset{q\times q}{(\mathbf{I}_q-\mathbf{P})}\underset{q\times 1}{\sqrt{T}\bar{h}\left(y_1^T;\theta_0\right)},
\end{aligned}$$

where the second asymptotic equality is due to eq. (A4), and

$$\mathbf{P}_q\equiv E(h_\theta)^\top\left[E(h_\theta)\Sigma_0^{-1}E(h_\theta)^\top\right]^{-1}E(h_\theta)\Sigma_0^{-1}$$

is the orthogonal projector in the space generated by the columns of $E(h_\theta)$ by the inner product Σ_0^{-1}. We have thus shown that,

$$\mathcal{C}_T\stackrel{d}{=}\sqrt{T}\bar{h}\left(y_1^T;\theta_0\right)^\top(\mathbf{I}_q-\mathbf{P}_q)^\top\Sigma_T^{-1}(\mathbf{I}-\mathbf{P}_q)\sqrt{T}\bar{h}\left(y_1^T;\theta_0\right).$$

But,

$$\sqrt{T}\bar{h}\left(y_1^T;\theta_0\right)\xrightarrow{d}N(0,\Sigma_0),$$

and by a classical result,

$$\mathcal{C}_T \xrightarrow{d} \chi^2(q-p).$$

EXAMPLE. Consider the classical system of Euler equations arising in the Lucas (1978) model,

$$E\left[\beta\frac{u'(c_{t+1})}{u'(c_t)}(1+r_{i,t+1})-1\right|\left.\mathcal{F}_t\right]=0, \quad i=1,\cdots,m_1,$$

where r_i is the return on asset i, m_1 is the number of assets and β is the psychological discount factor. If $u(x)=\frac{x^{1-\gamma}}{1-\gamma}$, and the model is literally taken (i.e. it is well-specified), there exist some β_0 and γ_0 such that the previous system can be written as:

$$E\left[\beta_0\left(\frac{c_{t+1}}{c_t}\right)^{-\gamma_0}(1+r_{i,t+1})-1\right|\left.\mathcal{F}_t\right]=0, \quad i=1,\cdots,m_1.$$

Here we have $p=2$, and to estimate the true parameter vector $\theta_0 \equiv (\beta_0,\gamma_0)$, we may build up a system of orthogonality conditions. Such orthogonality conditions are naturally suggested by the previous system,

$$E[h(y_t;\theta_0)]=0,$$

where

$$\underset{m_1\cdot m_2}{h(y_t;\theta)}=\begin{pmatrix} \left[\beta\left(\frac{c_{t+1}}{c_t}\right)^{-\gamma}(1+r_{1,t+1})-1\right] & \times & \begin{pmatrix}\text{instr.}_{1,t}\\ \vdots \\ \text{instr.}_{m_2,t}\end{pmatrix} \\ \vdots & \vdots & \vdots \\ \left[\beta\left(\frac{c_{t+1}}{c_t}\right)^{-\gamma}(1+r_{m_1,t+1})-1\right] & \times & \begin{pmatrix}\text{instr.}_{1,t}\\ \vdots \\ \text{instr.}_{m_2,t}\end{pmatrix}\end{pmatrix},$$

and $(\text{instr.}_{j,t})_{j=1}^{m_2}$ is the set of instruments used to produce the previous orthogonality restrictions. E.g.: past values of $\frac{c_{t+1}}{c_t}$, r_i, constants, etc.

10.7 Simulation-based estimators

Ideally, Maximum Likelihood (ML) estimation is the preferred estimation method of parametric Markov models because it leads to first-order efficiency. Yet economic theory often places model's restrictions that make these models impossible to estimate through maximum ML. Instead, GMM arises as a natural estimation method. Unfortunately, GMM is unfeasible in many situations of interest. Let us be a bit more specific. Assume that the data generating process is not IID, and that instead it is generated by the transition function,

$$y_{t+1}=H(y_t,\epsilon_{t+1},\theta_0), \tag{10.1}$$

where $H:\mathbb{R}^n\times\mathbb{R}^d\times\Theta\mapsto\mathbb{R}^n$, and ϵ_t is the vector of IID disturbances in $\mathbb{R}^d$. It is assumed that the econometrician knows the function H. Let $z_t=(y_t,y_{t-1},\cdots,y_{t-l+1})$, $l<\infty$. In many applications, the function $\bar{h}$ can be written as,

$$\bar{h}\left(y_1^T;\theta\right)=\frac{1}{T}\sum_{t=1}^{T}\left[f_t^*-E\left(f\left(z_t,\theta\right)\right)\right], \tag{10.2}$$

where,

$$f_t^* = f(z_t, \theta_0),$$

is a vector-valued moment function, or "observation function" - that summarize satisfactorily the data we may say. But if we are not even able to compute the expectation $E(f(z_t, \theta))$ in closed form (that is, $\theta \mapsto E(f(z, \theta))$ is not known analytically), the GMM estimator is unfeasible. Simulation-based methods can make the method of moments feasible in cases in which this expecation can not be computed in closed-form.

10.7.1 Background

The basic idea underlying simulation-based methods is very simple. Even if the moment conditions are too complex to be evaluated analytically, in many cases of interest the underlying model in eq. (10.1) can be simulated. Draw ϵ_t from its distribution, and save the simulated values $\hat{\epsilon}_t$. Compute recursively,

$$y_{t+1}^\theta = H\left(y_t^\theta, \hat{\epsilon}_{t+1}, \theta\right),$$

and create simulated moment functions as follows,

$$f_t^\theta \equiv f\left(z_t^\theta, \theta\right).$$

Consider the following parameter estimator,

$$\theta_T = \arg\min_{\theta \in \Theta} G_T(\theta)^\top W_T G_T(\theta), \tag{10.3}$$

where $G_T(\theta)$ is the simulated counterpart to $\bar{h}$ in eq. (10.2),

$$G_T(\theta) = \frac{1}{T} \sum_{t=1}^T \left(f_t^* - \frac{1}{S(T)} \sum_{s=1}^{S(T)} f_s^\theta \right),$$

where $S(T)$ is the simulated sample size, which is made dependent on the sample size T for the purpose of the asymptotic theory. In other terms, the estimator θ_T is the one that best matches the sample properties of the actual and simulated processes f_t^* and f_t^θ. This estimator is known as the SMM estimator.

Estimation methods relying on the IIP work very similarly. Instead of minimizing the distance of some moment conditions, the IIP relies on minimizing the parameters of an auxiliary (possibly misspecified) model. For example, consider the following auxiliary parameter estimator,

$$\beta_T = \arg\max_\beta \log L\left(y_1^T; \beta\right), \tag{10.4}$$

where L is the likelihood of some (possibly misspecified) model. Consider simulating S times the process y_t in eq. (10.1), and computing,

$$\beta_T^s(\theta) = \arg\max_\beta \log L(y_s(\theta)_1^T; \beta), \quad s = 1, \cdots, S,$$

where $y_s(\theta)_1^T = (y_t^{\theta,s})_{t=1}^T$ are the simulated variables (for $s = 1, \cdots, S$) when the parameter vector is θ. The IIP-based estimator is defined similarly as θ_T in eq. (10.3), but with the function G_T given by,

$$G_T(\theta) = \beta_T - \frac{1}{S} \sum_{s=1}^S \beta_T^s(\theta). \tag{10.5}$$

The diagram in Figure 10.1 illustrates the main ideas underlying the IIP.

Finally, the EMM estimator also works very similarly. It sets,

$$G_T(\theta) = \frac{1}{N}\sum_{n=1}^{N} \frac{\partial}{\partial \beta} \log f\left(y_n^\theta \middle| z_{n-1}^\theta ; \beta_T \right),$$

where $\frac{\partial}{\partial \beta} \log f\left(y \middle| z ; \beta \right)$ is the score of some auxiliary model, β_T is the Pseudo ML estimator of the auxiliary model, and finally $(y_n^\theta)_{n=1}^N$ is a long simulation (i.e. N is very high) of eq. (10.1) when the parameter vector is set equal to θ.

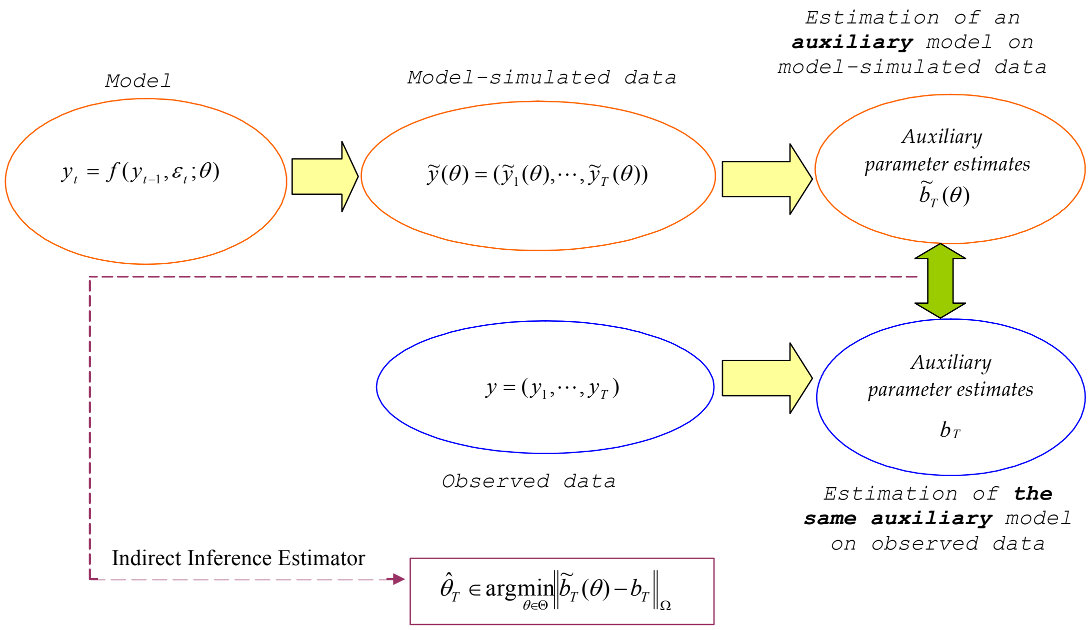


FIGURE 10.1. *The Indirect Inference principle*. Given the *true* model $y_t = f(y_{t-1}, \epsilon_t; \theta)$, an estimator of θ based on the indirect inference principle ($\hat{\theta}_T$ say) makes the parameters of some *auxiliary* model $\tilde{b}_T(\hat{\theta}_T)$ as close as possible to the parameters b_T of *the same auxiliary* model estimated on real data. That is, $\hat{\theta}_T = \arg\min \left\| \tilde{b}_T(\theta) - b_T \right\|_\Omega$, for some norm Ω.

10.7.2 Asymptotic normality for the SMM estimator

Let us derive heuristically asymptotic normality results for the SMM. Let,

$$\Sigma_0 = \sum_{j=-\infty}^{\infty} E\left[\left(f_t^* - E\left(f_t^*\right)\right)\left(f_{t-j}^* - E\left(f_{t-j}^*\right)\right)^\top \right],$$

and suppose that

$$W_T \xrightarrow{p} W_0 = \Sigma_0^{-1}.$$

We now demonstrate that under this condition, as $T \to \infty$ and $S(T) \to \infty$,

$$\sqrt{T}(\theta_T - \theta_0) \xrightarrow{d} N\left(\mathbf{0}_p, (1+\tau)\left[D_0^\top \Sigma_0^{-1} D_0\right]^{-1}\right), \tag{10.6}$$

where $\tau = \lim_{T\to\infty} \frac{T}{S(T)}$, $D_0 = E\left(\nabla_\theta G_\infty(\theta_0)\right) = E\left(\nabla_\theta f_\infty^{\theta_0}\right)$, and the notation G_∞ means that $G_.$ is drawn from its stationary distribution.

Indeed, the first order conditions satisfied by the SMM in eq. (10.3) are,

$$\mathbf{0}_p = \left[\nabla_\theta G_T(\theta_T)\right]^\top W_T G_T(\theta_T) = \left[\nabla_\theta G_T(\theta_T)\right]^\top W_T \cdot \left[G_T(\theta_0) + \nabla_\theta G_T(\theta_0)(\theta_T - \theta_0)\right] + o_p(1).$$

That is,

$$\begin{aligned}\sqrt{T}(\theta_T - \theta_0) &\stackrel{d}{=} -\left(\left[\nabla_\theta G_T(\theta_T)\right]^\top W_T \nabla_\theta G_T(\theta_0)\right)^{-1} \left[\nabla_\theta G_T(\theta_T)\right]^\top W_T \cdot \sqrt{T} G_T(\theta_0) \\ &\stackrel{d}{=} -\left(D_0^\top W_0 D_0\right)^{-1} D_0^\top W_0 \cdot \sqrt{T} G_T(\theta_T) = -\left(D_0^\top \Sigma_0^{-1} D_0\right)^{-1} D_0^\top \Sigma_0^{-1} \cdot \sqrt{T} G_T(\theta_0).\end{aligned} \tag{10.7}$$

We have,

$$\begin{aligned}
\sqrt{T}G_T(\theta_0) &= \sqrt{T}\cdot\frac{1}{T}\sum_{t=1}^{T}\left(f_t^* - \frac{1}{S(T)}\sum_{s=1}^{S(T)} f_s^{\theta_0}\right)\\
&= \frac{1}{\sqrt{T}}\sum_{t=1}^{T}\left(f_t^* - E\left(f_\infty^*\right)\right) - \frac{\sqrt{T}}{\sqrt{S(T)}}\cdot\frac{1}{\sqrt{S(T)}}\sum_{s=1}^{S(T)}\left(f_s^{\theta_0} - E\left(f_\infty^{\theta_0}\right)\right)\\
&\overset{d}{\to} N\left(0, (1+\tau)\,\Sigma_0\right),
\end{aligned}$$

where we have used the fact that $E\left(f_\infty^*\right) = E\left(f_\infty^{\theta_0}\right)$. By replacing this result into eq. (10.7) produces the convergence in eq. (10.6). If $\tau = \lim_{T\to\infty}\frac{T}{S(T)} = 0$ (i.e. if the number of simulations grows more fastly than the sample size), the SMM estimator is as efficient as the GMM estimator. Moreover, inspection of the asymptotic result in eq. (10.6) reveals that we must have $\tau = \lim_{T\to\infty}\frac{T}{S(T)} < \infty$. This condition means that the number of simulations $S(T)$ can not grow more slowly than the sample size.

10.7.3 Asymptotic normality for the IIP-based estimator

The IIP-based estimator works slightly differently. For this estimator, even if the number of simulations S is fixed, asymptotic normality obtains without the need to impose that S goes to infinity more fastly than the sample size. Basically, what really matters here is that ST goes to infinity. To demonstrate this claim, we need to derive the asymptotic theory for the IIP-based estimator. By eq. (10.7), and the discussion in Section 10.6.1, we know that asymptotically, the first order conditions satisfied by the IIP-based estimator are,

$$\sqrt{T}\left(\theta_T - \theta_0\right) \overset{d}{=} -\left(D_0^\top W_0 D_0\right)^{-1} D_0^\top W_0 \cdot \sqrt{T}G_T(\theta_0),$$

where G_T is as in eq. (10.5), $D_0 = \nabla_\theta b(\theta)$, and $b(\theta)$ is solution to the limiting problem corresponding to the estimator in (10.4), viz,

$$b(\theta) = \arg\max_\beta\left(\lim_{T\to\infty}\frac{1}{T}\log L\left(y_1^T;\beta\right)\right).$$

We need to find the distribution of G_T in eq. (10.5). We have,

$$\begin{aligned}
\sqrt{T}G_T(\theta_0) &= \frac{1}{S}\sum_{s=1}^{S}\sqrt{T}\left(\beta_T - \beta_T^s(\theta_0)\right)\\
&= \frac{1}{S}\sum_{s=1}^{S}\sqrt{T}\left[\left(\beta_T - \beta_0\right) - \left(\beta_T^s(\theta_0) - \beta_0\right)\right]\\
&= \sqrt{T}\left(\beta_T - \beta_0\right) - \frac{1}{S}\sum_{s=1}^{S}\sqrt{T}\left(\beta_T^s(\theta_0) - \beta_0\right),
\end{aligned}$$

where $\beta_0 = b(\theta_0)$. Hence, given the independence of the sample and the simulations,

$$\sqrt{T}G_T(\theta_0) \overset{d}{\to} N\left(0, \left(1+\frac{1}{S}\right)\cdot \text{asy.var}\left(\sqrt{T}\beta_T\right)\right).$$

That is, asymptotically S can be fixed with respect to T.

[Discuss the choice of the optimal weighting matrix]

10.7.4 Asymptotic normality for the EMM estimator

10.8 Appendix 1: Notions of convergence

Convergence in probability. *A sequence of random vectors* $\{x_T\}$ converges in probability *to the random vector* $\tilde{x}$ *if for each* $\epsilon > 0$, $\delta > 0$ *and each* $i = 1, 2, \cdots, N$, *there exists a* $T_{\epsilon,\delta}$ *such that for every* $T \geq T_{\epsilon,\delta}$,

$$P\left(|x_{Ti} - \tilde{x}_i| > \delta\right) < \epsilon.$$

This is succinctly written as $x_T \stackrel{p}{\to} \tilde{x}$, *or* $\text{plim } x_T = \bar{x}$, *if* $\tilde{x} \equiv \bar{x}$, *a constant.*

The previous notion generalizes in a straight forward manner the standard notion of a limit of a deterministic sequence - which tells us that a sequence $\{x_T\}$ converges to $\bar{x}$ if for every $\kappa > 0$ there exists a T_κ : for each $T \geq T_\kappa$ we have that $|x_T - \bar{x}| < \kappa$. And indeed, convergence in probability can also be rephrased as follows,

$$\lim_{T\to\infty} P\left(|x_{Ti} - \tilde{x}_i| > \delta\right) = 0.$$

Here is a stronger definition of convergence:

Almost sure convergence. *A sequence of random vectors* $\{x_T\}$ converges almost surely *to the random vector* $\tilde{x}$ *if, for each* $i = 1, 2, \cdots, N$, *we have*:

$$P\left(\omega : x_{Ti}(\omega) \to \tilde{x}_i\right) = 1,$$

where ω *denotes the entire random sequence* x_{Ti}. *This is succinctly written as* $x_T \stackrel{a.s.}{\to} \tilde{x}$.

Naturally, almost sure convergence implies convergence in probability. Convergence in probability means that for each $\epsilon > 0$, $\lim_{T\to\infty} P\left(\omega : |x_{Ti}(\omega) - \tilde{x}_i| < \epsilon\right) \to 1$. Almost sure convergence requires that $P\left(\lim_{T\to\infty} x_{Ti} \to \tilde{x}_i\right) = 1$ or that $\lim_{T'\to\infty} P\left(\sup_{T\geq T'} |x_{Ti} - \tilde{x}_i| > \delta\right) = \lim_{T'\to\infty} P\left(\bigcup_{T\geq T'} |x_{Ti} - \tilde{x}_i| > \delta\right)$ 0.

Next, assume that the second order moments of all x_i are finite. We also have:

Convergence in quadratic mean. *A sequence of random vectors* $\{x_T\}$ converges in quadratic mean *to the random vector* $\tilde{x}$ *if for each* $i = 1, 2, \cdots, N$, *we have*:

$$\lim_{T'\to\infty} E\left[(x_{Ti} - \tilde{x}_i)^2\right] \to 0.$$

This is succinctly written as $x_T \stackrel{q.m.}{\to} \tilde{x}$.

Remark. By Chebyshev's inequality,[9]

$$P\left(|x_{Ti} - \tilde{x}_i| > \delta\right) \leq \frac{E\left[(x_{Ti} - \tilde{x}_i)^2\right]}{\delta^2},$$

which shows that q.m.-convergence implies convergence in probability.

We now turn to a weaker notion of convergence:

[9] Let x be centered with density f. We have:

$$\begin{aligned} E\left(x^2\right) &= \int_{-\infty}^{\infty} x^2 f(x)\,dx \geq \int_{|x|>\delta} x^2 f(x)dx = \int_{-\infty}^{-\delta} x^2 f(x)dx + \int_{\delta}^{\infty} x^2 f(x)dx \\ &\geq \delta^2 \int_{-\infty}^{-\delta} f(x)dx + \delta^2 \int_{\delta}^{\infty} f(x)dx = \delta^2 P\left(|x| > \delta\right). \end{aligned}$$

Next, let $x = y - \bar{y}$ to complete the proof of the Chebyshev inequality.

Convergence in distribution. *Let* $\{f_T(x)\}_T$ *be the sequence of probability distributions (that is,* $f_T(x) = pr(x_T \leq x)$*) of the sequence of the random vectors* $\{x_T\}$*. Let* $\tilde{x}$ *be a random vector with probability distribution* $f(x)$*. A sequence* $\{x_T\}$ *converges in distribution to* $\tilde{x}$ *if, for each* $i = 1, 2, \cdots, N$*, we have:*

$$\lim_{T\to\infty} f_T(x) = f(x).$$

This is succinctly written as $x_T \xrightarrow{d} \tilde{x}$*.*

The following two results are extremely useful:

Slutzky's theorem. *If* $y_T \xrightarrow{p} \bar{y}$ *and* $x_T \xrightarrow{d} \tilde{x}$*, then:*

$$y_T \cdot x_T \xrightarrow{d} \bar{y} \cdot \tilde{x}.$$

Cramer-Wold device. *Let* λ *be a* N*-dimensional vector of constants. We have:*

$$x_T \xrightarrow{d} \tilde{x} \Leftrightarrow \lambda^\top \cdot x_T \xrightarrow{d} \lambda^\top \cdot \tilde{x}.$$

Here is one illustration of the Cramer-Wold device. If $\lambda^\top \cdot x_T \xrightarrow{d} N\left(0; \lambda^\top \Sigma \lambda\right)$, then $x_T \xrightarrow{d} N\left(0; \Sigma\right)$.

10.8.1 Laws of large numbers

The first two laws concern convergence in probability.

Weak law (No. 1) (Khinchine): *Let* $\{x_T\}$ *a i.i.d. (identically, independently distributed) sequence with* $E(x_T) = \mu < \infty \ \forall T$*. We have:*

$$\bar{x}_T \equiv \frac{1}{T}\sum_{t=1}^{T} x_t \xrightarrow{p} \mu.$$

Weak law (No. 2) (Chebyshev): *Let* $\{x_T\}$ *a sequence independently but not identically distributed, with* $E(x_T) = \mu_T < \infty$ *and* $E\left[(x_T - \mu_T)^2\right] = \sigma_T^2 < \infty$*. If* $\lim_{T\to\infty} \frac{1}{T^2}\sum_{t=1}^{T} \sigma_t^2 \to 0$*, then:*

$$\bar{x}_T \equiv \frac{1}{T}\sum_{t=1}^{T} x_t \xrightarrow{p} \bar{\mu}_T \equiv \frac{1}{T}\sum_{t=1}^{T} \mu_t.$$

Kolmogorov formulated the strong versions of the previous weak laws.

10.8.2 The central limit theorem

We state and prove the central limit theorem for scalar and i.i.d. processes. Let us be given a scalar i.i.d. sequence $\{x_T\}$, with $E(x_T) = \mu < \infty$ and $E\left[(x_T - \mu)^2\right] = \sigma^2 < \infty \ \forall T$. Let $\bar{x}_T \equiv \frac{1}{T}\sum_{t=1}^{T} x_t$. We have,

$$\frac{\sqrt{T}\left(\bar{x}_T - \mu\right)}{\sigma} \xrightarrow{d} N(0, 1).$$

The multidimensional version of this theorem only needs a change in notation.

The classic method of proof of the central limit theorem makes use of characteristic functions. Let:

$$\varphi(t) \equiv E\left(e^{\mathrm{i}tx}\right) = \int e^{\mathrm{i}tx} f(x)dx, \ i \equiv \sqrt{-1}.$$

We have $\frac{\partial^r}{\partial t^r}\varphi(t)\big|_{t=0} = \mathrm{i}^r m^{(r)}$, where $m^{(r)}$ is the r-th order moment. By a Taylor's expansion in the neighborhood of $t = 0$,

$$\begin{aligned}\varphi(t) &= \varphi(0) + \frac{\partial}{\partial t}\varphi(t)\bigg|_{t=0} \cdot t + \frac{1}{2}\frac{\partial^2}{\partial t^2}\varphi(t)\bigg|_{t=0} \cdot t^2 + \cdots \\ &= 1 + i \cdot m^{(1)} \cdot t - m^{(2)} \cdot \frac{1}{2}t^2 + \cdots\end{aligned}$$

Next, consider the random variable,

$$Y_T \equiv \frac{\sqrt{T}\left(\bar{x}_T - \mu\right)}{\sigma} = \frac{\sqrt{T}\left(\frac{1}{T}\sum_{t=1}^T x_t - \mu\right)}{\sigma} = \frac{1}{\sqrt{T}}\sum_{t=1}^T \frac{(x_t - \mu)}{\sigma}.$$

Its characteristic function is[10]

$$\varphi_{Y_T}(t) = \left[\varphi\left(\frac{t}{\sqrt{T}}\right)\right]^T = \left[1 - \frac{1}{2}\frac{t^2}{T} + o\left(T^{-1}\right)\right]^T.$$

Clearly, $\lim_{T\to\infty}\varphi_{Y_T}(t) = e^{-\frac{1}{2}t^2}$ - the characteristic function of a standard Gaussian variable.

10.9 Appendix 2: some results for dependent processes

A. *Weak dependence*

Let $\sigma_T^2 = var\left(\sum_{t=1}^T x_t\right)$, and suppose that $\sigma_T^2 = O(T)$, and $\sigma_T^2 = O\left(T^{-1}\right)$. If

$$\sigma_T^{-1}\sum_{t=1}^T (x_t - E(x_t)) \xrightarrow{d} N(0,1),$$

we say that $\{x_t\}$ is *weakly dependent*. The term "nonergodic" is reserved for those processes that exhibit such strong dependence that they do not even satisfy the LLN.

- Stationarity
- Weak dependence
- Ergodicity

B. *The central limit theorem for martingale difference sequences*

Let Let $\{x_t\}$ be a martingale difference sequence with

$$E\left(x_t^2\right) = \sigma_t^2 < \infty, \quad \forall t,$$

and define $\bar{x}_T \equiv \dfrac{1}{T}\sum_{t=1}^T x_t$, and $\bar{\sigma}_T^2 \equiv \dfrac{1}{T}\sum_{t=1}^T \sigma_t^2$. Let,

$$\lim_{T\to\infty}\frac{1}{T\bar{\sigma}_T^2}\sum_{t=1}^T x_t^2 \mathbb{I}_{|x_t|\geq\epsilon\cdot T\cdot\bar{\sigma}_T^2} = 0, \quad \forall\epsilon, \quad \text{and} \quad \frac{1}{T}\sum_{t=1}^T x_t^2 - \bar{\sigma}_T^2 \xrightarrow{p} 0.$$

Under the previous condition,

$$\frac{\sqrt{T}\cdot\bar{x}_T}{\bar{\sigma}_T} \xrightarrow{d} N(0,1).$$

[10] Indeed, $Y_T = \sum_{t=1}^T a_t$ where $a_t \equiv \frac{x_t - \mu}{\sqrt{T}\sigma}$ and a_t is i.i.d. with $E(a_t) = 0$ and $E(a_t^2) = \frac{1}{T}$ $\forall t$. The characteristic function of Y_T is then the product of the characteristic functions of a_t (which are the same): $\varphi_{Y_T}(t) = (\varphi_a(t))^T$, where $\varphi_a(t) = 1 - \frac{t^2}{2T} + \cdots$.

C. Applications to maximum likelihood estimation

We now use the central limit theorem for martingale difference sequences to show asymptotic normality of the MLE in the case of weakly dependent processes. We have,

$$\log L_T(\theta)=\sum_{t=1}^{T}\ell_t(\theta),\quad \ell_t(\theta)\equiv\ell(\theta;y_t|\,x_t).$$

The MLE satisfies the following first order conditions,

$$\mathbf{0}_p=\nabla_\theta\log L_T(\theta)|_{\theta=\hat\theta_T}\overset{d}{=}\sum_{t=1}^{T}\nabla_\theta\ell_t(\theta)|_{\theta=\theta_0}+\sum_{t=1}^{T}\nabla_{\theta\theta}\ell_t(\theta)|_{\theta=\theta_0}(\hat\theta_T-\theta_0),$$

whence

$$\sqrt{T}(\hat\theta_T-\theta_0)\overset{d}{=}-\left[\frac{1}{T}\sum_{t=1}^{T}\nabla_{\theta\theta}\ell_t(\theta_0)\right]^{-1}\cdot\frac{1}{\sqrt{T}}\sum_{t=1}^{T}\nabla_\theta\ell_t(\theta_0).$$

We have:

$$E_{\theta_0}\left[\nabla_\theta\ell_{t+1}(\theta_0)|\,F_t\right]=\mathbf{0}_p,$$

which shows that $\left\{\frac{\partial\ell_t(\theta_0)}{\partial\theta}\right\}$ is a martingale difference. Naturally, here we also have that:

$$E_{\theta_0}\left(|\nabla_\theta\ell_{t+1}(\theta_0)|_2|\,F_t\right)=-E_{\theta_0}\left(\nabla_{\theta\theta}\ell_{t+1}(\theta_0)|\,F_t\right)\equiv\mathcal{J}_t(\theta_0).$$

Next, for a given constant $c\in\mathbb{R}^p$, let:

$$x_t\equiv c^\top\nabla_\theta\ell_t(\theta_0).$$

Clearly, $\{x_t\}$ is also a martingale difference. Furthermore,

$$E_{\theta_0}\left(x_{t+1}^2|\,F_t\right)=-c^\top\mathcal{J}_t(\theta_0)\,c,$$

and because $\{x_t\}$ is a martingale difference, x_t and x_{t-i} are mutually uncorrelated,[11] all i. It follows that,

$$\begin{aligned}var\left(\sum_{t=1}^{T}x_t\right)&=\sum_{t=1}^{T}E\left(x_t^2\right)\\&=\sum_{t=1}^{T}c^\top E_{\theta_0}\left(|\nabla_\theta\ell_t(\theta_0)|_2\right)c\\&=\sum_{t=1}^{T}c^\top E_{\theta_0}\left[E_{\theta_0}\left(|\nabla_\theta\ell_t(\theta_0)|_2|\,F_{t-1}\right)\right]c\\&=-\sum_{t=1}^{T}c^\top E_{\theta_0}\left[\mathcal{J}_{t-1}(\theta_0)\right]c\\&=-c^\top\left[\sum_{t=1}^{T}E_{\theta_0}\left(\mathcal{J}_{t-1}(\theta_0)\right)\right]c.\end{aligned}$$

[11] $E(x_tx_{t-i})=E\left[E\left(x_t\cdot x_{t-i}|\,F_{t-i}\right)\right]=E\left[E\left(x_t|\,F_{t-i}\right)\cdot x_{t-i}\right]=0.$

Next, define as in subsection A.6.2:

$$\bar{x}_T \equiv \frac{1}{T}\sum_{t=1}^{T} x_t \quad \text{and} \quad \bar{\sigma}_T^2 \equiv \frac{1}{T}\sum_{t=1}^{T} E\left(x_t^2\right) = -c^\top \left[\frac{1}{T}\sum_{t=1}^{T} E_{\theta_0}\left(\mathcal{J}_{t-1}\left(\theta_0\right)\right)\right] c.$$

Under the conditions of subsection A.6.2,

$$\frac{\sqrt{T}\bar{x}_T}{\bar{\sigma}_T} \xrightarrow{d} N\left(0,1\right).$$

By the Cramer-Wold device,

$$\left[\frac{1}{T}\sum_{t=1}^{T} E_{\theta_0}\left(\mathcal{J}_{t-1}\left(\theta_0\right)\right)\right]^{-1/2} \frac{1}{\sqrt{T}}\sum_{t=1}^{T} \nabla_\theta \ell_t(\theta_0) \xrightarrow{d} N\left(0, \mathbf{I}_p\right).$$

Namely, the conditions that need to be satisfied are,

$$\frac{1}{T}\sum_{t=1}^{T} \nabla_{\theta\theta}\ell_t\left(\theta_0\right) - \frac{1}{T}\sum_{t=1}^{T} E_{\theta_0}\left[\mathcal{J}_{t-1}\left(\theta_0\right)\right] \xrightarrow{p} 0, \quad \text{and} \quad p\lim \frac{1}{T}\sum_{t=1}^{T} E_{\theta_0}\left[\mathcal{J}_{t-1}\left(\theta_0\right)\right] = \mathcal{J}_\infty\left(\theta_0\right) \text{ (say)}.$$

In this case, it follows by eq. (A5) that,

$$\sqrt{T}(\hat{\theta}_T - \theta_0) \xrightarrow{d} N\left(\mathbf{0}_p, \mathcal{J}_\infty\left(\theta_0\right)^{-1}\right).$$

References

Amemiya, T. (1985): *Advanced Econometrics.* Cambridge, Mass.: Harvard University Press.

Fisher, R.A. (1912): "On an Absolute Criterion for Fitting Frequency Curves," *Messages of Mathematics*: 41, 155-157.

Gauss, C.F. (1816): "Bestimmung der Genanigkeit der Beobachtungen," *Zeitschrift für Astronomie und Verwandte Wissenschaften*: 1, 185-196.

Hansen, L.P. (1982): "Large Sample Properties of Generalized Method of Moments Estimators," *Econometrica*: 50, 1029-1054.

Neyman, J. and E.S. Pearson (1928): "On the Use and Interpretation of Certain Test Criteria for Purposes of Statistical Inference," *Biometrika*: 20A, 175-240, 263-294.

Newey, W.K. and D.L. McFadden (1994): "Large Sample Estimation and Hypothesis Testing," in Engle, R.F. and D. L. McFadden (eds.), *Handbook of Econometrics*, vol. 4, chapter 36, 2111-2245. Amsterdam: Elsevier.

Pearson, K. (1894): "Contributions to the Mathematical Theory of Evolution," *Philosophical Transactions of the Royal Society of London*, Series A: 185, 71-78.

11

Estimating and testing dynamic asset pricing models

11.1 Asset pricing, prediction functions, and statistical inference

We develop conditions[1] ensuring the feasibility of the estimation methods when applied to general equilibrium models, that is as soon as an unobservable multidimensional process is estimated in conjunction with predictions functions suggested by standard asset pricing theories. We assume that the data generating process is a multidimensional partially observed diffusion process solution to,

$$dy\left(\tau\right)=b\left(y\left(\tau\right);\theta\right)d\tau+\Sigma\left(y\left(\tau\right);\theta\right)dW\left(\tau\right), \tag{11.1}$$

where W is a multidimensional process and (b,Σ) satisfy some regularity conditions we single out below. This appendix analyzes situations in which the original partially observed system (11.1) can be estimated by augmenting it with a number of observable deterministic functions of the state. In many situations of interest, such deterministic functions are suggested by asset pricing theories in a natural way. Typical examples include derivative asset price functions or any deterministic function(als) of asset prices (e.g., asset returns, bond yields, implied volatility, etc.). The idea to use predictions of asset pricing theories to improve the fit of models with unobservable factors is not new (see, e.g., Christensen (1992), Pastorello, Renault and Touzi (2000), Chernov and Ghysels (2000), Singleton (2001, sections 3.2 and 3.3)), and Pastorello, Patilea and Renault (2003). In this appendix, we provide a theoretical description of the mechanism leading to efficiency within the class of our estimators.

We consider a standard Markov pricing setting. For fixed $t\geq0$, we let M be the expiration date of a contingent claim with rational price process $c=\{c(y(\tau),M-\tau)\}_{\tau\in[t,M)}$, and let $\{z(y(\tau))\}_{\tau\in[t,M]}$ and $\Pi(y)$ be the associated intermediate payoff process and final payoff function, respectively. Let $\partial/\partial\tau+L$ be the usual infinitesimal generator of (11.1) taken under the risk-neutral measure. In a frictionless economy without arbitrage opportunities, c is the solution to the following partial differential equation:

$$\begin{cases} 0=\left(\dfrac{\partial}{\partial\tau}+L-R\right)c(y,M-\tau)+z(y),\ \forall(y,\tau)\in Y\times[t,M) \\ c(y,0)=\Pi(y),\ \forall y\in Y \end{cases} \tag{11.2}$$

[1] This section is based on the unpublished appendix of my joint work with Filippo Altissimo (See Altissimo and Mele (2006)).

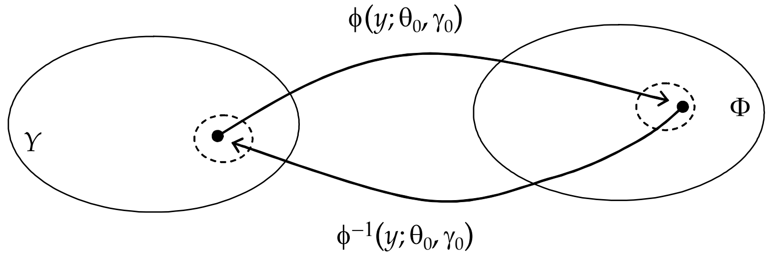


FIGURE 11.1. *Asset pricing, the Markov property, and statistical efficiency.* Y is the domain on which the partially observed primitive state process $y \equiv (y^o \;\; y^u)^\top$ takes values, Φ is the domain on which the observed system $\phi \equiv (y^o \; C(y))^\top$ takes values in Markovian economies, and $C(y)$ is a contingent claim price process in $\mathbb{R}^{d-q^*}$. Let $\phi^c = (y^o, c(y,\ell_1), \cdots, c(y,\ell_{d-q^*}))$, where $\{c(y,\ell_j)\}_{j=1}^{d-q^*}$ forms an intertemporal cohort of contingent claim prices, as in definition G.1. If *local restrictions* of ϕ are one-to-one and onto, the CD-SNE applied to ϕ^c is feasible. If ϕ is also *globally* invertible, the CD-SNE applied to ϕ^c achieves first-order asymptotic efficiency.

where $R \equiv R(y)$ is the short-term rate. We call *prediction function* any continuous and twice differentiable function $c(y; M-\tau)$ solution to the partial differential equation (11.2).

We now augment system (11.1) with $d-q^*$ prediction functions. Precisely, we let:

$$C(\tau) \equiv \left(c\left(y(\tau), M_1 - \tau\right), \cdots, c\left(y(\tau), M_{d-q^*} - \tau\right)\right), \quad \tau \in [t, M_1]$$

where $\{M_i\}_{i=1}^{d-q^*}$ is an increasing sequence of fixed maturity dates. Furthermore, we define the measurable vector valued function:

$$\phi\left(y(\tau); \theta, \gamma\right) \equiv \left(y^o(\tau), C\left(y(\tau)\right)\right), \quad \tau \in [t, M_1], \quad (\theta, \gamma) \in \Theta \times \Gamma,$$

where $\Gamma \subset \mathbb{R}^{p_\gamma}$ is a compact parameter set containing additional parameters. These new parameters arise from the change of measure leading to the pricing model (11.2), and are now part of our estimation problem.

We assume that the pricing model (11.2) is correctly specified. That is, all contingent claim prices in the economy are taken to be generated by the prediction function $c(y, M-\tau)$ for some $(\theta_0, \gamma_0) \in \Theta \times \Gamma$. For simplicity, we also consider a stylized situation in which all contingent claims have the same contractual characteristics specified by $\mathcal{C} \equiv (z, \Pi)$. More generally, one may define a series of classes of contingent claims $\{\mathcal{C}_j\}_{j=1}^{J}$, where class of contingent claims j has contractual characteristics specified by $\mathcal{C}_j \equiv (z_j, \Pi_j)$.[2] The number of prediction functions that we would introduce in this case would be equal to $d - q^* = \sum_{j=1}^{J} M^j$, where M^j is the number of prediction functions within class of assets j. To keep the presentation simple, we do not consider such a more general situation here.

Our objective is to provide estimators of the parameter vector (θ_0, γ_0) under which observations were generated. In exactly the same spirit as for the estimators considered in the main text, we want our CD-SN estimator of (θ_0, γ_0) to make the finite dimensional distributions of ϕ implied by model (11.1) and (11.2) as close as possible to their sample counterparts. Let $\Phi \subseteq \mathbb{R}^d$ be the domain on which ϕ takes values. As illustrated in Figure 11.1, our program is to move from the "unfeasible" domain Y of the original state variables in y (observables and

[2] As an example, assets belonging to class $\mathcal{C}_1$ can be European options; assets belonging to class $\mathcal{C}_1$ can be bonds; and so on.

not) to the domain Φ on which all observable variables take value. Ideally, we would like to implement such a change in domain in order to recover as much information as possible on the original unobserved process in (11.1). Clearly, ϕ is fully revealing whenever it is globally invertible. However, we will show that our methods can be implemented even when ϕ is only locally one-to-one. Further intuition on this distinction will be provided after the statement of theorem G.1 below.

An important feature of the theory in this appendix is that it does not hinge upon the availability of contingent prices data covering the same sample period covered by the observables in (11.1). First, the price of a given contingent claim is typically not available for a long sample period. As an example, available option data often include option prices with a life span smaller than the usual sample span of the underlying asset prices; in contrast, it is common to observe long time series of option prices having the same maturity. Second, the price of a single contingent claim depends on time-to-maturity of the claim; therefore, it does not satisfy the stationarity assumptions maintained in this paper. To address these issues, we deal with data on assets having the same characteristics at each point in time. Precisely, consider the data generated by the following random processes:

Definition G.1. (Intertertemporal (ℓ, N)-cohort of contingent claim prices) *Given a prediction function $c(y; M-\tau)$ and a N-dimensional vector $\ell \equiv (\ell_1, \cdots, \ell_N)$ of fixed maturities, an* intertemporal (ℓ, N)-cohort of contingent claim prices *is any collection of contingent claim price processes $c(\tau, \ell) \equiv (c(y(\tau), \ell_1), \cdots, c(y(\tau), \ell_N))$ $(\tau \geq 0)$ generated by the pricing model* (11.2).

Consider for example a sample realization of three-months at-the-money option prices, or a sample realization of six-months zero-coupon bond prices. Long sequences such as the ones in these examples are common to observe. If these sequences were generated by (11.2), as in definition G.1, they would be deterministic functions of y, and hence stationary. We now develop conditions ensuring both feasibility and first-order efficiency of the CD-SNE procedure as applied to this kind of data. Let $\bar{a}$ denote the matrix having the first q^* rows of Σ, where a is the diffusion matrix in (G1). Let ∇C denote the Jacobian of C with respect to y. We have:

Theorem 11.1. (Asset pricing and Cramer-Rao lower bound) *Suppose to observe an intertemporal $(\ell, d-q^*)$-cohort of contingent claim prices $c(\tau, \ell)$, and that there exist prediction functions C in $\mathbb{R}^{d-q^*}$ with the property that for $\theta = \theta_0$ and $\gamma = \gamma_0$,*

$$\begin{pmatrix} \bar{a}(\tau) \cdot \Sigma(\tau)^{-1} \\ \nabla C(\tau) \end{pmatrix} \neq 0, \quad P \otimes d\tau\text{-}a.s. \quad all\ \tau \in [t, t+1], \tag{G3}$$

where C satisfies the initial condition $C(t) = c(t, \ell) \equiv (c(y(t), \ell_1), \cdots, c(y(t), \ell_{d-q^*}))$. Let $(z, v) \equiv (\phi_t^c, \phi_{t-1}^c)$, where $\phi_t^c = (y^o(t), c(y(t), \ell_1), \cdots, c(y(t), \ell_{d-q^*}))$. Then, under the assumptions in theorem 3, the CD-SNE has the same properties as in theorem 2, with the variance terms being taken with respect to the fields generated by ϕ_t^c. Finally, suppose that ϕ_t^c is Markov, and set $w_T(z, v) = \pi_T(z)^2 / \pi_T(z, v) \mathbb{T}_{T,\alpha}(z, v)$. Then, the CD-SNE attains the Cramer-Rao lower bound (with respect to the fields generated by ϕ_t^c) as $S \to \infty$.

According to Theorem 11.1, our CD-SNE is feasible whenever ϕ is locally invertible for a time span equal to the sampling interval. As Figure 3 illustrates, condition (G3) is satisfied whenever ϕ is locally one-to-one and onto.[3] If ϕ is also globally invertible for the same time span, ϕ^c is Markov. The last part of this theorem then says that in this case, the CD-SNE is asymptotically efficient. We emphasize that such an efficiency result is simply about first-order efficiency in the *joint* estimation of θ *and* γ given the observations on ϕ^c. We are not claiming that our estimator is first-order efficient in the estimation of θ in the case in which y is fully observable.

Naturally, condition (G3) does not ensure that ϕ is globally one-to-one and onto. In other terms, ϕ might have many locally invertible restrictions.[4] In practice, ϕ might fail to be globally invertible because monotonicity properties of ϕ may break down in multidimensional diffusion models. In models with stochastic volatility, for example, option prices can be decreasing in the underlying asset price (see Bergman, Grundy and Wiener (1996)); and in the corresponding stochastic volatility yield curve models, medium-long term bond prices can be increasing in the short-term rate (see Mele (2003)). Intuitively, these pathologies may arise because there is no guarantee that the solution to a stochastic differential system is nondecreasing in the initial condition of one if its components - as it is instead the case in the scalar case.

When all components of vector y^o represent the prices of assets actively traded in frictionless markets, (G3) corresponds to a condition ensuring market completeness in the sense of Harrison and Pliska (1983). As an example, condition (G3) for Heston's (1993) model is $\partial c/\partial\sigma \neq 0$ $P \otimes d\tau$-a.s, where σ denotes instantaneous volatility of the price process. This condition is satisfied by the Heston's model. In fact, Romano and Touzi (1997) showed that within a fairly general class of stochastic volatility models, option prices are *always* strictly increasing in σ whenever they are convex in Q. Theorem G.1 can be used to implement efficient estimators in other complex multidimensional models. Consider for example a three-factor model of the yield curve. Consider a state-vector (r, σ, ℓ), where r is the short-term rate and σ, ℓ are additional factors (such as, say, instantaneous short-term rate volatility and a central tendency factor). Let $u^{(i)} = u\left(r(\tau), \sigma(\tau), \ell(\tau); M_i - \tau\right)$ be the time τ rational price of a pure discount bond expiring at $M_i \geq \tau$, $i = 1, 2$, and take $M_1 < M_2$. Let $\phi \equiv (r, u^{(1)}, u^{(2)})$. Condition (G3) for this model is then,

$$u_\sigma^{(1)} u_\ell^{(2)} - u_\ell^{(1)} u_\sigma^{(2)} \neq 0, \quad P \otimes dt\text{-a.s.} \quad \tau \in [t, t+1], \tag{G4}$$

where subscripts denote partial derivatives. It is easily checked that this same condition must be satisfied by models with correlated Brownian motions and by yet more general models. Classes of models of the short-term rate for which condition (G4) holds are more intricate to identify than in the European option pricing literature mentioned above (see Mele (2003)).

11.2 Term structure models

Let $r\left(t\right)$ be the short-term rate process, solution to the following stochastic differential equation,

$$dr\left(t\right) = \kappa\left(\mu - r\left(t\right)\right)dt + \sqrt{v\left(t\right)}r\left(t\right)^\eta\, dW\left(t\right), \quad t \geq 0, \tag{11.3}$$

where $W\left(t\right)$ is a standard Brownian motion, and κ, μ and η are three positive constants. Suppose, also, that the instantaneous volatility process $\sqrt{v\left(t\right)}r\left(t\right)^\eta$ is such that $v\left(t\right)$ is solution

[3] Local invertibility of ϕ means that for every $y \in Y$, there exists an open set Y_* containing y such that the restriction of ϕ to Y_* is invertible. And ϕ is locally invertible on Y_* if $\det J\phi \neq 0$ (where $J\phi$ is the Jacobian of ϕ), which is condition (G3).

[4] As an example, consider the mapping $\mathbb{R}^2 \mapsto \mathbb{R}^2$ defined as $\phi(y_1, y_2) = (e^{y_1}\cos y_2, e^{y_1}\sin y_2)$. The Jacobian satisfies $\det J\phi(y_1, y_2) = e^{2y_1}$, yet ϕ is 2π-periodic with respect to y_2. For example, $\phi(0, 2\pi) = \phi(0, 0)$.

to,

$$dv(t) = \beta(\alpha - v(t))\,dt + \xi v(t)^{\vartheta}\left(\rho dW(t) + \sqrt{1-\rho^2}dU(t)\right), \quad t \geq 0, \tag{11.4}$$

where $U(t)$ is another standard Brownian motion; β, α, ξ and ϑ are four positive constants, and ρ is a constant such that $|\rho| < 1$.

11.2.1 The level effect

Which empirical regularities would the short-term rate model in Eqs. (11.3)-(11.4) address? Which sign of the correlation coefficient ρ would be consistent with historical episodes such as the Monetary Experiment of the Federal Reserve System between October 1979 and October 1982?

The short-term rate model in Eqs. (11.3)-(11.4) would address two empirical regularities.

1) The volatility of the short-term rate is not constant over time. Rather, it seems to be driven by an additional source of randomness. All in all, the short-term process seems to be generated by the stochastic volatility model in Eqs. (11.3)-(11.4), in which the volatility component $v(t)$ is driven by a source of randomness only partially correlated with the source of randomness driving the short-term rate process itself.

2) The volatility of the short-term rate is increasing in the level of the short-term rate. This phenomenon is known as the "level effect". Perhaps, periods of high interest rates arise because of erratic liquidity. (Erratic liquidity would command a high risk-premium and so a high LIBOR rate say.) But precisely because of erratic liquidity, interest rates are also very volatile. The short-term rate model in Eqs. (11.3)-(11.4) is a very useful reduced form able to capture these effects through the two parameters: η and ρ. If the parameter η is greater than zero, the instantaneous interest rate volatility increases with the level of the interest level. If the "correlation" coefficient $\rho > 0$, the interest rate *volatility* is also partly related to the sources of interest rate volatility not directly related to the *level* of the interest rate.

During the Monetary Experiment, the FED decided to target money supply, rather than interest rates. So the high volatility of money demand mechanically translated to high interest rate volatility as a result of the market clearing. Moreover, the quantity of monetary base was kept deliberately low - to fight against inflation. So the US experienced both high interest rate volatility and high interest rates (see for example Andersen and Lund (1997, J. Econometrics) for an empirical study). There is no empirical study about the issues related to the sign of the correlation coefficient ρ. Here is a possible argument. A rolling window estimation suggestive that the level of ρ changed a lot around the Monetary Experiment would mean that the bulk of interest rate volatility was not entirely due to the mechanical effects related to the FED behavior.

11.2.2 The simplest estimation case

Next, suppose we wish to estimate the parameter vector $\theta = [\kappa, \mu, \eta, \beta, \alpha, \xi, \vartheta, \rho]^{\top}$ of the model in Eqs. (11.3)-(11.4). Under which circumstances would Maximum Likelihood be a feasible estimation method?

The ML estimator would be feasible under two sets of conditions. *First*, the model in Eqs. (11.3)-(11.4) should not have stochastic volatility at all, viz, $\beta = \xi = 0$; in this case, the short-term rate would be solution to,

$$dr(t) = \kappa(\mu - r(t))\,dt + \bar{\sigma} r(t)^{\eta}\,dW(t), \quad t \geq 0,$$

where $\bar{\sigma}$ is now a constant. *Second*, the value of the elasticity parameter η is important. If $\eta = 0$, the short-term rate process is the Gaussian one proposed by Vasicek (1977, J. Fin. Economics). If $\eta = \frac{1}{2}$, we obtain the square-root process introduced in Financial Economics by Cox, Ingersoll and Ross (1985, Econometrica) (CIR henceforth). In the Vasicek case, the transition density of r is Gaussian, and in the CIR case, the transition density of r is a noncentral chi-square. So in both the Vasicek and CIR, we may write down the likelihood function of the diffusion process. Therefore, ML estimation is possible in these two cases.

In the more general case, we have to go for simulation methods [...]

11.2.3 More general models

Estimating the model in Eqs. (11.3)-(11.4) is certainly instructive. Yet a more important question is to examine the term-structure implications of this model. More generally, how would the estimation procedure outlined in the previous subsection change if the task is to estimate a Markov model of the term-structure of interest rates? There are three steps.

11.2.3.1 Step 1

Collect data on the term structure of interest rates. We will need to use data on two maturities (say a time series of riskless 6 months and 5 years interest rates).

11.2.3.2 Step 2

Let us consider a model of the entire term-structure of interest rates. By the fundamental theorem of asset pricing, and the Markov property of the diffusion, the price of a riskless bond predicted by the model is,

$$u^j\left(r\left(t\right),v\left(t\right)\right) \equiv u\left(r\left(t\right),v\left(t\right),N_j - t\right) = \mathbb{E}\left(e^{-\int_t^{N_j} r(s)ds} \middle| r\left(t\right),v\left(t\right)\right), \tag{11.5}$$

where $\mathbb{E}\left(.\right)$ is the conditional expectation taken under the risk-neutral probability, and N_j is a sequence of expiration dates. Naturally, the previous formula relies on some assumptions about risk-aversion correction. (Some of these assumptions may be of a reduced-form nature; others may rely on the specification of preferences, beliefs, markets and technology. But we do not need to be more precise at this level of generality. In turn, these assumptions entail that the pricing formula in Eq. (11.5) depends on some additional risk-adjustment parameter vector, say λ. Precisely, the Radon-Nykodim derivative of the risk-neutral probability with respect to the physical probability is given by $\exp\left(-\frac{1}{2}\int\left\|\Lambda\left(t\right)\right\|^2 dt - \int\Lambda\left(t\right)dZ\left(t\right)\right)$, where $Z = \left[W\ U\right]^\top$, W and U are the two Brownian motions in Eqs. (11.3)-(11.4), and $\Lambda\left(t\right)$ is some process adapted to Z, which is taken to be of the form $\Lambda\left(t\right) \equiv \Lambda^{\mathrm{m}}\left(r\left(t\right),v\left(t\right);\lambda\right)$, for some vector valued function Λ^{m} and some parameter vector λ. The function Λ^{m} makes risk-adjustment corrections dependent on the current value of the state vector $\left(r\left(t\right),v\left(t\right)\right)$, and thus makes the model Markov.

So the estimation problem is actually one in which we have to estimate both the "physical" parameter vector $\theta = \left[\kappa,\mu,\eta,\beta,\alpha,\xi,\vartheta,\rho\right]^\top$ and the "risk-adjustment" parameter vector λ.

Next, compute interest rates corresponding to two maturities,

$$R^j\left(r\left(t\right),v\left(t\right);\theta,\lambda\right) = -\frac{1}{N_j}\log u^j\left(r\left(t\right),v\left(t\right)\right),\quad j = 1,2, \tag{11.6}$$

where the bond prices are computed through Eq. (11.5), and where the notation $R^j\left(r,v;\theta,\lambda\right)$ emphasizes that the theoretical term-structure depends on the parameter vector $\left(\theta,\lambda\right)$. We can now use the data ($R^j_\$$ say) and the model predictions about the data (R^j), create moment conditions, and proceed to estimate the parameter vector (θ, λ) through some method of moments (provided the moments are enough to make (θ, λ) identifiable). But there are two difficulties. *First*, the volatility process $v(t)$ is not observable by the econometrician. *Second*, the bond pricing formula in Eq. (11.5) does not generally admit a closed-form.

The first difficulty can be overcome through inference methods based on simulations. Here is an outline of these methods that could be used here. Simulate the system in Eqs. (11.3)-(11.4) for a given value of the parameter vector (θ, λ). For each simulation, compute a time series of interest rates R^j from Eq. (11.6). Use these simulated data to create moment conditions. The parameter estimator is the value of (θ, λ) which minimizes some norm of these moment conditions obtained through the simulations.

The next step discusses how to address the second difficulty.

11.2.3.3 Step 3

The use of *affine* models would considerably simplify the analysis. Affine models place restrictions on the data generating process in Eqs. (11.3)-(11.4) and in the risk-aversion corrections in Eq. (11.5) in such a way that the term structure in Eq. (11.6) is,

$$R^j\left(r\left(t\right), v\left(t\right); \theta, \lambda\right) = A\left(j; \theta, \lambda\right) + \mathbf{B}\left(j; \theta, \lambda\right) \cdot \mathbf{y}\left(t\right), \quad j = 1, 2,$$

where $A(j; \theta, \lambda)$ and $\mathbf{B}(j; \theta, \lambda)$ are some functions of the maturity N_j ($\mathbf{B}$ is vector valued), and generally depend on the parameter vector (θ, λ); and finally the state vector $\mathbf{y} = [r\ v]^\top$. (Namely, an affine model obtains once $\eta = 0$, $\vartheta = \frac{1}{2}$, and the function Λ^{m} is affine.) So once Eqs. (11.3)-(11.4) are simulated, the computation of a time series of interest rates R^j is straight forward.

11.3 Appendix

Proof of Theorem 1.11. Let $\pi_t \equiv \pi_t\left(\phi\left(y(t+1), \mathbf{M}-(t+1)\mathbf{1}_{d-q^*}\right) | \, \phi\left(y(t), \mathbf{M}-t\mathbf{1}_{d-q^*}\right)\right)$ denote the transition density of

$$\phi\left(y(t), \mathbf{M}-t\mathbf{1}_{d-q^*}\right) \equiv \phi\left(y(t)\right) \equiv \left(y^o(t), c(y(t), M_1-t), \cdots, c(y(t), M_{d-q^*}-t)\right),$$

where we have emphasized the dependence of ϕ on the time-to-expiration vector:

$$\mathbf{M}-t\mathbf{1}_{d-q^*} \equiv (M_1-t, \cdots, M_{d-q^*}-t).$$

By $\Sigma(\tau)$ full rank $P \otimes d\tau$-a.s., and Itô's lemma, ϕ satisfies, for $\tau \in [t, t+1]$,

$$\begin{cases} dy^o(\tau) &= b^o(\tau)d\tau + F(\tau)\Sigma(\tau)dW(\tau) \\ dc(\tau) &= b^c(\tau)d\tau + \nabla c(\tau)\Sigma(\tau)dW(\tau) \end{cases}$$

where b^o and b^c are, respectively, q^*-dimensional and $(d-q^*)$-dimensional measurable functions, and $F(\tau) \equiv \bar{a}(\tau)\cdot\Sigma(\tau)^{-1}$ $P\otimes d\tau$-a.s. Under condition (G3), π_t is not degenerate. Furthermore, $C\left(y(t); \ell\right) \equiv C(t)$ is deterministic in $\ell \equiv (\ell_1, \cdots, \ell_{d-q^*})$. That is, for all $(\bar{c}, \bar{c}^+) \in \mathbb{R}^d \times \mathbb{R}^d$, there exists a function μ such that for any neighbourhood $N(\bar{c}^+)$ of $\bar{c}^+$, there exists another neighborhood $N(\mu(\bar{c}^+))$ of $\mu(\bar{c}^+)$ such that,

$$\begin{aligned} & \left\{\omega \in \Omega : \phi\left(y(t+1), \mathbf{M}-(t+1)\mathbf{1}_{d-q^*}\right) \in N(\bar{c}^+) \middle| \ \phi\left(y(t), \mathbf{M}-t\mathbf{1}_{d-q^*}\right) = \bar{c}\right\} \\ = \ & \left\{\omega \in \Omega : (y^o(t+1), c(y(t+1), M_1-t)), \cdots, c(y(t+1), M_{d-q^*}-t)) \in N(\mu(\bar{c}^+)) \right. \\ & \qquad \left. |\phi\left(y(t), \mathbf{M}-t\mathbf{1}_{d-q^*}\right) = \bar{c}\right\} \\ = \ & \left\{\omega \in \Omega : (y^o(t+1), c(y(t+1), M_1-t)), \cdots, c(y(t+1), M_{d-q^*}-t)) \in N(\mu(\bar{c}^+)) \right. \\ & \qquad \left. |(y^o(t), c(y(t), M_1-t), \cdots, c(y(t), M_{d-q^*}-t)) = \bar{c}\right\} \end{aligned}$$

where the last equality follows by the definition of ϕ. In particular, the transition laws of ϕ_t^c given ϕ_{t-1}^c are not degenerate; and ϕ_t^c is stationary. The feasibility of the CD-SNE is proved. The efficiency claim follows by the Markov property of ϕ, and the usual score martingale difference argument. ■

References

Altissimo, F. and A. Mele (2006). "Simulated Nonparametric Estimation of Dynamic Models."

Arnold, L., 1992, *Stochastic Differential Equations*: *Theory and Applications.* Malabar, Florida: Krieger Publishing Company.

Bergman, Y. Z., B. D. Grundy, and Z. Wiener, 1996, "General Properties of Option Prices," *Journal of Finance*, 51, 1573-1610.

Chernov, M. and E. Ghysels, 2000, "A Study towards a Unified Approach to the Joint Estimation of Objective and Risk-Neutral Measures for the Purpose of Options Valuation," *Journal of Financial Economics*, 56, 407-458.

Christensen, B. J., 1992, "Asset Prices and the Empirical Martingale Model," working paper, New York University.

Harrison, J. M. and S. R. Pliska, 1983, "A Stochastic Calculus Model of Continuous Trading: Complete Markets," *Stochastic Processes and their Applications* 15, 313-316.

Heston, S., 1993, "A Closed-Form Solution for Options with Stochastic Volatility with Applications to Bond and Currency Options," *Review of Financial Studies*, 6, 327-343.

Karatzas, I. and S. Shreve, 1991, *Brownian Motion and Stochastic Calculus*, Berlin: Springer Verlag.

Kloeden, P.E. and E. Platen, 1999, *Numerical Solutions of Stochastic Differential Equations*, Berlin: Springer Verlag.

Mele, A., 2003, "Fundamental Properties of Bond Prices in Models of the Short-Term Rate," *Review of Financial Studies*, 16, 679-716.

Pastorello, S., E. Renault and N. Touzi, 2000, "Statistical Inference for Random-Variance Option Pricing," *Journal of Business and Economic Statistics* 18, 358-367.

Pastorello, S., V. Patilea, and E. Renault, 2003, "Iterative and Recursive Estimation in Structural Non Adaptive Models," *Journal of Business and Economic Statistics*, 21, 449-509.

Romano, M. and N. Touzi, 1997, "Contingent Claims and Market Completeness in a Stochastic Volatility Model," *Mathematical Finance*, 7, 399-412.

Singleton, K.J., 2001, "Estimation of Affine Asset Pricing Models Using the Empirical Characteristic Function," *Journal of Econometrics*, 102, 111-141.

Appendixes

Mathematical appendix

This appendix provides the reader with a few concepts used throughout the *Lectures*.

A.1 Foundational issues in probability theory

A.1.1 Heuristic considerations

When we try to imagine a dynamic representation of the world, we rapidly crash into the formidable difficulties that our mind encounters in exactly extrapolating future events. In the previous sections, we tried to overcome such difficulties by introducing two-period economies with two and/or more states of nature. The resulting model naturally led to the notion of rational expectations. Now we would like to face a structure of events more complex than the structure of events considered in the previous chapter. The task of this section is to provide some elements corresponding to such a need. The elements we shall be handling here are essentially the ones that are needed for a rigorous treatment of probability theory. Kolmogorov (1993)[5] was the first to organize such a theory by linking the concept of probability with the concept of measure in the integration theory.

The starting point is what is commonly referred to as a *Nature Experiment*, denoted as $\mathcal{E}$, whose outcome is obviously not known a priori. $\mathcal{E}$ can be interpreted as the result of tossing a coin, or as the quantity of pencils that will be produced in the US in the next ten years. Then we introduce a state space Ω, which represents the set of all possible outcomes of $\mathcal{E}$. As an example, if $\mathcal{E}$ consists in tossing twice a coin, the outcomes are: HT, HH, TH, TT, and $\Omega \equiv \{\text{HT,HH,TH,TT}\}$. These specific outcomes are also termed as *elementary events* or *atoms*, i.e. they can not be expressed as the logical sum of any other event that can be generated by $\mathcal{E}$. It is clear, however, that we may wish to be interested in more complex events than simply elementary events as for instance, {HT,HH} ("H at the first tossing"), {HT,HH,TH} ("at least once H") etc... To take non-elementary events into account, we introduce a collection of parts of Ω that exactly includes the pieces of information that are of interest to us. Such a collection is usually referred to as Boole's algebra on Ω, or tribe, or σ-algebra on Ω (according to the case of

[5] Kolmogorov, A. (1933): *Foundations of the Theory of Probability* (in Russian). Published in English in 1950: New York, Chelsea.

pertinence: see below for further details) and will have to satisfy some fundamental properties. The very fundamental property that has to be satisfied by an algebra is that at each time that we apply a complementation, union and intersection operation on its elements, the result must be another element that also belongs to the same algebra.

It turns out that the previous stability property of an algebra is essential on an axiomatic standpoint at the very starting point of probability theory. Indeed, the fact that an event will take place or not necessarily implies the realization of other events (e.g., the elementary events), and the definition domain of probabilities must carefully integrate such an aspect. In other terms, an algebra must be built up in such a way that one must always be able to say if any event (elementary or not) has taken place or not after observing $\mathcal{E}$.

Let us suppose for instance that Ω has a finite number of elements. A natural choice of an algebra could be the set of all partitions of Ω, denoted as 2^{Ω}. Such a notation is also natural because the number of its elements is just $2^{\#\Omega}$, where $\#\Omega$ is the number of elements of Ω. For example, if $\Omega = \{a, b, c\}$, then $2^{\Omega} = \{\emptyset, \{a\}, \{b\}, \{c\}, \{a, b\}, \{a, c\}, \{b, c\}, \Omega\}$. (The example of tossing a coin twice is similar but more cumbersome to describe since it involves $2^4 = 16$ elements.) Such an example shows that it is natural to introduce a "probability" P on 2^{Ω}, that has the property that $P(\emptyset) = 0$, $P(\Omega) = 1$, etc... This is possible because 2^{Ω} is stable with respect to the operations of complementation, union and intersection.

Choosing 2^{Ω}, however, does not always lead to a well-posed algebra: sometimes 2^{Ω} can include unuseful pieces of information or, simply, it can be ill-defined: as an example, Ω can be uncountable because $\Omega = \mathbb{R}^d$. In such a case, the construction strategy will be "minimal", and will consist in taking into account the smallest algebra containing the elements of interest.

Next section presents the standard setup that one typically needs to address these issues. Such a setup allows one to take account of complex operations consisting, for instance, in "transferring" purely qualitative events (e.g., a Central Bank allied or not with Government) on to a space of purely quantitative events (the real line, for instance): in this last space, an event may be built-up via the parts constituting the original space. It is exactly such an idea that underlies the general equilibrium representation of the financial markets made by Arrow (cf. the previous chapter): after all, a financial asset can be thought as being a random variable defined on states of the world. Such an approach, coupled with the rational expectations assumption, also permits to model the dynamics of a given economy in the modern approach of macroeconomic fluctuations. Such an approach, too, is, directly or indirectly, an intellectual extension of the original contribution of Arrow. Of course, this is also the case of the models developed by Lucas and Prescott at the very beginning of the 1970s, not to say the dynamic financial models formulated by Black and Scholes (1973) and Merton (1973) (cf. the next chapter) or the models exhibiting the existence of non-informative equilibria (e.g.: observation of the same equilibrium price in situations where there were mutually exclusive portions of the state-space).

A.1.2 The reference setup

We introduce a *measurable space* associated with $\mathcal{E}$: $(\Omega, \mathcal{F}, P)$. $(\Omega, \mathcal{F}, P)$ will be referred to as *probability space* if P is a probability measure—see below for the definition of a probability measure—. Here Ω is the set (or space) of events and $\mathcal{F}$ is *a* tribe on Ω, or *a* σ-algebra on Ω (or simply *a* σ-field) satisfying the following properties:

i) $\Omega \in \mathcal{F}$;

ii) $A \in \mathcal{F} \Rightarrow A^c \in \mathcal{F}$, where $A^c = \Omega \backslash A$ is the complement of A in Ω;

iii) $A_i \in \mathcal{F}$, $i = 1, 2, \cdots \Rightarrow \bigcup_{i=1}^{\infty} A_i \in \mathcal{F}$, where $\{A_i\}$ is a countable sequence of disjoint events.[6]

Next, let $\mathcal{G}$ a subset of the set of parts of Ω. We say that the tribe $\sigma(\mathcal{G})$ generated by $\mathcal{G}$ is the intersection of all tribes containing $\mathcal{G}$. $\sigma(\mathcal{G})$ can thus be conceived as the smallest tribe containing $\mathcal{G}$. We call Borel tribe, and note it as $\mathcal{B}(\mathbb{R}^d)$, the tribe generated by the open sets of $\mathbb{R}^d$.

P is a measure on $(\Omega, \mathcal{F})$ if it is a function s.t. $\mathcal{F} \mapsto [0, \infty]$ that satisfies the condition $P\left(\bigcup_{i=1}^{\infty} A_i\right) = \sum_{i=1}^{\infty} P(A_i)$ for all countable sequences of disjoints sets $A_i \in \mathcal{F}$. Such a property is known as the σ-additivity property. The mass of a measure is equal to $P(\Omega)$. $(\Omega, \mathcal{F}, P)$ is a *probability space* if P is a probability measure on $(\Omega, \mathcal{F})$, id est if it is a function s.t. $\mathcal{F} \mapsto [0, 1]$, if it is σ-additive, and if it satisfies $P(\emptyset) = 0$ and $P(\Omega) = 1$ (unit mass).

Given a measurable space $(\Omega, \mathcal{F}, P)$, we say that a function $y : \Omega \mapsto \mathbb{R}^d$ is $\mathcal{F}$-measurable if:

$$\forall B \in \mathcal{B}(\mathbb{R}^d), \quad y^{-1}(B) \equiv \{\omega \in \Omega \ / \ y(\omega) \in B\} \in \mathcal{F}.$$

Given a probability space, a real-valued random variable is a function $\Omega \mapsto \mathbb{R}^d$ which is $\mathcal{F}$-measurable. Note that while originally defined on such "amorphic" spaces as $(\Omega, \mathcal{F}, P)$, random variables are subsequently redefined on spaces having the form $(\mathbb{R}^d, \mathcal{B}(\mathbb{R}^d), P_y)$ where:

$$P_y(B) = P\left(y^{-1}(B)\right), \quad \forall B \in \mathcal{B}(\mathbb{R}^d),$$

is a probability measure on $(\mathbb{R}^d, \mathcal{B}(\mathbb{R}^d))$. ($P_y$ is also referred to as *distribution* of y.) This is exactly the "transfer" method mentioned towards the end of the previous section. We shall note:

$$E(y) \equiv \int_{\Omega} y(\omega) dP(\omega) = \int_{\mathbb{R}^d} y dP_y(y). \tag{11.7}$$

A stochastic process in $\mathbb{R}^d$ is a family of random variables—id est they are originally defined on $(\Omega, \mathcal{F})$ and then redefined on $(\mathbb{R}^d, \mathcal{B}(\mathbb{R}^d))$—that are indexed by time: $\{x_t\}_{t \in \mathbb{T}}$; here $\mathbb{T}$ can be $[0, \infty)$ but also $[0, T]$, $T < \infty$ or $\mathbb{N}$. Considered as a "function" of the points $\omega \in \Omega$ for a fixed $t_* \in \mathbb{T}$, the mapping $\omega \mapsto x_{t_*}(\omega)$ is a random variable; considered as a "function" of $t \in \mathbb{T}$ for a fixed point $\omega_* \in \Omega$, the mapping $t \mapsto x_t(\omega_*)$ is the trajectory of the stochastic process corresponding to the realization of ω_*. In other terms, the set $\{(t, \omega) \ / \ x_t(\omega) \in B\}$, $\forall B \in \mathcal{B}(\mathbb{R}^d)$, belongs to the product σ-algebra $\mathcal{B}(\mathbb{T}) \otimes \mathcal{F}$, where $\mathcal{B}(\mathbb{T})$ is the Borel tribe generated by $\mathbb{T}$: this is the equivalent of the measurability property of a random variable, and in fact we say that in this case, $\{x_t\}_{t \in \mathbb{T}}$ is *jointly* $\mathcal{B}(\mathbb{T}) \otimes \mathcal{F}$-measurable to mean that it is measurable with respect to the smallest σ-algebra on $\mathbb{T} \times \Omega$ containing the sets of the form $B \times F$, where $B \in \mathcal{B}(\mathbb{T})$ and $F \in \mathcal{F}$.

Economic reasons (*not* mathematical reasons) often make it desirable to assume that the systems presenting an interest to us are those that are not subject to loss of memory and that are such that the unfolding of time reveals new pieces of information. To meet such needs, $(\Omega, \mathcal{F})$ is endowed with a *filtration*, id est a family of nondecreasing σ-algebras $\{\mathcal{F}_t\}_{t \in \mathbb{T}}$ of $\mathcal{F}$, id est such that $\mathcal{F}_t \subseteq \mathcal{F}_s \subseteq \mathcal{F}$, for $0 \leq t < s$. As an example, if y represents the equilibrium price

[6] One obtains a *Boole algebra* by replacing such a third condition with a property of stability of finite unions.

vector of d financial assets, in the first financial models the information system of reference was built-up on a filtration of the form:

$$\mathcal{F}_t^y = \sigma(y_s, 0 \leq t \leq s),$$

id est the smallest σ-algebra with respect to which y_s is measurable for all $s \in [0, t]$;[7] these asset prices could then be able to guarantee the full spanning of any $\mathcal{F}_T$-measurable ($0 < T < \infty$) random variable, which meant that markets are *complete*. As explained in the previous chapter, models with *incomplete* markets seem to be more realistic: in these models, agents face the realization of $\mathcal{F}_T^+$-measurable random variables, where the filtration $\{\mathcal{F}_t^+\}_{t\in\mathbb{T}}$ now is bigger that $\{\mathcal{F}_t\}_{t\in\mathbb{T}}$: cf. chapters 4 and 5 for more precise details.

Finally, we say that the stochastic process $\{x_t\}_{t\in\mathcal{T}}$ is adapted to the filtration $\{\mathcal{F}_t\}_{t\in\mathbb{T}}$, or $\mathcal{F}$-adapted, if x_t is a $\mathcal{F}_t$-measurable random variable. For all $s \geq t$, we shall denote as $E(y_s/\mathcal{F}_t)$ the conditional expectation of y_s given $\mathcal{F}_t$, id est the expectation of y_s, formally defined as in (3.8), but taken under the conditional probability defined as:

$$P\left\{y_s \leq y^+ \ / \ \mathcal{F}_t\right\} = \frac{P\left(\{\omega : y_s(\omega) \leq y^+\} \cap \mathcal{F}_t\right)}{P\left\{\mathcal{F}_t\right\}}.$$

We shall also have the occasion to define conditional probability of the form $P\{y_s \leq y^+ / y_t = y\}$, and when the support is continuous, with an abuse of notation we shall continue to write $P\{y_s \leq y^+ / y_t = y\}$ or simply $P\{y^+ \ / \ y\}$ instead of the more precise notation $P\{y_s \leq y^+ / y_t = I_y\}$, where I_y is an arbitrarily small neighborhood of y.

A.2 Stochastic calculus

A.3 Contraction theorem

This is a result that can also be used to establish stability of a discrete time system, although the typical use economists make of it is made to produce results for problems arising in dynamic programming.

Definition 4A.1 (Metric Space). ... *and the triangle inequality*: $d(a, b) \leq d(a, c) + d(c, b)$.

Definition 4A.2 (Cauchy sequence). *Let* $\{x_n\}$ *be a sequence of points in a metric space* $\mathcal{M}$, *and suppose that* $d(x_n, x_m) \to_{\substack{n\to\infty \\ m\to\infty}} 0$. *Then* $\{x_n\}$ *is a Cauchy's sequence.*

If $\mathcal{M} = \mathbb{R}$, all Cauchy's sequences are convergent, and the metric spaces displaying such a property are said to be *complete*.

Definition 4A.3 (Contraction). *If* $\forall x, y \in \mathcal{M} : d(f(x), f(y)) \leq kd(x, y)$, $k < 1$, *then* f *is a* k*-contraction.*

Theorem 4A.3. *Let* $\mathcal{M}$ *be a complete metric space and* f *a* k*-contraction of* $\mathcal{M}$ *into itself. Then* f *admits one and only one fixed point.*

[7] The usual interpretation of an event $A \in \mathcal{F}_t^y$ is that at time t, we must be able to say whether A actually took place or not.

PROOF. Let $x_0 \in \mathcal{M}$ arbitrary, and let $x_n = f(x_{n-1})$. Since f is a k-contraction,

$$\begin{cases} d(x_n, x_{n+1}) &= d(f(x_{n-1}), f(x_n)) \le kd(x_{n-1}, x_n) \\ d(x_{n-1}, x_n) &\le kd(x_{n-2}, x_{n-1}) \end{cases}$$

whence:

$$kd(x_{n-1}, x_n) \le k^2 d(x_{n-2}, x_{n-1}).$$

By replacing the second inequality into the first,

$$d(x_n, x_{n+1}) \le k^2 d(x_{n-2}, x_{n-1}),$$

and in general,

$$d(x_n, x_{n+1}) \le k^n d(x_0, x_1).$$

By the triangle inequality,

$$d(x_n, x_{n+2}) \le d(x_n, x_{n+1}) + d(x_{n+1}, x_{n+2}),$$

and

$$d(x_{n+2}, x_{n+4}) \le d(x_{n+2}, x_{n+3}) + d(x_{n+3}, x_{n+4}).$$

By combining these inequalities,

$$d(x_n, x_{n+2}) + d(x_{n+2}, x_{n+4}) \le d(x_n, x_{n+1}) + d(x_{n+1}, x_{n+2}) + d(x_{n+2}, x_{n+3}) + d(x_{n+3}, x_{n+4}),$$

and since $d(x_n, x_{n+4}) \le d(x_n, x_{n+2}) + d(x_{n+2}, x_{n+4})$,

$$d(x_n, x_{n+4}) \le \sum_{i=n}^{n+3} d(x_i, x_{i+1}),$$

In general we have:

$$d(x_n, x_{n+p}) \le \sum_{i=n}^{n+p-1} d(x_i, x_{i+1}),$$

Since $d(x_i, x_{i+1}) \le k^i d(x_0, x_1)$,

$$d(x_n, x_{n+p}) \le d(x_0, x_1) \sum_{i=n}^{n+p-1} k^i = d(x_0, x_1) \cdot \frac{(1-k^p)k^n}{1-k}.$$

Since $k < 1$, $\frac{(1-k^p)k^n}{1-k} \to_{\substack{p\to\infty \\ n\to\infty}} 0$, and the previous inequality then reveals that $d(x_n, x_{n+p}) \to_{\substack{p\to\infty \\ n\to\infty}} 0$. Therefore, $\{x_n\} \to_{n\to\infty} \overline{x}$, say. It's a Cauchy's sequence.

Now we are in a position to show that $\bar{x} : \bar{x} = f(\bar{x})$.

By the triangle inequality,

$$d(\bar{x}, f(\bar{x})) \le d(\bar{x}, x_{n+1}) + d(x_{n+1}, f(\bar{x})) = d(\bar{x}, x_{n+1}) + d(f(x_n), f(\bar{x})).$$

Since $\{x_n\} \to_{n\to\infty} \bar{x}$ and f is continuous, $\forall \epsilon > 0$, $\exists n_\epsilon > 0 : d(\bar{x}, x_{n+1}) < \frac{1}{2}\epsilon$ and $d(f(x_n), f(\bar{x})) < \frac{1}{2}\epsilon$. It follows that $d(\bar{x}, f(\bar{x})) < \epsilon$.

Next we show that $\bar{x}$ is unique. Suppose on the contrary that there exists another $\hat{x} : \hat{x} = f(\hat{x})$. We would have:

$$d(\hat{x}, \bar{x}) = d(f(\hat{x}), f(\bar{x})) \le kd(\hat{x}, \bar{x}) < d(\hat{x}, \bar{x}),$$

which is a contradiction. $\|$

A.4 Optimization of continuous time systems

We have to:

$$\left| \begin{array}{l} J\left(k(t),t,T\right) \equiv \max\limits_{(v(\tau))_{\tau=t}^{T}} E\left\{ \int_t^T e^{-\rho(\tau-t)} u(k(\tau),v(\tau))d\tau + e^{-\rho(T-t)} B(k(T)) \right\} \\ \text{s.t. } dk(\tau) = T\left(k(\tau),v(\tau)\right)d\tau + \sigma\left(k(\tau),v(\tau)\right)dW(\tau) \\ \qquad k(t) = k_t, \text{ given.} \end{array} \right. \tag{3A4.1}$$

where $\{W(\tau)\}_{\tau=t}^{T}$ is a standard Brownian Motion, and:

$$J\left(k(T),T,T\right) = B(k(T)).$$

Here the control is v, and the observation of the trajectory of k would provide us with information that could make us change the initial chosen control v. Clearly, we must rule out controls depending on future observation of k, and a very natural and relatively simple way to also make the control be fed back with new information provided by k, is to confine attention to controls $v(t,\omega)$ that are a deterministic function of the state at time τ:

$$v(\tau;\omega) = \widehat{v}\left(\tau, k(\tau;\omega)\right),$$

where the deterministic function $\widehat{v}$ is usually referred to as a *feedback*. Clearly, in this case the *controlled* state evolves as:

$$dk(\tau) = T\left(k(\tau), \widehat{v}\left(\tau, k(\tau)\right)\right)d\tau + \sigma\left(k(\tau), \widehat{v}\left(\tau, k(\tau)\right)\right)dW(\tau),$$

and it is thus Markovian.

The stochastic programming principle also applies here:

$$\begin{aligned}
& J\left(k(t),t,T\right) \\
= {} & \max_v E\left\{ \int_t^T e^{-\rho(\tau-t)} u(k(\tau),v(\tau))d\tau + e^{-\rho(T-t)} B(k(T)) \right\} \\
= {} & \max_v E\left\{ \int_t^{t+\Delta t} e^{-\rho(\tau-t)} u(k(\tau),v(\tau))d\tau \right. \\
& \qquad \left. + \int_{t+\Delta t}^T e^{-\rho(\tau-t)} u(k(\tau),v(\tau))d\tau + e^{-\rho(T-t)} B(k(T)) \right\} \\
= {} & \max_v E\left\{ \int_t^{t+\Delta t} e^{-\rho(\tau-t)} u(k(\tau),v(\tau))d\tau \right. \\
& \qquad \left. + e^{-\rho\cdot\Delta t} \left(\int_{t+\Delta t}^T e^{-\rho(\tau-t-\Delta t)} u(k(\tau),v(\tau))d\tau + e^{-\rho(T-t-\Delta t)} B(k(T)) \right) \right\} \\
= {} & \max_v E\left\{ \int_t^{t+\Delta t} e^{-\rho(\tau-t)} u(k(\tau),v(\tau))d\tau \right. \\
& \left. + e^{-\rho\cdot\Delta t} E\left(\int_{t+\Delta t}^T e^{-\rho(\tau-t-\Delta t)} u(k(\tau),v(\tau))d\tau + e^{-\rho(T-t-\Delta t)} B(k(T)) \Big/ \mathcal{F}(t+\Delta t) \right) \right\} \\
= {} & \max_v E\left\{ \int_t^{t+\Delta t} e^{-\rho(\tau-t)} u(k(\tau),v(\tau))d\tau + e^{-\rho\cdot\Delta t} J\left(k(t+\Delta t), t+\Delta t, T\right) \right\},
\end{aligned}$$

where the fourth line follows from the law of iterated expectations. Rearranging terms and dividing throughout by Δt leaves:

$$\begin{aligned} 0 &= \max_v E\left\{ \frac{\int_t^{t+\Delta t} e^{-\rho(\tau-t)} u(k(\tau),v(\tau))d\tau}{\Delta t} + \frac{e^{-\rho\cdot\Delta t} J\left(k(t+\Delta t),t+\Delta t,T\right) - J\left(k(t),t,T\right)}{\Delta t} \right\} \\ &= \max_v E\left\{ \frac{\int_t^{t+\Delta t} e^{-\rho(\tau-t)} u(k(\tau),v(\tau))d\tau}{\Delta t} \right. \\ &\qquad \left. + e^{-\rho\cdot\Delta t} \frac{J\left(k(t+\Delta t),t+\Delta t,T\right) - J\left(k(t),t,T\right)}{\Delta t} - \left(\frac{1-e^{-\rho\cdot\Delta t}}{\Delta t}\right) J\left(k(t),t,T\right) \right\}. \end{aligned}$$

Sending Δt to zero and applying Itô's lemma produces the result that under mild regularity conditions,

$$\begin{cases} 0 &= \max_v \left\{ u(k,v) + \left(\frac{\partial}{\partial\tau} + L\right) J\left(k,\tau,T\right) - \rho J\left(k,\tau,T\right) \right\} \\ J\left(k(T),T,T\right) &= B(k(T)) \end{cases} \tag{3A4.2}$$

where

$$LJ(k,\tau,T) = \frac{\partial J\left(k,\tau,T\right)}{\partial k} T\left(k,v\right) + \frac{1}{2} \frac{\partial J^2\left(k,\tau,T\right)}{\partial k^2} \sigma\left(k,v\right)^2.$$

The first order conditions for (4A3.2) are:

$$0 = u_v\left(k,\widehat{v}\right) + \frac{\partial J\left(k,\tau,T\right)}{\partial k} T_v\left(k,\widehat{v}\right) + \frac{\partial J^2\left(k,\tau,T\right)}{\partial k^2} \sigma\left(k,\widehat{v}\right)\sigma_v\left(k,\widehat{v}\right),$$

which implicitly define the solution $\widehat{v} = \widehat{v}\left(\tau,k\right)$ with:

$$0 = u_v\left(k,\widehat{v}\left(\tau,k\right)\right) + \frac{\partial J\left(k,\tau,T\right)}{\partial k} T_v\left(k,\widehat{v}\left(\tau,k\right)\right) + \frac{\partial J^2\left(k,\tau,T\right)}{\partial k^2} \sigma\left(k,\widehat{v}\left(\tau,k\right)\right)\sigma_v\left(k,\widehat{v}\left(\tau,k\right)\right). \tag{3A4.3}$$

Also, the first line in (4A3.2) can be written as:

$$\begin{cases} 0 &= u\left(k,\widehat{v}\left(\tau,k\right)\right) + \left(\frac{\partial}{\partial\tau} + L\right) J\left(k,\tau,T\right) - \rho J\left(k,\tau,T\right) \\ LJ\left(k,\tau,T\right) &= \frac{\partial J\left(k,\tau,T\right)}{\partial k} T\left(k,\widehat{v}\left(\tau,k\right)\right) + \frac{1}{2} \frac{\partial J^2\left(k,\tau,T\right)}{\partial k^2} \sigma\left(k,\widehat{v}\left(\tau,k\right)\right)^2 \end{cases}$$

Next, define the marginal indirect utility with respect to the state variable in problem (4A3.1),

$$\lambda\left(k,\tau,T\right) \equiv \frac{\partial J\left(k,\tau,T\right)}{\partial k}.$$

We have

$$\mu\left(k,\tau,T\right) \equiv \frac{\partial\lambda\left(k,\tau,T\right)}{\partial k} = \frac{\partial^2 J\left(k,\tau,T\right)}{\partial k^2},$$

and we can thus write:

$$\begin{aligned}
& \rho J(k,\tau,T) - \frac{\partial J(k,\tau,T)}{\partial \tau} \\
= \; & u(k,\widehat{v}(\tau,k)) + \lambda(k,\tau,T)\,T(k,\widehat{v}(\tau,k)) + \frac{1}{2}\frac{\partial \lambda(k,\tau,T)}{\partial k}\sigma(k,\widehat{v}(\tau,k))^2 \\
\equiv \; & \widehat{H}\left(\tau,k,\frac{\partial J}{\partial k},\frac{\partial^2 J}{\partial k^2}\right),
\end{aligned}$$

where

$$\begin{aligned}
& \widehat{H}(\tau,k,\lambda,\mu) \equiv \widehat{H}\left(\tau,k,\frac{\partial J}{\partial k},\frac{\partial^2 J}{\partial k^2}\right) \\
\equiv \; & u(k,\widehat{v}(\tau,k)) + \lambda(k,\tau,T)\,T(k,\widehat{v}(\tau,k)) + \frac{1}{2}\mu(k,\tau,T)\,\sigma(k,\widehat{v}(\tau,k))^2,
\end{aligned}$$

and $\widehat{H}$ is usually referred to as (optimized) Hamiltonian, and the previous equation is the *Bellman equation* for diffusion processes.

By Itô's lemma:

$$d\lambda = \left(\frac{\partial \lambda}{\partial \tau} + \frac{\partial \lambda}{\partial k}T + \frac{1}{2}\frac{\partial^2 \lambda}{\partial k^2}\sigma^2\right)d\tau + \frac{\partial \lambda}{\partial k}\sigma dW. \tag{3A4.4}$$

On the other hand, by differentiating every term in the Bellman equation

$$\rho J - \frac{\partial J}{\partial \tau} = \widehat{H}(\tau,k,\lambda(\tau,k),\mu(\tau,k))$$

with respect to k, we get:

$$\begin{aligned}
\rho\frac{\partial J}{\partial k} - \frac{\partial^2 J}{\partial k \partial \tau} &= \rho\lambda - \frac{\partial \lambda}{\partial \tau} \\
&= \frac{\partial \widehat{H}}{\partial k} + \frac{\partial \widehat{H}}{\partial \lambda}\frac{\partial \lambda(\tau,k)}{\partial k} + \frac{\partial \widehat{H}}{\partial \mu}\frac{\partial \mu(\tau,k)}{\partial k} \\
&= \frac{\partial \widehat{H}}{\partial k} + T\frac{\partial \lambda(\tau,k)}{\partial k} + \frac{1}{2}\sigma^2\frac{\partial \mu(\tau,k)}{\partial k},
\end{aligned}$$

or,

$$\rho\lambda - \frac{\partial \widehat{H}}{\partial k} = \frac{\partial \lambda}{\partial \tau} + T\frac{\partial \lambda}{\partial k} + \frac{1}{2}\sigma^2\frac{\partial \mu}{\partial k}.$$

By plugging this into (3A4.4) we get:

$$d\lambda = \left(\rho\lambda - \frac{\partial \widehat{H}}{\partial k}\right)d\tau + \frac{\partial \lambda}{\partial k}\sigma dW,$$

from which we get:

$$\rho\lambda - \frac{\partial \widehat{H}}{\partial k} = \frac{E(d\lambda)}{d\tau} = \frac{\partial \lambda}{\partial \tau} + \frac{\partial \lambda}{\partial k}T + \frac{1}{2}\frac{\partial^2 \lambda}{\partial k^2}\sigma^2.$$

The infinite horizon case

$$\left| \begin{array}{l} J\left(k(t)\right) \equiv \max\limits_{(v(\tau))_{\tau=t}^{\infty}} E\left\{\int_t^{\infty} e^{-\rho(\tau-t)} u(k(\tau), v(\tau)) d\tau\right\} \\ \\ \text{s.t. } dk(\tau) = T\left(k(\tau), v(\tau)\right) d\tau + \sigma\left(k(\tau), v(\tau)\right) dW(\tau) \\ \qquad k(t) = k_t, \text{ given.} \end{array} \right.$$

is treated similarly. We have

$$\begin{cases} 0 & = & \max\limits_{v}\left\{u(k,v) + LJ\left(k,\tau,T\right) - \rho J\left(k,\tau,T\right)\right\} \\ \\ \lim_{T\to\infty} e^{-\rho T} E\left\{J\left(k(T)\right)\right\} & = & 0 \end{cases}$$

and denoting again $\lambda = \frac{\partial J}{\partial k}$, we have:

$$d\lambda = \left(\rho\lambda - \frac{\partial \widehat{H}}{\partial k}\right) d\tau + \frac{\partial \lambda}{\partial k}\sigma dW,$$

from which we get:

$$\rho\lambda - \frac{\partial \widehat{H}}{\partial k} = \frac{E\left(d\lambda\right)}{d\tau} = \frac{\partial \lambda}{\partial k} T + \frac{1}{2}\frac{\partial^2 \lambda}{\partial k^2}\sigma^2.$$

Generally then the problem can be solved by implementing the following two steps:

- The first step should be very simple in many cases: it consists in maximizing the Hamiltonian,

  $$H \equiv u(k,v) + \lambda \cdot T\left(k,v\right) + \frac{1}{2}\cdot\mu\cdot\sigma\left(k,v\right)^2,$$

  viz.,

  $$\max_v\left\{u(k,v) + \lambda \cdot T\left(k,v\right) + \frac{1}{2}\cdot\mu\cdot\sigma\left(k,v\right)^2\right\}.$$

  In the deterministic macroeconomic models that we saw before (where $\sigma \equiv 0$), for instance, we got first order conditions that we subsequently used to characterize the equilibrium dynamics of the economy, along with the condition:

  $$\rho\lambda - \frac{\partial \widehat{H}}{\partial k} = \frac{\partial \lambda}{\partial k} T + \frac{1}{2}\frac{\partial^2 \lambda}{\partial k^2}\sigma^2.$$

- Step two involves artistic skills. It consists in solving the previous partial differential equation for J. In implementing step two, don't forget to use the boundary condition of the problem that

  $$\lim_{T\to\infty} e^{-\rho T} J(k(T)) = 0.$$

To summarize, the first order conditions are:

$$\begin{cases} \dfrac{\partial H}{\partial v} & = & 0 & \text{which gives us } \widehat{H} \\ \\ \dfrac{\partial H}{\partial \lambda} & = & T & \left(\dfrac{\partial H}{\partial \lambda} = \dot{k} \text{ if } \sigma \equiv 0\right) \\ \\ d\lambda & = & \left(\rho\lambda - \dfrac{\partial \widehat{H}}{\partial k}\right) d\tau + \sigma\left(k, \widehat{v}\right) dW & \left(\rho\lambda - \dfrac{\partial \widehat{H}}{\partial k} = \dot{\lambda} \text{ if } \sigma \equiv 0\right) \end{cases}$$

A.5 On linear functionals

A continuous linear functional f on a space S is a linear map from S onto the set of complex numbers C, i.e. $f : S \mapsto C$, and it is such that to every $\varphi \in S$, it corresponds the complex number denoted as (f, φ) under the following conditions:[8]

1. *Linearity*: $(f, \alpha_1\varphi_1 + \alpha_2\varphi_2) = \alpha_1(f, \varphi_1) + \alpha_2(f, \varphi_2)$ for every complex numbers α_1 and α_2 and functions φ belonging to S.
2. *Continuity*: If a sequence of functions $\varphi_1, \varphi_2, \cdots, \varphi_n$ tends to zero in S, the sequence of complex numbers $(f, \varphi_1), (f, \varphi_2), \cdots, (f, \varphi_n)$ also tends to zero.

Any functional satisfying the previous conditions is also called *tempered distribution*, and the set of all tempered distributions will be denoted as S'.

Example 1 (The price system in an infinite commodity space): Let f be a locally integrable function bounded by a power of $|x|$ as $|x| \to \infty$. We associate to this function a functional, also noted as f, via the formula:

$$(f, \varphi) = \int f(x)\varphi(x)dx, \quad \varphi \in S.$$

Example 2 (δ-distribution): Condider the functional which associates the number $\varphi(0)$ to every function $\varphi \in S$, which is denoted as

$$(\delta, \varphi) = \varphi(0).$$

An abusive but practical way to denote this is also $\int \delta(x)\varphi(x)dx = \varphi(0)$.

Next, let V be a Hilbert space, i.e. a real vector space endowed with a scalar product denoted as (u, v) such that V is complete with respect to the associated norm $\|v\| = \sqrt{(v, v)}$, which essentially means that every Cauchy sequence does converge. An example is $L^2(\Omega)$ endowed with the scalar product $\int u(x)v(x)dx$, where Ω is an open set of $\mathbb{R}^d$ or, very simply, $\mathbb{R}^d$ endowed with the Euclidian scalar product. A set $K \subset V$ is convex iff $\theta u + (1-\theta)v \in K$ $\forall u, v \in K$, $\forall \theta \in [0, 1]$.

Theorem 1 (Projection on a convex set): *Let K be a nonempty, closed and convex part of V. Then, for each $u_0 \in V$, there exists a unique point $u \in K$ such that*:

$$\|u - u_0\| = \inf_{v \in K} \|v - u_0\|.$$

Furthermore, u is the unique point of K such that:

$$(u - u_0, u - v) \leq 0, \quad \forall v \in K. \tag{5A1}$$

[8] Here S is a vector space of the complex valued functions $\varphi(x)$ of several complex variables x satisfying the following conditions: functions φ are infinitely differentiable and, when x tends to infinity, φ and their derivatives of all orders tend to zero at a rate higher than $|x|^{-1}$ (e.g., e^{-x^2}, $x^n e^{-x^2}$).

For evident reasons, u is called the projector of u_0 on K.

THEOREM 2 (The Riesz representation of a linear functional): *For every linear functional L continuous on V, there exists a unique $u \in V$ such that*:

$$L(v) = (u, v), \quad \forall v \in V.$$

EXERCISE 1: Use theorem 2 to derive the first part of the fundamental theorem of finance.

EXERCISE 2: Use theorem 2 to establish Pareto optimality in the standard setting of the Real Business Cycles models of chapter 3. [Hint: use the evaluation equilibria approach pioneered by Debreu (1953)]

THEOREM 3 (Separation of a point from a convex set, again: see also chapter 1): *Let K a nonempty convex part of V, and $u_0 \notin K$. Then, there exists a closed hyperplan in V which strictly separates u_0 and K, i.e. there exists a linear functional $L \in V'$* (the space of linear functionals which are continuous on V) *and $\alpha \in \mathbb{R}$ such that*:

$$L(u_0) < \alpha < L(v), \quad \forall v \in K.$$

PROOF: Let u be the projector of u_0 on K. Since $u_0 \notin K$, $u - u_0 \neq 0$. Let L be the linear functional defined, for each $w \in V$, as:

$$L(w) = (u - u_0, w),$$

and let

$$\alpha = \frac{L(u) + L(u_0)}{2}.$$

By (5A1),

$$L(u - v) \equiv (u - u_0, u - v) \leq 0,$$

and since $L(u - v) = L(u) - L(v)$, it follows that:

$$L(u) \leq L(v).$$

Furthermore, $L(u_0) < L(u)$ because $L(u - u_0) = (u - u_0, \underbrace{u - u_0}_{\notin K}) > 0$, and then:

$$L(u_0) < \alpha < L(u) \leq L(v), \quad \forall v \in K.$$

||

THEOREM 4 (Farkas lemma): *Let $a_1, \cdots, a_M$ be fixed elements of V, and consider the sets*:

$$\begin{cases} \mathbb{K} = \{v \in V, \quad (a_i, v) \leq 0 \text{ for } 1 \leq i \leq M\} \\ \widehat{\mathbb{K}} = \left\{v \in V, \quad \exists \lambda_1, \cdots, \lambda_M \geq 0, \; v = -\sum_{i=1}^M \lambda_i a_i\right\} \end{cases}$$

Then, for every $p \in V$, we have that:

$$(p, w) \geq 0 \quad \forall w \in \mathbb{K} \Rightarrow p \in \widehat{\mathbb{K}}.$$