Financial Econometrics

S.T. Rachev

Manuscript: Markus Höchstötter

Course taught as part of the FE module S-2 at the Hector School of Enigineering and Management

University Fridericiana Karlsruhe

Contents

Chapter 1

Introduction

One of the major tasks of financial research is the prediction of price changes. It is natural to draw on past information from the series of price chages to form reasonable conclusions. Whether absolute or as logarithms, proce chnges have long been believed to be serially independent. A refinement results from the assumption that price changes or the logarithms of which are normally distributed, hence, are Brownian motion. This long believed hypothetical independence lead to view prices as random walks. Formally, the evolution of prices can be stated by $P_t = P_{t-1} + a_t$ with P_t the observed proce at time t and a_t serially uncorrelated error terms with zero expectation. Consequently, the price changes are serially uncorrelated themselves. The price at time t can then be written as $P_t = \sum_{i=1}^{t} a_i$. Very often, the a_t are modeled as normal random variables with finite, constant variance. Laid as a foundation by Bachelier in his 1900 PhD dissertation, prices were thought to ressemble a Brownian motion type random walk. Later on, the autocorrelative structures of prices were analyzed in the context of ARIMA models.

For some prices, not necessarily stock prices, other models are deemed more appropriate than random walks or from the class of ARIMA structures.

In addition to the order of integration, a great variety of models for trends and compositions of the prices can be thought of, instead. An important factor plays the memory of the series.

Great critique of the random walk model resulted from the development of the theory of efficient markets and the notion of economic equilibria. Sequential price changes could no longer be considered merely independent. Instead, the theroy of martingales was introduced. Let the return of, for example, a stock price be defined by

$$r_{t+1} = \frac{P_{t+1} + D_t - P_t}{P_t} \tag{1.1}$$

where D_t is the dividend paid druing period t. Further, let us assume that the return has a constant expected value, $E_t(r_{t+1}) = r$. This leads to the idea of the price in t being its discounted expected value, i.e.

$$P_t = \frac{E_t(P_{t+1} + D_t)}{1 + r}. \tag{1.2}$$

Additionally, the understanding of reinvesting proceeds from the stock such as dividends into a fund's position in the stock at time t, x_t, leads to the idea of considering the fund's position as a martingale when discounted by the expected return, i.e.

$$x_t = \frac{E_t(x_{t+1})}{1 + r}. \tag{1.3}$$

The price itself, then, is a submartingale when the dividend price ration is constant, d, and $r > d$, i.e.

$$P_t = \frac{E_t(P_{t+1})}{(1 + r - d)} < E_t(P_{t+1}). \tag{1.4}$$

Hence, under the martingale assumption, the conditional first order moment of the difference $P_{t+1} - P_t$ does not depend on the information at t. This is less restrictive than the random walk assumption where not only the first order

moments but also higher conditional moments are independent of the information $\mathcal{F}_t$. Particularly, the observable quite and noisy times, respectively, and the inherent autocorrelation of conditional variances of price movements can be caught by martingales which is not feasible for random walks. This aspect of Martingale theory is consequently extended to non-linear stochastic processes such as Engle's (1982) autoregressive conditionally heteroscedastic, or ARCH, model in order to provide a structure for higher conditional moments. Essential, in this respect, are tests for non-linearity. Also, multivariate versions of ARCH are introduced as well as multivariate regression techniques. The time series are analyzed with respect to stationarity. Tests for misspecification are also carried through as well as dynamic regression methods.

Later on in the text, integrated financial time series are analyzed. For this purpose, cointegrated processes and testing for cointegration is introduced which is carried on to cointegrated systems and their estimation.

Chapter 2

Stationary and nonstationary time-series models

2.1 Stochastic processes, ergodicity and stationarity

2.1.1 Stochastic Processes

For analyzing financial time series, it is advisable to regard the observed series, $(x_1, x_2, \ldots, x_T)$, as a realization, denoted $\{x_t\}_1^T$, of a stochastic process[1], $\{X_t\}_{-\infty}^{\infty}$. In the following, the index set, though, will be restricted to that of the realization, i.e. $\mathcal{T} = (1, T)$. The realization is related to the stochastic process in the same manner as a sample is related to the entire population. The stochastic process is identified by a T-dimensional probability distribution. In general, the focus will be on the T first moments, $E(x_1), E(x_2), \ldots, E(x_T)$, the T variances, $V(x_1), V(x_2), \ldots, V(x_T)$, and the $T(T-1)/2$ covariances, $Cov(x_i, x_j)$. Though unrealistic, the assumption of

[1]For simplicty, written as x_t, in the sequel.

joint normality of the distribution would characterize the stochastic process exhaustively. If the process, however, can be modeled as a linear combination of past values of the process itself and other processes, and Gaussianity cannot be assumed, then it could still be described by the above expectations. But the number of observations is not sufficient to infer on the remaining $T + T(T+1)/2$ parameters. THere are too many parameters to estimate for, then.

2.1.2 Ergodicity

When there is, however, only a finite realization available, the possibility of inference depends on the *ergodicity* of the series. That means that sample moments approach their population counterparts as the number of observations becomes infinite. Formally,

Definition 2.1.1. Ergodic A time series is ergodic iff its sample moments converge to their population moments in probability.

In the following, processes are assumed to be ergodic to simplify matters.

2.1.3 Stationarity

A process has very pleasant features if it is *stationary*. Let the mean and the variance of a process be constant, respectively. Moreover, for a given k, let the autocovariance function

$$\gamma_k = Cov(x_t, x_{t-k})$$

and the autocorrelation function (ACF)

$$\rho_k = \frac{Cov(x_t, x_{t-k})}{V(x_t)V(x_{t-k})} = \frac{\gamma_k}{\gamma_0}$$

be the same for all t and only dependend on the lag k, then the process is said to be *weakly stationary* or, since only the first two moments are considered, *covariance stationary*. If additionally for arbitrary m, however, the joint probability distribution of the process at any set of times, $t_1, t_2, \ldots, t_m$, is the same as $t_{1+k}, t_{2+k}, \ldots, t_{m+k}$ for any value of k, i.e. unaffected by time shifts, then the process is said to *strictly stationary*. When moments exist, that is finite, then strict staionarity implies weak stationarity.

Because of the symmetry assumption of the ACF, i.e. $\gamma_{-k} = \gamma_k$, only the half of positive k is conventionally given. In the analysis of x_t, the ACF expresses the serial dependence structure[2] of the process and, hence, the memory within the proecess. Together with mean, $\mu = E(x_t)$, and variance, $\sigma^2 = \gamma_0 = V(x_t)$, the ACF is essential in describing the realization.

2.2 ARMA

2.2.1 Linear filter

By the Wold decomposition, any weakly stationary and purely non-deterministic process, $x - \mu$, is a linear combination of a sequence of random variables that are serially uncorrelated. Purely non-deterministic means that any perfectly predictable quantity, μ, has been subtracted from the process x_t. Hence it can be written in the form

$$x_t - \mu = a_t + \psi_1 a_{t-1} + \psi_2 a_{t-2} + \ldots = \sum_{j=0}^{\infty} \psi_j a_{t-j}, \ \psi_0 = 1 \qquad (2.1)$$

where the a_j are uncorrelated random variables.Let us assume in the following, that, without loss generality, $\mu = 0$ and set $\psi_j = \phi^j$.

[2]In other words, influence from the past.

2.2.2 AR(1)

2.1 can be expressed as the *first-order autoregressive* or AR(1) process

$$\begin{aligned} x_t &= a_t + \phi a_{t-1} + \phi^2 + \ldots = a_t + \phi(a_{t-1} + \phi a_{t-2} + \ldots) \\ &= \phi x_{t-1} + a_t. \end{aligned} \tag{2.2}$$

Now, we introduce the *lag* operator, B, to be operator shifting from observation x_t to x_{t-1} in the way $Bx_t \equiv x_{t-1}$. This can be extended to multiple lags by $B^d x_t \equiv x_{t-d}$. Thus, 2.2.2 can be written as

$$(1 - \phi B)x_t = a_t. \tag{2.3}$$

or, equivalently,

$$\begin{aligned} x_t &= (1 - \phi B)^{-1} a_t = (1 + \phi B + \phi^2 B^2 + \ldots)a_t \\ &= a_t + \phi a_{t-1} + \phi^2 a_{t-2} + \ldots \end{aligned} \tag{2.4}$$

which converges for $|\phi| < 1$.

The ACF of the AR(1) process can obtained from

$$\gamma_k - \phi\gamma_{k-1} = E(a_t x_{t-k}) \tag{2.5}$$

by, first, multiplying both sides by x_{t-k} and, then, taking expectations of both sides. Because of the independence of a_t and x_{t-k}, 2.5 can be rewritten as

$$\gamma_k = \phi\gamma_{k-1}, \ \forall \ k. \tag{2.6}$$

Carrying on the recursion to $\gamma_k = \phi^k \gamma_0$ with ACF $\rho_k = \phi^k$, one can see for $\phi > 0$ that the ACF decays exponentially to zero whereas for $\phi < 0$, the ACF decays in an oscillating way to zero.

2.2.3 MA(1)

Consider next the so called *first order moving average* or MA(1) model

$$x_t = a_t - \theta a_{t-1}. \tag{2.7}$$

Analogously, it can be written as

$$x_t = (1 - \theta B)a_t.$$

The autocovariance function is

$$\begin{aligned} \gamma_0 &= \sigma^2(1 + \theta^2) \\ \gamma_1 &= -\sigma^2\theta \\ \gamma_k &= 0 \text{ for } k > 1. \end{aligned}$$

Consequently, the ACF is

$$\begin{aligned} \rho_1 &= \frac{-\theta}{1 + \theta^2} \text{ and} \\ \rho_k &= 0, \text{ for } k > 1. \end{aligned} \tag{2.8}$$

The lack of memory of this type of process beyond lag 1 and, hence, the abrupt jump to zero in the ACF for $k \geq 2$ is in drastic contrast to the rather smooth decay of the ACF of an AR(1) process.

Rewriting ρ_1 from 2.2.3 as $\theta^2\rho_1 + \theta + \rho_1 = 0$, generates real solutions for the first order coefficients if $-.5 < \rho_1 < .5$. Always two θ satisfy this equation such that always two corresponding MA(1) processes have the same ACF.

All MA models are stationary. If for the MA(1), $|\theta| < 1$, the model is said to *invertible*. That means that it can be transformed, or rather inverted, into an autoregressive process of the form

$$x_t = \pi_1 x_{t-1} + \pi_2 x_{t-2} + \ldots + a_t$$

with $\sum |\pi_j| < \infty$ by $a_t = x_t(1 - \theta B)^{-1}$. The following relation holds $\pi_j = -\theta^j$, and one can see that the *invertibility* condition, $|\theta| < 1$, is necessary for a_t to have finite (second) moments.

2.2.4 General AR and MA processes

Naturally, the orders of the respective processes are not restricted to the value one. The autoregressive process of general order p can be written as

$$\begin{gathered} x_t - \phi_1 x_{t-1} - \phi_2 x_{t-2} - \ldots - \phi_p x_{t-p} = a_t \text{ or} \\ \left(1 - \phi_1 B - \phi_2 B^2 - \ldots - \phi_p B^p\right) = a_t \text{ or shorter} \\ \phi(B) x_t = a_t. \end{gathered} \tag{2.9}$$

So, the generalized MA representation, $x_t = \psi(B)a_t$, can be found through division of $\phi(B)$ of both sides of 2.2.4. The ψ converge in the case of stationarity, i.e. when the characteristic roots of

$$\phi(B) = (1 - g_1 B)(1 - g_2 B) \ldots (1 - g_p) = 0 \tag{2.10}$$

all lie outside the unit circle or, equivalently, when $|g_i| < 1,\ i = 1, 2, \ldots, p$. The solution to the difference equation

$$\phi(B)\rho_k = 0, k > 0, \tag{2.11}$$

which is a linear combination of the roots,

$$\rho_k = A_1 g_1^k + A_2 g_2^k + \ldots + A_p g_p^k,$$

governs the ACF. Its shape is consequently a blend of either damped exponentials or damped sine waves in the case of real roots or complex roots, respectively.

The AR(2) process may serve as an example,

$$(1 - \phi_1 B - \phi_2 B^2) = a_t.$$

From its characteristic equation, $\phi(B) = 0$, the roots

$$g_{1,2} = \frac{1}{2}\left(\phi_1 \pm (\phi_1^2 + 4\phi_2)^{1/2}\right) \tag{2.12}$$

can be obtained as real or complex numbers. For the process to be stationary, it is necessary that $|g_1| < 1$ and $|g_2| < 1$. Hence, the respective restrictions are

$$\begin{aligned} \phi_1 + \phi_2 &< 1, \\ -\phi_1 + \phi_2 &< 1, \text{ and} \\ -1 < \phi_2 &< 1. \end{aligned} \tag{2.13}$$

From (2.12) it can be seen that the roots are complex if $\phi_1^2 + 4\phi_2 < 0$.

The ACF, in this case, for the four possible $(\phi_1, \phi 2)$ are shown in figure 2.1 . As stated before, the ACF is a blend of damped, possibly, oscillating exponentials. In the complex case, it is a blend of damped, possibly, oscillating sine waves. Plots of time series generated for the four different AR(2) processes can be seen in figure 2.2 . The a_t are $NID(0, 25)$. In case of the real roots, the graphs may appear either smooth or jagged depending on the signs of the roots whereas the graphs seem periodic when the roots are complex.

It is sometimes difficult to infer upon the correct order of an AR(p) process from looking at the ACF. Hence, the *partial autocorrelation function*, or PACF, should be referred to, alternatively. It determines the part of the correlation between x_t and x_{t-k} which is actually the correlation which both, x_t and x_{t-k}, have with the intermediate variables, $x_{t-1}, x_{t-2}, \ldots, x_{t-k+1}$.

More formally, the partial autocorrelation is the coefficient ϕ_{kk} of the AR(k) process

$$x_t = \phi_{k1}x_{t-1} + \phi_{k2}x_{t-2} + \ldots + \phi_{kk}x_{t-k} + a_t, \tag{2.14}$$

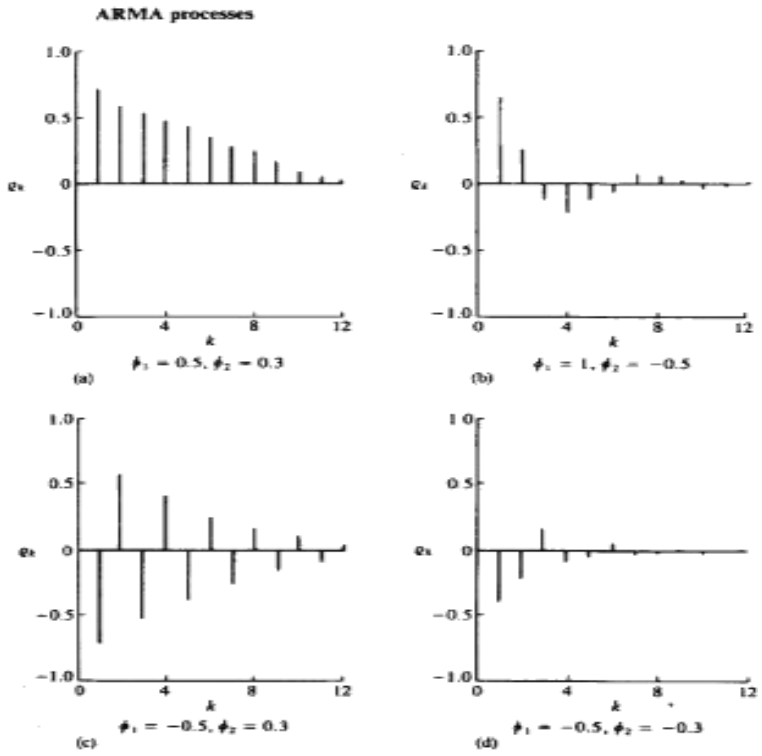


Figure 2.1: ACFs of various AR(2) processes.

i.e. the autoregression of x_t on k lagged terms. The ϕ_{kk} are commonly obtained through solving the so called *Yule-Walker* equations given from succesively multiplying $x_t\phi(B)$ by $x_{t-1}, x_{t-2}, \ldots, x_{t-k}$ and taking expectations.

This leads to the system

$$\phi_{kk} = \frac{\begin{vmatrix} 1 & \rho_1 & \dots & \rho_{k-2} & \rho_1 \\ \rho_1 & 1 & \dots & \rho_{k-3} & \rho_2 \\ & & \ddots & & \\ \rho_{k-1} & \rho_{k-2} & \dots & \rho_1 & \rho_k \end{vmatrix}}{\begin{vmatrix} 1 & \rho_1 & \dots & \rho_{k-2} & \rho_{k-1} \\ \rho_1 & 1 & \dots & \rho_{k-3} & \rho_{k-2} \\ & & \ddots & & \\ \rho_{k-1} & \rho_{k-2} & \dots & \rho_1 & 1 \end{vmatrix}}.$$

So, for particular p,

$$\phi_{11} = \rho_1 = \phi,\ \phi_{kk} =, k > 1,\ \text{when } p = 1,$$
$$\phi_{11} = \rho_1, \phi_{22} = \frac{\rho_2 - \rho_1^2}{1 - \rho_1^2},\ \phi_{kk} =, k > 2,\ \text{when } p = 2,$$
$$\phi_{11} \neq 0, \phi_{22} \neq 0, \dots, \phi_{pp} \neq 0,\ \phi_{kk} =, k > p.$$

The AR(p) process is consequently described by a gradual decay of the ACF which is a combination of damped exponentials and sine waves, and an abrupt vanish of the PACF for lags greater than p.

For the general MA(q),

$$x_t = a_t - \theta_1 a_{t-1} - \ldots - \theta_q a_{t-q},$$

the ACF is of the form

$$\rho_k = \frac{-\theta_k + \theta_1\theta_{k+1} + \ldots + \theta_{q-k}\theta_q}{1 + \theta_1^2 + \ldots + \theta_q^2},\ k = 1, 2, \dots, q,$$
$$\rho_k = 0,\ k > q.$$

In contrast to the ARp), the ACF vanishes for lags greater than q. Hence, there is no memory beyond q periods.

Now, for the MA(q) to be invertible, the characteristic roots solve

$$(1 - \theta_1 B - \ldots - \theta_q B^q) = (1 - h_1 B) \ldots (1 - h_q B) = 0 \tag{2.15}$$

for $|h_i| < 1,\ i = 1, 2, \ldots, q$. In figure 2.3,

, two MA(2) processes are displayed for $a_t \sim N(0, 25)$. They are very similarly jagged as an AR(2) with real roots $g_1 \cdot g_2 < 0$. The PACF, now, vanishes slowly in contrast to the AR(p). A formal expression of which is rather complicated. They, however, look similar to the ACfs of AR processes. There appears to be a reciprocity in the respective behavior of the AR process' ACF and PACF and the MA process' ACF and PACF.

2.3 ARMA(p,q) and ARIMA(p,d,q)

A natural consequence may be to consider mixtures of the above AR and MA processes. This brings us straight to the autoregressive-moving average models. Consider an AR(1) and an MA(1) process, respectively. When combining both, the result is a *first-order autoregressive-moving average*, or ARMA(1,1) model of the form

$$x_t - \phi x_{t-1} = a_t - \theta a_{t-1}. \tag{2.16}$$

Division of both sides by $(1 - \phi B)$ leads to the corresponding MA(∞) model

$$\begin{aligned} x_t = \psi(b) a_t &= \left(\sum_{i=0}^{\infty} \phi_i B^i (1 - \theta B) a_t \right) \\ &= a_t + (\phi - \theta) \sum_{i=0}^{\infty} \phi^{i-1} a_{t-i} \\ \text{where } \psi(B) &= \frac{1 - \theta B}{1 - \phi B}. \end{aligned} \tag{2.17}$$

Conversely, the ARMA(1,1) process can be transformed into an AR(∞) representation with weights

$$\pi(B) = \frac{1 - \phi B}{1 - \theta B}$$

such that

$$a_t = \pi(B)x_t = \left(\sum_{i=0}^{\infty} \theta_i B^i \right) (1 - \phi B)x_t$$

which is equivalent to

$$x_t = (\phi - \theta) \sum_{i=0}^{\infty} \theta^{i-1} x_{t-i} + a_t.$$

The process is stationary if $|\phi| < 1$ and, alternatively, invertible if $|\theta| < 1$.

To obtain the ACF from an ARMA(1,1) representation, consider multiplying both sides of (2.16) by x_{t-k} and taking expectations. This yields

$$\begin{aligned} \gamma_k &= \phi\gamma_{k-1}, \ k > 1, \\ \gamma_0 - \phi\gamma_1 &= \sigma^2 - \theta(\phi - \theta)\sigma^2, \ k = 0, \\ \gamma_1 - \phi\gamma_0 &= -\theta\sigma^2, \end{aligned}$$

respectively, from the fact that $x_{t-k}a_{t-j}$ has zero expectation for $k > j$. From these equations, the ACF can be retrieved after a few steps such that

$$\begin{aligned} \rho_1 &= \frac{(1 - \phi\theta)(\phi - \theta)}{1 + \theta^2 - 2\phi\theta}. \\ \rho_k &= \phi\rho_{k-1}, \ k > 1 \end{aligned}$$

For $k \geq 1$, this ressembles the structure of the ACF of an AR(1) process in that it decays at an exponential rate. There is a difference, however. With $\phi > \theta, \ \phi, \theta \neq 0$, the ρ_1 can be much smaller than ϕ which one recalls to be the ACF of an AR(1) process.

When AR and MA are both of higher order, the combination of them, consequently, leads to a higher order ARMA process. So, if AR(p) and MA(q), the resulting model is ARMA(p, q) with the form

$$x_t - \phi_1 x_{t-1} - \ldots - \phi_p x_{t-p} = a_t - \theta_1 a_{t-1} - \ldots - \theta_q a_{t-q},$$
$$\text{or short } \phi(B)x_t = \theta(B)a_t.$$

As to the stationarity and invertibility conditions, they depend on those of the respective AR and MA processes forming the ARMA process. Beyond a certain number of lags, that is $q - p$, the ACF displays the shape of that of an AR(p) process whereas the PACF will decline like that of an MA(q) process after $p - q$ lags.[3]

It has been silently preconditioned in the previous that the process has zero mean, i.e. $\mu = 0$. This can easily be altered, however, by simply replacing x_t with the, now, mean-corrected $x_t - \mu$, such that the ARMA(p,q) model turns into

$$\phi(B)(x_t - \mu) = \theta(B)a_t.$$

2.3.1 Linear stochastic processes

In this short paragraph, attention is given to the innovations, $\{a_t\}$. Up to now, it has been merely said that they are uncorrelated and identically distributed with zero mean and finite variance. There is a little more to it, though. In other words, they are white noise, i.e. $a_t \sim WN(0, \sigma^2)$. It is also possible that the a_t are additionally independent which renders the sequence *strict* white noise, denoted $a_t \sim SWN(0, \sigma^2)$. Consequently, if a stationary process x_t is designed as a linear filter of the strict white noise, i.e. $x_t = \psi(B)a_t,\ a_t \sim SWN(0, \sigma^2)$, then x_t is referred to as a *linear* process.

[3]That is if $q - p > 0$ or $p - q > 0$, respectively.

The resulting stationary process from a filter of a white noise process need not be linear itself.

2.3.2 Non-stationary processes and ARIMA models

Up to now, variance as well as mean have been supposed to be constant which is the case for a weakly stationary process. But this may contradict reality since many financial series exhibit moments that vary with time.

Let us first consider a non-stationarity in variance. Consider the process

$$x_t = \mu_t + \epsilon_t \tag{2.18}$$

where μ_t is the *deterministic* mean level and ϵ_t is the error term. Additionally, let the variance of the error term depend on the mean level in the functional relationship $V(x_t) = V(\epsilon_t) = h^2(\mu_t)\sigma^2$ and h is a known function. Now, one has to find a function g to transform the data into $g(x_t)$ which renders the variance of the transformed process constant. To do this, consider the Taylor expansion of $g(x_t)$ around the mean level to obtain $g(x_t) = g(\mu_t) + (x_t + \mu_t)g^{'}(\mu_t)$ with $g^{'}$ denoting first derivatives. Hence, one obtains an approximation of the variance by

$$V[g(x_t)] \approx [g^{'}(\mu_t)^2]h^2(\mu_t)\sigma^2.$$

The transform function, g, then is found from

$$g^{'}(\mu_t) = \frac{1}{h(\mu_t)}.$$

Next, we approach the problem if the mean is variant in time. In 2.18, μ_t can be thought of to change in many ways. Assume that it is a polynmial in time t of order d. Thus, the process can be described by

$$x_t = \sum_{i=0}^{d} \beta_j t^j + \psi(B)a_t$$

with

$$E[x_t] = \sum_{j=0}^{d} \beta_j t^j,$$

hence, deterministic. Consider the example where $d = 1$ such that

$$x_t = \beta_0 + \beta_1 t + a_t.$$

First-differencing, that is subtracting x_{t-1} from both sides removes the time component, i.e.

$$x_t - x_{t-1} = \beta_1 + a_t - a_{t-1}. \tag{2.19}$$

This is an ARMA(1,1) process which is neither stationary nor invertible because of the unit roots of both, AR and MA components. Introducing the *first-differencing operator*, ∇, we can rewrite 2.19 in the fashion

$$w_t \equiv \nabla x_t = x_t - x_{t-1} = (1 - B)x_t = \beta_1 + \nabla a_t.$$

While the MA(1) part of the newly generated process, w_t, is not invertible, w_t itself is stationary.

In general, the trend may be a polynomial of order d and the error process of 2.18 may be generated by the ARMA process

$$\phi(B)\epsilon_t = \theta(B)a_t.$$

Then, first-differencing x_t d times, we obtain the new process

$$\nabla^d x_t = \theta_0 + \frac{\nabla^d \theta(B)}{\phi(B)} a_t$$

with $\theta_0 = d!\beta_d$. The MA part has, now, d unit roots, and the variance of x_t is found to be invariant to t. This can be seen in an example for, both, linear and quadratic trends. The respective models are

$$x_t = 10 * 2t + a_t \text{ and}$$
$$x_{=}10 + 5t - .03t^2 + a_t$$

with white noise specified by $a_t \sim iid\ N(0, 9)$. This can be seen in figure 2.5.

Consider next ARMA process with non-stationary AR part. This may be demonstrated with the AR(1) model

$$x_t = \phi x_{t-1} + a_t = x_t = x_0\phi^t + \sum_{i=0}^{t} \phi^i a_{t-i}. \tag{2.20}$$

where $\phi > 1$. Suppose the process started at x_0. Now, at each time τ prior to t, (2.20) yields the conditional expectation of x_t in terms of x_0 and the past shocks, $a_0, \ldots, a_{\tau-1}$. Hence, the process has a *stochastic* trend which is altered each period by a new shock. The variance of the process can be given by

$$V(x_t) = \sigma^2 \frac{\phi^{2(t+N+1)} - 1}{\phi^2 - 1}$$

which is increasing in time and growing to infinity due to $\phi > 1$. Since there is a trend in mean and variance, the process is referred to as *explosive*.

Generating a process as in (2.20) with $x_0 = 10, \phi = 1.05$ and $a_t \sim\ iiN(0, 9)$, behaves as demonstrated in figure 2.6.

The series describes an exponential curve after starting out similarly to the previous AR(1) samples. This feature will remain when additional AR and MA terms are added as long as the process is non-stationary.

We, now, know that the AR(1) process is stationary if $\phi < 1$, explosive if $\phi > 1$, and a so called *random walk* for $\phi = 1$, i.e.

$$x_t = x_{t-1} + a_t.$$

Including a constant θ_0, adds a drift to the random walk. Consequently, the process at time t is

$$x_t = x_0 + t \cdot \theta_0 + \sum_{i=0}^{t} a_{t-i}$$

with expected value $\mu_t = E(x_t) = x_0 + t\theta_0$ and variance $\gamma_{0,t} = V(x_t) = t \cdot \sigma^2$. The covariance can be written in the form

$$\gamma_{k,t} = Cov(x_t, x_{t-k}) = (t-k)\sigma^2, \ k \geq 0$$

yielding the correlation between a_t and x_{t-k}

$$\rho_{k,t} = \sqrt{\frac{t-k}{t}}$$

which approaches unity for increasing t. This renders the process very smooth yet still non-stationary. An example is given in figure 2.3.2. Generated are two random walk processes with $x_0 = 10$ in common as well as $a_t \sim \ iid \ N(0, 9)$. In the first part, the random walk has zero drift whereas in the second part, $\theta_0 = 2$. Notice that the two plots have basically nothing in common, any more.

Consider next *integrated* processes of which the random walk, $\nabla x_t = \theta_0 + a_t$, is one proxy. In general, the series may need d times first differencing to become stationary. It is then said to be integrated of order d. If this newly obtained series is governed by an ARMA(p, q) process

$$\phi(B)\nabla^d x_t = \theta_0 + \theta(B)a_t, \tag{2.21}$$

it is said to be an *autoregressive-integrated-moving average*, or ARIMA(p, d, q), process. In reality, d will take on values of, mostly, zero, one or, sometimes, two. Recall the ARMA(1,1) process

$$x - \phi x_{t-1} = a_t - \theta a_{t-1}$$

where we know the ACF to equal

$$\rho_1 = \frac{(1-\phi\theta)(\phi-\theta)}{1+\theta^2-2\phi\theta}$$

$$\rho_k = \phi\rho_{k-1}, \ k > 1.$$

Now, let ϕ approach one. Then, the resulting ARIMA(0,1,1) process

$$\nabla x_t = a_t - \theta a_{t-1}$$

has correlations all tending toward unity.

In (2.21), let $\theta_0 = 0$. We introduce a new quantity, w_t, such that

$$\phi(B)w_t = \theta(B)a_t \tag{2.22}$$

is a stationary and invertible ARMA process with $w_t = \nabla^d x_t$. To obtain x_t from this, we set $x_t = S^d w_t$ where S is the infinite summation

$$S = (1 + B + B^2 + \ldots) = (1 - B)^{-1} = \nabla^{-1}.$$

It is reffered to as the *integral* operator giving the name to the integrated process x_t because it it obtained by *integrating* w_t d times.

The x_t is said to be *homogeneous non-stationary* as mentioned in Box and Jenkins (1976). That is, many financial time series seem to be pretty much independent of the present level, x_t, in their local behavior. This is in contrast to AR(1) processes whose local behavior, for $|\phi| < 1$ and $|\phi| > 1$, depends stronly on the present levels, respectively. Hence, financial time series appear to behave locally homogeneously though they are non-stationary. To compensate for this discrepancy, consider that the ARMA process must have the property

$$\phi(B)(x_t + c) = \phi(B)x_t$$

implying that $\phi(B)c = 0$ which is equivalent to $\phi(1) = 0$. Hence, $\phi(B)$ can be factorized with a unit root such that

$$\phi(B) = \phi_1(B)(1 - B) = \phi_1(B)\nabla.$$

Consequently, (2.22) can be rewritten in the fashion

$$\phi_1(B)w_t = \theta(B)a_t$$

with $w_t = \nabla x_t$. In order to sustain homogeneous non-stationrity, w_t is required to not explode. Thus, $\phi_1(B)w_t$ is either stationary or there is an additional unit rrot in ϕ_1 such that $\phi_2(B) = \phi_1(B)(1-B)$ yielding $\phi_2(B)w_t^* = \nabla^2 x_t = \theta(B)a_t$ with stationary autoregressive part. If this is still not enough, it might be required to first difference up to d times. If the resulting process should be stationary, then, the autoregressive operators of homogeneous non-stationary series are required to be of the form $\phi(B)\nabla^d$. An example is given in figure 2.3.2. Here, $\nabla^2 x_t = a_t$ with $a_t \sim iid\ N(0, 2)$ and $x_0 = x_1 = 10$. Note how the level and slope alter randomly.

So far with $\theta_0 = 0$, the series driven by an ARIMA process displayed stochastic trends. We, now, introduce a deterministic trend by a non-zero drift parameter such that $\mu_t = E(x_t) = \beta_0 + \theta_0 t$. Here, $\beta_0 = x_0$. Including a constant in the model for, generally, dth differences results in a polynomial trend of order d. For $\theta_0 \neq 0$,

$$E(w_t) = E(\nabla^d x_t) = \nu_w = \frac{\theta_0}{1 - \phi_1 - \phi_2 - \ldots - \phi_p}$$

is different from zero. As a example, consider the series in figure 2.3.2. Here, $\nabla^2 x_t = 2 + a_t$ and, again, $a_t \sim\ iid\ N(0, 2)$ and $x_0 = x_1 = 10$. The trend of the original series is, now, of a quadratic type completely overshadowing the noise term.

As concluding remarks, when $\theta_0 = 0$, a stochastic trend is the consequence. On the other hand, if $\theta_0 \neq 0$, a polynomial in t of order d represents the deterministic trend. The resulting model, then, can be written as

$$\phi(B)\nabla^d x_t = \phi(1)\beta_d d! + \nabla^d \theta(B)a_t.$$

By the exactly d unit roots of the MA part, it is visible that the original noise of the process, x_t, is stationary. If this were not so, the noise would be non-stationary.

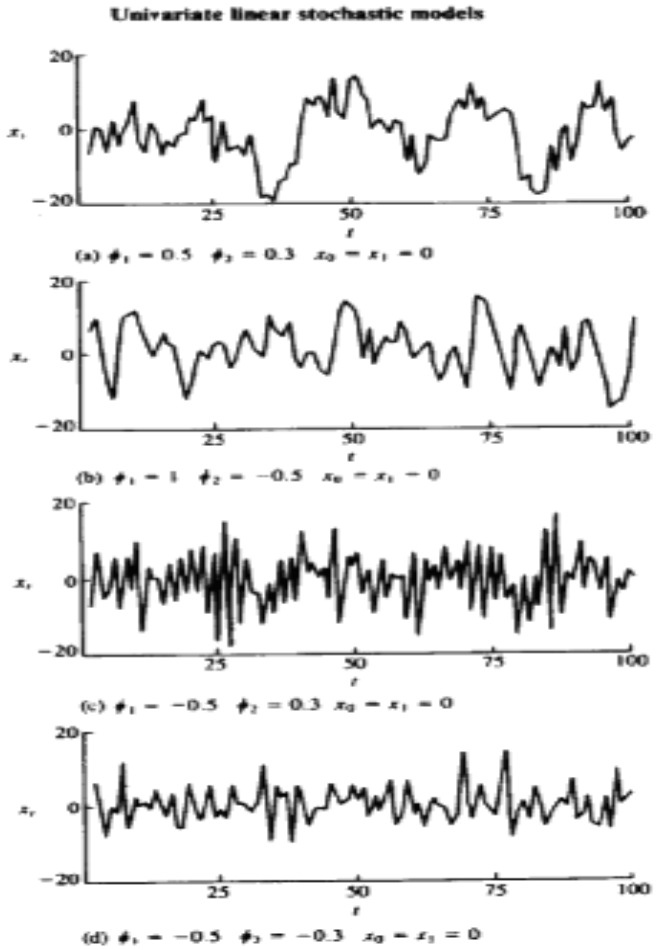


Figure 2.2: Simulations of various AR(2) processes.

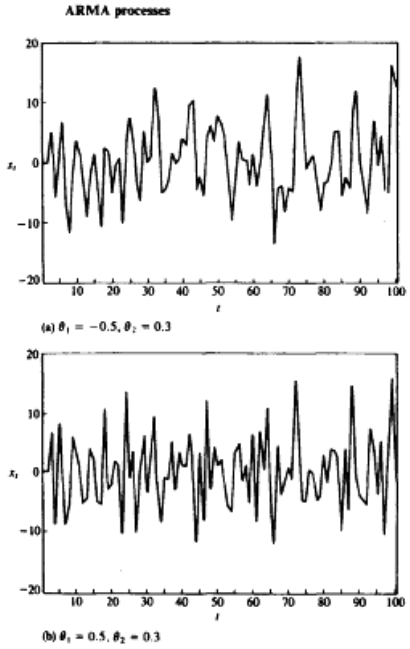


Figure 2.3: Simulations of MA(2) processes.

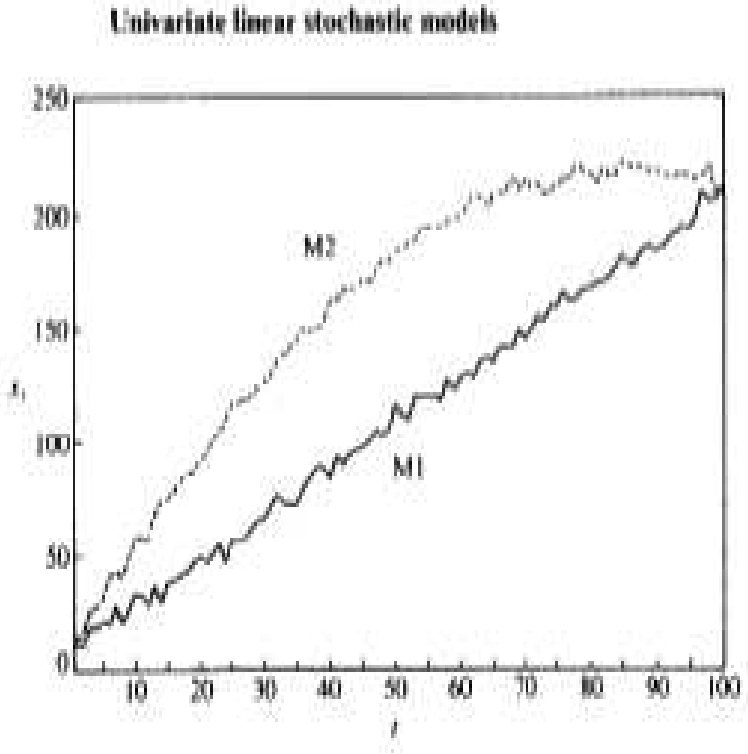


Figure 2.4: Linear and quadratic trends. $M1 : x_t = 10 + 2t + a_t$, $M2 : x_t = 10 + 5t - 0.03t^2 + a_t$. $a_t \sim NID(0, 9)$

Figure 2.5:

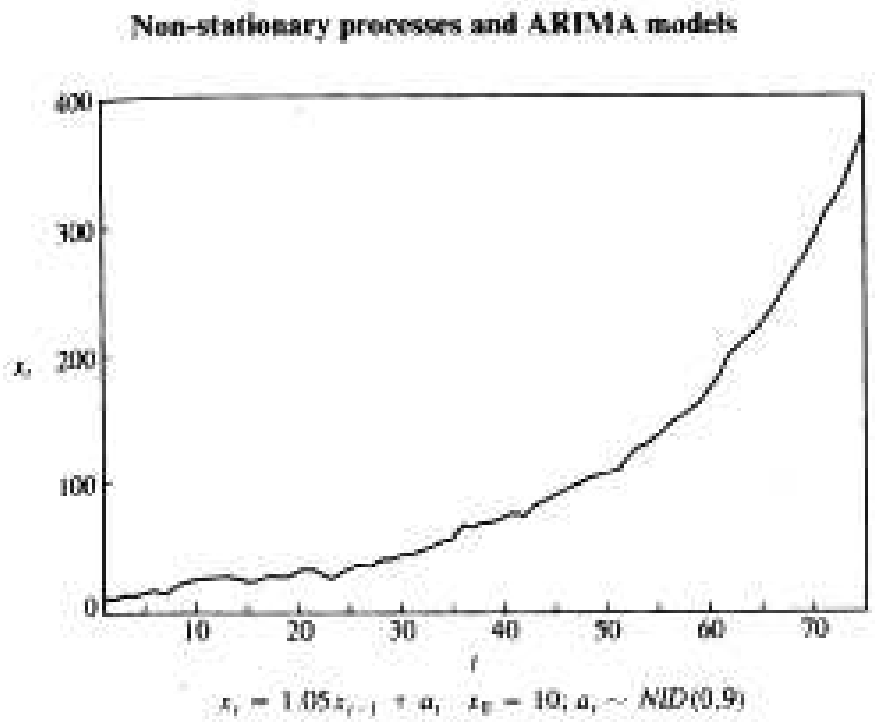
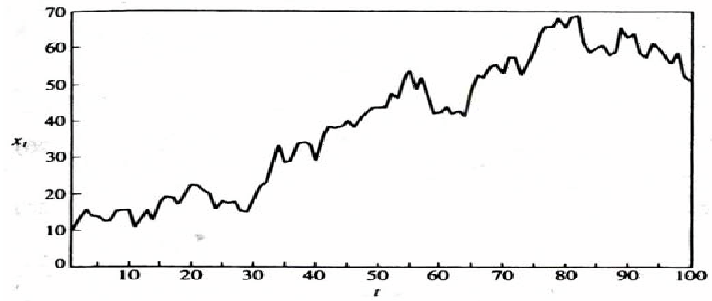
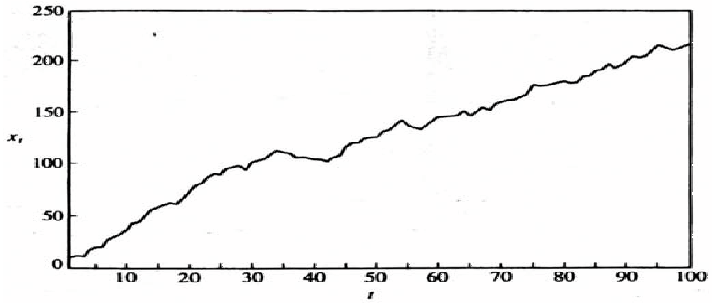


Figure 2.6: Explosive AR(1) model.

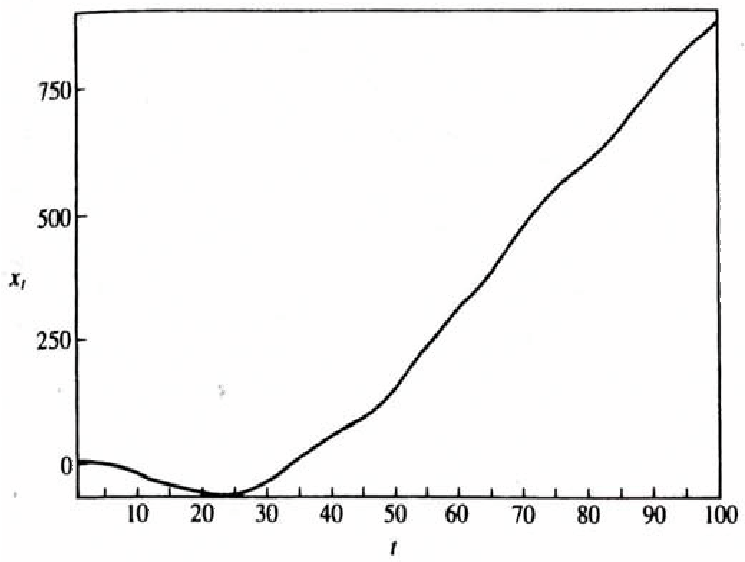


(a) $x_t = x_{t-1} + a_t$, $a_t \sim NID(0, 9)$, $x_0 = 10$



(b) $x_t = 2 + x_{t-1} + a_t$, $a_t \sim NID(0, 9)$, $x_0 = 10$

Figure 2.10 Random walks



$(1 - B)^2 x_t = a_t, x_0 = x_1 = 10, a_t \sim NID(0, 2)$

Figure 2.11 'Second difference' model

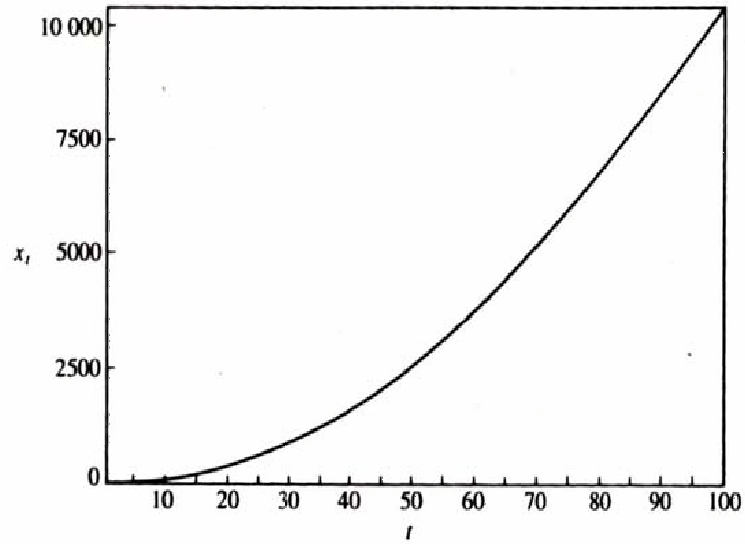


Figure 2.12 'Second difference with drift' model

Chapter 3

Modeling financial time-series and the corresponding data analyses (e.g. estimation, forecasting, testing).

3.1 ARMA model building

3.1.1 Estimation of ACF, PACF

In order to obtain a suitable ARMA model for the given time series, it is important to find estimates for the mean, μ, variance, σ^2, and autocorrelations, ρ_k. Since we assume that our data is stationary and ergodic, the sample mean,

$$\bar{x} = \frac{1}{T}\sum_{t=1}^{T} x_t,$$

and sample variance,

$$s^2 = \frac{1}{T}\sum_{t=1}^{T}(x_t - \bar{x})^2,$$

can serve as adequate estimates, respectively. For the ACF, the *sample autocorrelation function* or SACF is applied with the sample autocorrelation at lag k such as

$$r_k = \frac{\sum_{t=k+1}^{T}(x_t - \bar{x})(x_{t-k} - \bar{x})}{Ts^2},\ k = 1, 2, \ldots.$$

Consider indpendent observations from a population distributed with finite variance such that $\rho_k = 0,\ k \neq 0$. Then, the approximate variance of the r_k is given by T^{-1}. For large T, the $\sqrt{T}r_k \ \sim \ N(0,1)$ approximately. Confidence bounds are consequently provided by $2 \times T^{0.5}$. Thus, when the estimates ,r_k, exceed these bounds aboslutely, the corresponding lags k can be considered *significant*.

Consider next the generalization where the $\rho_k = 0$ for $k > q$. The variance of the corresponding r_k is then

$$V(r_k) = T^{-1}\left(1 + 2\rho_1^2 + \ldots + 2\rho_q^2\right) \tag{3.1}$$

for $k > q$. Replacing the ρ_j in (3.1) by the sample estimates, r_j, give the respective estimates of the variance of the $r_1, r_2, \ldots$ for increasing q, i.e. $T^{-1}, T^{-1}(1 + 2r_1^2), \ldots, T^{-1}(1 + 2r_1^2 + \ldots + 2r_{k-1}^2)$. Notice that the confidence bounds gradually widen as the lag order increases.

The *sample partial autocorrelation function* or SPACF at lags k equals the last estimated autocorrelation coefficients, $\hat{\phi}_{kk}$, respectively. If the true process is AR(p), then for $k > p$, the $\sqrt{T}\hat{\phi}_{kk} \ \sim \ N(0,1)$, approximately.

3.1.2 Model building

To provide methods for assessing the appropriateness of competing models for given series, we consider to approaches. The first is attributed to Box and

ARMA model building

k	r_k	$s.e.(r_k)$	$Q(k)$		$Q^*(k)$	
1	0.048	0.093	0.27	[0.61]	0.27	[0.61]
2	-0.160	0.093	3.21	[0.20]	3.32	[0.19]
3	0.096	0.096	4.28	[0.23]	4.44	[0.22]
4	-0.040	0.097	4.47	[0.35]	4.64	[0.33]
5	-0.053	0.097	4.79	[0.44]	4.98	[0.42]
6	0.014	0.097	4.81	[0.57]	5.00	[0.54]
7	0.139	0.097	7.04	[0.42]	7.42	[0.39]
8	-0.109	0.099	8.41	[0.39]	8.91	[0.35]
9	-0.024	0.100	8.47	[0.49]	8.99	[0.44]
10	0.051	0.100	8.77	[0.55]	9.31	[0.50]
11	-0.144	0.102	11.14	[0.43]	11.98	[0.36]
12	-0.097	0.102	12.22	[0.43]	13.21	[0.35]

Note: Figures in ... give $P(x_k^2 > Q(k), Q^*(k))$

Table 3.1: SACF of real S & P 500 returns and accompanying statistics.

Jenkins (1976). It suggests to match the obtained SCAFs and SPACFs with the theoretical ACFs and PACFs. Once a good fit is found, the respcetive ϕ_is, θ_is, and σs are computed.

The alternative approach is to select a set of possible (p, q) combinations and estimate the parametzers accordingly. Finally, the model is chosen for which a particular selection criterion attains its minimum. The procedure just presented is to be demonstrated in detail in the next few examples.

Example 3.1.1. This example has been given in Mills (1993). Consider the S&P 500 U.S. stock index for the years between 1872 and 1986. The plot in Figure 3.1 suggests annual returns with a constant mean at 8.21 percent. The overall picture of the graph of the annual returns gives the impression that the series might be stationary. The estimated autocorrelation coefficients,

ARMA model building

k	R_k	$s.e.(r_k)$	$\bar{\phi}_{kk}$	$s.e.(\phi_{kk})$
1	0.829	0.082	0.829	0.082
2	0.672	0.126	-0.048	0.082
3	0.547	0.148	0.009	0.082
4	0.435	0.161	-0.034	0.082
5	0.346	0.169	0.005	0.082
6	0.279	0.174	0.012	0.082
7	0.189	0.176	-0.116	0.082
8	0.154	0.178	0.114	0.082
9	0.145	0.179	0.047	0.082
10	0.164	0.180	0.100	0.082
11	0.185	0.181	0.028	0.082
12	0.207	0.182	0.038	0.082

Table 3.2: SACF and SPACF of UK spread

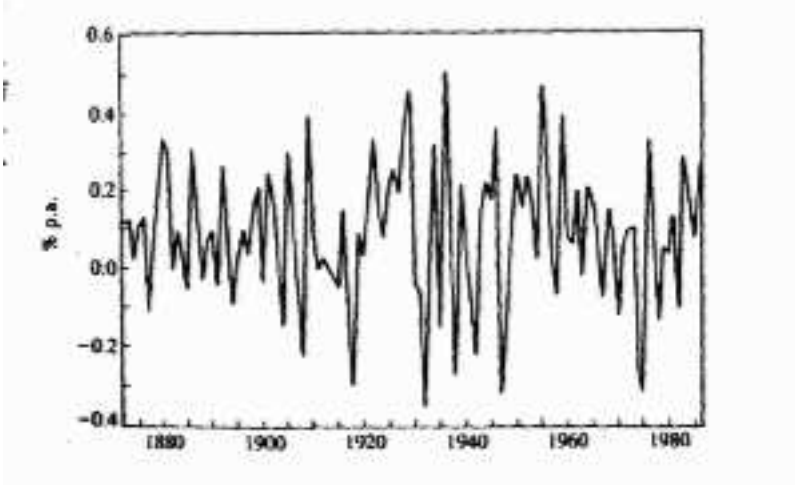


Figure 3.1: SACF of real S & P 500 returns.

r_k, with their standard errors from (3.1) obtained for various lags, k, even support the consideration that the returns might be white noise.

To verify the white noise assumption, a so called "Portmanteau" test statistic along with its improved version can be applied. The statistic is attributed to Box and Pierce. It is defined by

$$Q_1(k) = T \sum_{i=1}^{k} r_i^2. \tag{3.2}$$

When the series is white noise, $Q_1(k) \overset{\text{a}}{\sim} \chi^2(k)$ where $\overset{\text{a}}{\sim}$ indicates asymptotical distributional behavior. This estimator, (3.2) , however, was found, by simulation, to overestimate the significance levels. Hence, Ljung and Box (1978) presented the corrected statistic

$$Q_2(k) = T(T+2) \sum_{i=1}^{k} (T-i)^{-1} r_i^2. \tag{3.3}$$

Now, (3.3)$\overset{\text{a}}{\sim} \chi^2(k)$. The marginal significance levels can be retrieved from Table 3.1. Based on them, the null hypothesis of white noise bahvior of the S&P 500 annual returns cannot be rejected.

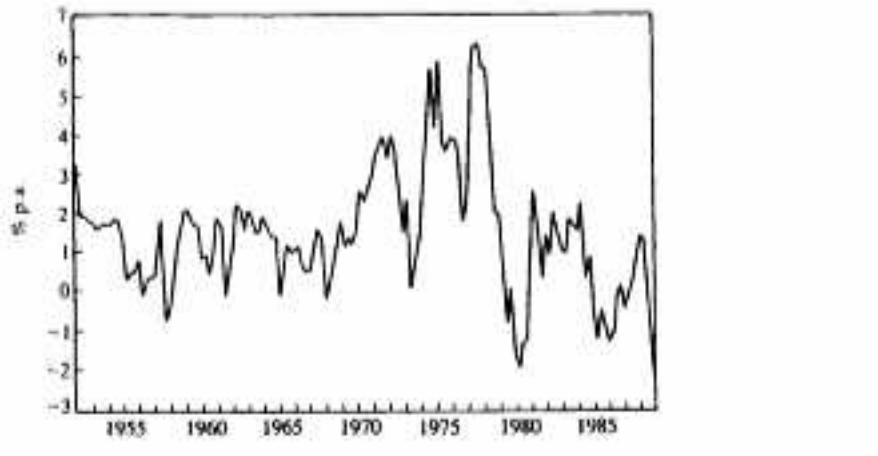


Figure 3.2: ARMA model building

Example 3.1.2. Consider next the UK interest rate spread between long and short interest rates which is important for the assessment of the structure of interest rates. The two interest rates are obtained from the 20 year UK gilts and the 91 day Treasury Bill, respectively. The series consists of quarterly observations between 1952 and 1988.

As can be seen from Table 3.2, the SACF and SPACF are given for lags up to $k = 12$. For $K = 1, \ldots, 5$, the r_k are significant. This is reflected in the plot of Figure 3.2 which is smoother than for white noise. One can tell by the Portemanteau statistics,(3.2) and (3.3), respectively, for lag $k = 12$ that the series is probably not white noise. Interestingly, the SACF declines up until lag $k = 9$ and increases thereafter. This might imply complex roots. Looking at the only significant value of the SPACF, $\hat{\phi}_{kk}$, an AR(1) appears to be appropriate. The estimated model is obtained from OLS regression as

$$\begin{gathered} x_t = 0.176 + 0.856 x_{t-1} + \hat{a}_t, \ \hat{\sigma}^2 = 0.870 \\ (0.098)\ (0.045) \end{gathered}$$

with standard errors given in parentheses. Now, to assess the correctness of

the model, test the white noise behavior of the newly obtained residuals, $\hat{a}_t$ by means of the Portemanteau statistics, (3.2) and (3.3), respectivly. Note, however, that the degrees of freedom need to be adjusted for an ARMA(p, q) to $k-p-q$ which is 11 in our case. The low values of 12.26 and 13.21 suggest that the model is adequately determined.

Instead of the Portemanteau test, one could overfit the model. Take, for example, an AR(2) or ARMA(1,1) structure. By means of OLS, this leads to the estimates

$$\begin{aligned} x_t &= 0.197 + 0.927x_{t-1} - 0.079x_{t-2} + \hat{a}_t, \ \sigma = 0.869 \\ &\quad (0.101)\ (0.054) \qquad (0.084) \\ x_t &= 0.213 + 0.831x_{t-1} + \hat{a}_t - 0.092\hat{a}_{t-1}, \ \hat{\sigma} = 0.870. \\ &\quad (0.104)\quad (0.051) \qquad\qquad (0.095) \end{aligned}$$

However, the t-statistics of the additional coefficients are insignificant in both equations, hence, rendering the AR(1) model the preferred one.

Example 3.1.3. The *Financial Times Actuaries All Share* index serves as the third and last example in this context. The period of monthly observations extends from 1965 through 1990. We obtain SACF and SPACF values as seen in Table 3.3. The $Q(12)$ is relatively low with 19.3, thus insignificant at levels of $p = 0.05$ and less. But the values of both, r_k and $\hat{\phi}_{kk}$, exceed their respective 5% critical values at lags $k = 1$ and 3. This calls for an ARMA structure with low p and q.

Now, the selection criteria announced previously come into play. Let us consider the Akaike's (1974) information criterion (AIC). It is defined as

$$AIC(p, q) = \ln \hat{\sigma}^2 + \frac{2}{T}(p + q).$$

Alternatively, we could use Schwarz's (1978) criterion which is

$$BIC(p, q) = \ln \hat{\sigma}^2 + \frac{\ln T}{T}(p + q).$$

ARMA model bulding

k	R_k	$s.e.(r_k)$	$\bar{\phi_{kk}}$	$s.e.(\phi_{kk})$
1	0.148	0.057	0.148	0.057
2	-0.061	0.059	0.085	0.057
3	0.117	0.060	0.143	0.057
4	0.067	0.061	0.020	0.057
5	-0.082	0.062	-0.079	0.057
6	0.013	0.062	0.034	0.057
7	0.041	0.063	0.008	0.057
8	-0.011	0.063	0.002	0.057
9	0.087	0.064	0.102	0.057
10	0.021	0.064	-0.030	0.057
11	-0.008	0.064	0.012	0.057
12	0.026	0.064	0.010	0.057

Table 3.3: SACF and SPACF of FTA ALL Share Nominal Returns

These are the two criteria we want to focus on despite the fact that there is a vast number of alternatives.

The most suitable model is then chosen by the (p, q) pair that minimizes the respective criterion. Formally, this can be written as

$$AIC(p^*_{AIC}, q^*_{AIC}) = min_{(p,q)} AIC(p, q),\ p \in P, q \in Q, \text{ or}$$
$$BIC(p^*_{BIC}, q^*_{BIC}) = min_{(p,q)} BIC(p, q),\ p \in P, q \in Q,$$

where $P = \{0, 1, \ldots, p^{max}\}$ and $Q = \{0, 1, \ldots, q^{max}\}$. There, however, is no unique rule on how to select the perfect p^{max} and q^{max}, respectively, to assure that the correct (p^0, q^0) are included. Even if inclusion be the case, there is no guarantee that (p^0, q^0) will be chosen based on the respective criterion.

A problem arises with the AIC in that the selected order, (p^*_{AIC}, q^*_{AIC}), will always be baised upwardly, hence, systematically leading to overparametrization. This is not the case with the BIC. Even though for $T \to \infty$, the (p^*_{BIC}, q^*_{BIC}) are such that $p^*_{BIC} \geq p^0$ and $q^*_{BIC} \geq q^0$, with probability one, the true model will eventually be determined due to the BIC's strong consistency. This is a resulty of the additional factor, $\ln T$ variable in T which is missing in the AIC.

The SACF and SPACF give rise to the assumption that adequate orders might be provided by setting $P = Q = 3$. Then, Table 3.4 shows the corresponding AIC and BIC values, respectively. The best order according to the AIC is (3,0) so that the resulting process is an AR(3) process. On the other hand, the preferable lag orders chosen by the BIC are (0,1) so that the proces is MA(1). Estimating the coefficients yields the following models

$$\begin{array}{ccccccc} x_t = & 4.41 & + 0.173 x_{t-1} & - 0.0108 x_{t-2} & + 0.144 x_{t-3} & + \hat{a}_t, & \hat{\sigma} = 6.25 \\ & (0.62) & (0.057) & (0.057) & (0.057) & & \end{array}$$

$$\begin{array}{ccccc} x_t = & 5.58 & + \hat{a}_t & + 0.186 \hat{a}_{t-1}, & \hat{\sigma} = 6.29 \\ & (0.35) & & (0.056) & \end{array}$$

Univeriate linear stochastic models

p \ q	0	1	2	3
AIC				
0	3.701	3.684	3.685	3.688
1	3.689	3.685	3.694	3.696
2	3.691	3.695	3.704	3.699
3	3.683	3.693	3.698	3.707
BIC				
0	3.701	3.696	3.709	3.724
1	3.701	3.709	3.730	3.744
2	3.715	3.731	3.752	3.759
3	3.719	3.741	3.758	3.779

Table 3.4: Model selection criteria for nominal returns

In the AR(3) model, the estimated mean is obtained from $4.41/(1-0.173+0.0108-0.144)$ to be equal to 5.58 which is the mean from the MA(1) model. Beause the unit roots of the MA(1) polynomial, $(1+0.186B)$, is outside the unit circle, it can be inverted to yield the AR polynomial

$$\begin{aligned}(1+0.186)^{-1} &= (1-0.186B+0.186^2B^2-0.186^3B^3+\ldots)\\ &\approx (1-0.186B+0.035B^2-0.006B^3)\end{aligned}$$

indicating an autoregressive structure which is exponentially declining rather than a limitation to three lags. Considering $(1+0.186)^{-1}=0.843$, though, which roughly is equal to the sum of the coefficients of the AR(3) model, $\Phi(1)$, both models appear equivalently suitable. This is confirmed by diagnostic checks with $Q_2(12)=6.07$ and 15.30 which corresponds to significance levels of 0.91 and 0.23, respectively. Based on this finding, the AIC chosen model may appear preferable.

We generate a portfolio of models based on several pairs, (p, q). Using the AIC, for example, we define a new statistic

$$\mathfrak{R} = \exp\left\{-\frac{1}{2}T(AIC(p^*_{AIC}, q^*_{AIC}) - AIC(p, q))\right\}$$

which compares alternative models to the AIC selected one. $\mathfrak{R}$ then provides a classififcation of alternative models with respect to ARMA(p^*_{AIC}, q^*_{AIC}). According to Poskitt and Tremayne (1987), a value of $\mathfrak{R}$ less $\sqrt{10}$ qualifies the portfolio ARMA(p, q) as a close competitor to ARMA(p^*_{AIC}, q^*_{AIC}). With this rule, based on the AIC, this yields six competitors, $(3, 0), (0, 1), (0, 2), (1, 1)$, $(0, 3)$, and $(1, 0)$ whereas, based on the BIC, we obtain $(0, 1), (1, 0)$, and $(0, 0)$.

We find that the theoretical properties of the AIC and BIC, respectively, are reflected in these results. Even though the AIC has the tendency to overparameterize, an autoregressive structure of order three does not seem too implausible for financial returns if one takes into consideration effects from, for example, dividends.

3.2 Integrated model building

Let us have a look at what happens when it is deemed necessary to approach the not always easy task of first-differencing the series, possibly $d > 0$ times, until we obtain $w_t = \delta^d x_t$ which we believe is stationary. In the following, we will assume that the series has been adequately first-differenced. Nothing stops us from parameter estimation in the fashion we were introduced to in the previous section, now. A few examples considering ARIMA processes are following next.

Example 3.2.1. In this example, we, once again, consider the spread of UK intrest rates from the previous example. Let us assume, however, that it is $I(1)$, hence, we are treating the autocorrelation structure of $w_t = \delta x_t$. From

ARIMA modelling

k	R_k	$s.e.(r_k)$	$\bar{\phi_{kk}}$	$s.e.(\phi_{kk})$
1	0.002	0.082	0.002	0.082
2	-0.036	0.082	-0.036	0.082
3	-0.040	0.082	-0.040	0.082
4	-0.090	0.082	-0.092	0.082
5	-0.090	0.084	-0.094	0.082
6	0.063	0.085	0.054	0.082
7	-0.192	0.086	-0.211	0.082
8	-0.068	0.092	-0.085	0.082
9	-0.101	0.093	-0.143	0.082
10	-0.038	0.094	-0.076	0.082
11	-0.002	0.094	-0.065	0.082
12	0.173	0.094	0.097	0.082

Table 3.5: SACF and SPACF of the first difference of the UK interest rate spread

the portmanteau statistic $Q_2(12) = 16.74$, we cannot reject the white noise hypothesis for the sample. In Table 3.5, we have the estimates of r_k and $\hat{\phi}_{kk}$ along with their respective standard errors, respectively. The mean is found insignificant rendering the series driftless.

Example 3.2.2. Next, we, once again, focus on the U.S. Dollar/Sterling exchange rate. Weekly observations of the rates themselves as well as the first differences can be seen in Figure 3.3.

Recall that the period is January 1980 through December 1988 generating 470 observations. It becomes quite evident from the plot that the first differences appear to be zero mean, stationary innovations of the rates behaving like a driftless random walk. When analyzing the SACF and SPACF of the differences, respectively, one finds that they ressemble white noise. The first

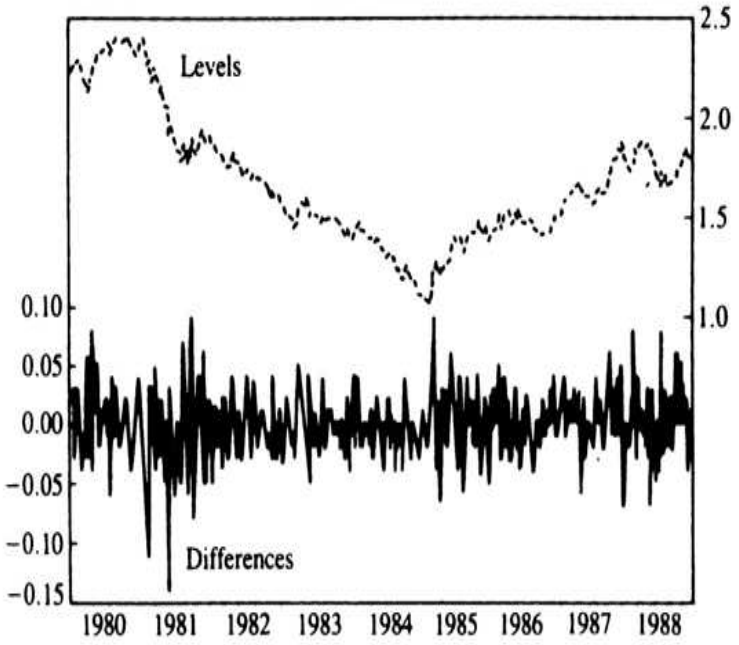


Figure 3.3: Dollar/sterling exchange rate weekly (1980-8)

24 entries for SACF and SPACF are all insignificant. The portmanteau statistic of $Q_2(24) = 20.3$ supports this hypothesis. The insignificant estimate of the mean equal -.0009 and according standard error of .0013 reassures that the series, in fact, is driftless.

Example 3.2.3. Finally, we are analyzing the observations from the FTA All Share index of the period from January 1965 to December 1990. As can be seen from Figure 3.2.3,

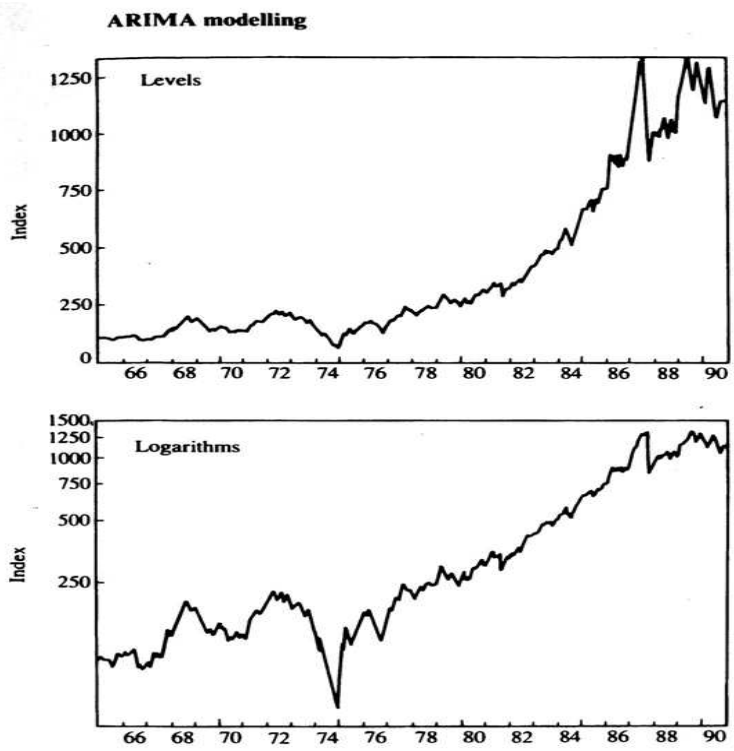


Figure 3.4: FTA All Share index (monthly 1965-90)

k	R_k	$s.e.(r_k)$	$\bar{\phi_{kk}}$	$s.e.(\phi_{kk})$
1	0.113	0.57	0.113	0.057
2	-0.103	0.58	-0.117	0.057
3	0.093	0.60	0.122	0.057
4	0.061	0.61	0.021	0.057
5	-0.102	0.61	-0.092	0.057
6	0.036	0.62	-0.011	0.057
7	0.044	0.62	0.020	0.057
8	-0.047	0.63	-0.046	0.057
9	0.075	0.63	0.115	0.057
10	0.021	0.064	-0.031	0.057
11	-0.041	0.064	-0.020	0.057
12	0.019	0.064	0.023	0.057

Table 3.6: SACF and SPACF of the first difference of the FTA ALL Share Index

the series appears to have a well pronounced upward drift. It is however, not linear. The noise accompanying the series seems to be increasing in absolute value as the level of the index becomes higher. When taking logarithms, the transformed series displays a linear trend accompanied by innovations with constant variance.

The trend is then removed by first-differencing. The corresponding SACF and SPACF, respectively, can be seen in Table 3.6.

No real structure can be inferred from the r_k. But the $\hat{\phi}_{kk}$ hint at an AR(3) process. This assumption is supported by residual diagnostic tests together with overfitting. The model, then, is given by

$$\nabla x_t = \underset{(0.0038)}{0.0067} + \underset{(0.057)}{0.141}\nabla x_{t-1} - \underset{(0.057)}{0.133}\nabla x_{t-2} + \underset{(0.057)}{0.123}\nabla x_{t-3} + \hat{a}_t,\ \hat{\sigma} = 0.0646.$$
$$Q(12) = 7.62$$

The monthly ∇x_t have estimated mean of 0.0077 which translates into an annual growth rate of 0.0925.

3.3 Forecasting

In this section, we consider forecasting values x_{T+h} for the ARIMA(p, d, q) process

$$\phi(B)\nabla^d x_t = \theta_0 + \theta(B)a_t$$

with observations $\{x_t\}_{1-d}^{T}$. We, now, introduce

$$\alpha(B) = \phi(B)\nabla^d = (1 - \alpha_1 B - \alpha_2 B^2 - \ldots - \alpha_{p+d}B^{p+d}).$$

We denote the *minimum mean square error* (MMSE) of x_{T+h} made at T by $f_{T,h}$ which is computed as

$$f_{T,h} = E(\alpha_1 x_{T+h-1} + \alpha_2 x_{T+h-2} + \ldots + \alpha_{p+d}x_{T+h-p-d} + \theta_0 + a_{T+h} - \theta_1 a_{T+h-1} - \ldots - \theta_q a_{T+h-q} | x_T, x_{T-1}, \ldots).$$

It is

$$E(x_{T+j}|x_T, x_{T-1}, \ldots) = \begin{cases} x_{T+j}, & j \leq 0 \\ f_{T,j}, & j > 0 \end{cases}$$

and

$$E(a_{T+j}|x_T, x_{T-1}, \ldots) = \begin{cases} a_{T+j}, & j \leq 0 \\ 0, & j > 0. \end{cases}$$

In words, this means that for the evaluation of $f_{T,h}$, the expected values of values lying in the past are replaced by the observed values x_{T+j} and a_{T+j}, respectively, and the future expected values are replaced by the forecast values, $f_{T,j}$ and 0, respectively.

Illustration of this is provided by three examples. First, consider the AR(1) model $(a - \phi B)x_t = \theta_0 + a_t$. We, now, have $\alpha(B) = (1 - \phi B)$ so that

$$x_{T+h} = \phi x_{T+h-1} + \theta_0 + a_{T+h} \tag{3.4}$$

which in return provides us with

$$f_{T,h} = \phi f_{T,h-1} + \theta_0 \tag{3.5}$$

$$= \phi^h x_T + \theta_0(1 + \phi + \phi^2 + \ldots + \phi^{h-1}). \tag{3.6}$$

Letting $h \to \infty$ and assuming stationarity by $|\phi| < 1$, we obtain

$$f_{T,h} = \frac{\theta_0}{1 - \phi} = E(x_t) = \mu. \tag{3.7}$$

This tells us that for estimation of values in the far future, the best forecast is the mean of the process.

As a second example, we focus on an ARIMA(0,1,1) model with $\nabla x_t = (1 - \theta B)a_t$ so that $\alpha(B) = (1 - B)$. This gives

$$x_{T+h} = x_{T+h-1} + a_{T+h} - \theta a_{T+h-1}.$$

When $h = 1$,

$$f_{T,1} = x_T - \theta a_T.$$

For $h = 2$,

$$f_{T,2} = f_{T,1} = x_T - \theta a_T.$$

For a general h, we obtain

$$f_{T,h} = f_{T,h-1},\ h > 1.$$

As a consequence, forecasts at time T for any period $T + h$ in the future are of value $f_{T,1}$.

Alternatively, from

$$f_{T,h} = x_T - \theta a_T$$
$$a_T = (1 - B)(1 - \theta B)^{-1} x_T,$$

we obtain for the h-step ahead forecast

$$f_{T,h} = (1 - \theta)(1 - \theta B)^{-1} x_T$$
$$= (1 - \theta)(x_T + \theta x_{T-1} + \theta^2 x_{T-2} + \ldots).$$

This is an exponentially weighted moving average of all known values of the process.

As the third example, consider the ARIMA(0,2,2) model $\nabla^2 x_t = (1 - \theta_1 B - \theta_2 B^2) a_t$ with $\alpha(B) = (1 - B)^2 = (1 - 2B + B^2)$ such that

$$x_{T+h} = 2x_{T+h-1} - x_{T+h-2} + a_{T+h} - \theta_1 a_{T+h-1} - \theta_2 a_{T+h-2}.$$

Now, for $h = 1$,

$$f_{T,1} = 2x_T - x_{T-1} - \theta_1 a_T - \theta_2 a_{T-1}.$$

For $h = 2$,

$$f_{T,2} = 2f_{T,1} - x_T \theta_2 a_T,$$

and for $h = 3$,

$$f_{T,3} = 2f_{T,2} - f_{T,1}.$$

For a general $h \geq 3$, we have, finally,

$$f_{T,h} = 2f_{T,h-1} - f_{T,h-2}.$$

This gives a straight line through $f_{T,1}$ and $f_{T,2}$ as loci for all forecasts made at time T with errors

$$e_{T,h} = x_{T+h} - f_{T,h} = a_{T+h} + \Psi_1 a_{T+h-1} + \ldots + \Psi_{h-1} a_{T+1}.$$

The $\Psi_1, \ldots, \Psi_{h-1}$ are obtained from $\Psi(B) = \alpha^{-1}(B)\theta(B)$ which yields a variance of the forecast error equal to

$$V(e_{T,h}) = \sigma^2(1 + \Psi_1^2 + \Psi_2^2 + \ldots + \Psi_{h-1}^2).$$

This is inline with intuition in the sense that the uncertainty expressed by the variance is a composition of the unforeseeable errors or shocks between periods T and $T + h$. In particular, for $h = 1$, the one-step ahead forecast error is

$$e_{T,1} = x_{T,1} - f_{T,1} = a_{T+1}.$$

This shows that the one-step ahead errors of a MMSE forecast have to be uncorrelated. This, however, is not generally the case for h-step ahead forecasts made at different times T or forecasts of varying lead times, h.

Considering the AR(1) model from before, the variance of the h-step ahead forecast is given by

$$\begin{aligned} V(e_{T,h}) &= \sigma^2(1 + \psi_2 + \psi_4 + \ldots + \psi^{2(h-1)}) \\ &= \sigma^2(1 + \phi^2 + \phi^4 + \ldots + \phi^{2(h-1)}) \\ &= \sigma^2 \frac{(1 - \phi^{2h})}{(1 - \phi^2)} \end{aligned}$$

due to the fact that, here, $\psi_j = \phi^j$. For $h \to \infty$, the variance approaches a constant from below which is $\sigma^2/(1 - \phi^2)$. This quantity expresses the expected variation of the process about the infinite time forecast, μ.

The ARIMA(0,1,1) model has the h-step ahead forecast error variance

$$V(e_{T,h}) = \sigma^2(1 + (h-1)(1-\theta)^2)$$

where $\psi_j = 1 - \theta,\ j = 1, 2, \ldots$. This variance is linearly increasing in h. For the ARIMA(0,2,2) with weights $\psi_j = 1 + \theta_2 + j(1 - \theta_1 - \theta_2),\ j = 1, 2, \ldots$, the variance of the h-step ahead forecast error is

$$V(e_{T,h}) = \sigma^2(1 + (h-1)(1+\theta_2)^2 + 6^{-1}h(h-1)(2h-1)(1-\theta_1-\theta_2)^2) \\ +h(h-1)(1+\theta_2)(1-\theta_1-\theta_2).$$

Thus, this variance, too, is increasing in h.

Using a few examples, we will present the dependence structure of successive forecasts and their error variances for varying orders of integration.

First, let us have a look at the previous AR(1) model of the UK interest rate spread with the parameter estimates $\hat{\phi} = 0.856$, $\hat{\theta}_0 = 0.176$, and $\hat{\sigma} = 0.870$. The spread at T is observed to be $x_T = -2.27$. We, consequently, obtain the h-step forecast being

$$f_{T+h} = 0.856^h(-2.27) + 0.176(1 + 0.856^2 + 0.856^4 + \ldots + 0.856^{2(h-1)}).$$

In the limit, this is 1.23 which is equal to the sample mean of the spread. The corresponding forecast error variance is

$$V(e_{T,h}) = 0.870^2(1 + 0.856^2 + 0.856^4 + \ldots + +0.856^{2(h-1)}).$$

This approaches the constant $0.870^2/(1 - 0.856^2) = 2.832$.

Consider next the random walk ARIMA(0,1,1) model of the spread. Recall that $\theta = 0$ and $\hat{\sigma} = 0.894$, in this model. The forecasts are then, for all h, $f_{T+h} = -2.27$. This deviates from the sample mean. The corresponding forecast error variance is $V(e_{T,h}) = \sigma^2 h = 0.80h$ which is increasing in h. The differences between autoregressive models with and without unit roots with respect to their respective forecasts have become obvious, now.

A random walk model was also fit to the U.S. Dollar/Sterling exchange rate. We know that we have a constant forecast for any projection period h. It is the value of $x_T = 1.81$. But the according variance is, of course, increasing in h. More exactly, it is the linear function $V(e_{T,h}) = 0.0073h$ with $\hat{\sigma} = 0.027$.

The logarithms of the FTA All Share index were fitted by the ARIMA(3,1,0) model

$$\phi(B) = (1 - 0.141B + 0.133B^2 - 0.123B^3).$$

This yields

$$\alpha(B) = (1 - 1.141B + 0.274B^2 - 0.256B^3 + 0.123B^4).$$

The forecasts, then, are by the recursion

$$\begin{aligned}
f_{T,1} &= 1.141x_T - 0.274x_{T-1} + 0.256x_{T-2} - 0.123x_{T-3} + 0.0067,\\
f_{T,2} &= 1.141f_{T,1} - 0.274x_T + 0.256x_{T-1} - 0.123x_{T-2} + 0.0067,\\
f_{T,3} &= 1.141f_{T,2} - 0.274f_{T,1} + 0.256x_{T-1} - 0.123x_{T-3} + 0.0067,\\
f_{T,4} &= 1.141f_{T,3} - 0.274f_{T,2} + 0.256f_{T,1} - 0.123x_{T-3} + 0.0067,
\end{aligned}$$

and

$$f_{T,h} = 1.141f_{T,h-1} - 0.274f_{T,h-2} + 0.256f_{T,h-3} - 0.123f_{T,h-4} + 0.0067,$$

for $h \geq 5$. The polynomial

$$\Psi(B) = 1 + 1.141B + 1.028B^2 + 1.229B^3 + 1.392B^4 + \dots \tag{3.8}$$

together with $\hat{\sigma} = 0.646$ suffices to compute the according forecast error variance which is increasing in h.

In the retrospective at time $T + h$, the forecast structure yields

$$\begin{aligned}
x_{T+h} - 1.141x_{T+h-1} + 0.274x_{T+h-2} - 0.256x_{T+h-3} + 0.123x_{T+h-4}\\
= 0.0067 + a_{T+h}.
\end{aligned}$$

At time T, this difference equation yields

$$x_{T+h} = \sum_{i=1}^{4} b_i^{(T)} f_i(h) + 0.0067 \sum_{j=T+1}^{T+h} \psi_{T+h-j}.$$

We obtain ψ from (3.8). The $f_i(h),\ i = 1, \ldots, 4$ are the roots of the polynomial $\alpha(B)$, respectively, each with exponent h. The roots are $1, 0.45$ and $(0.15 \pm 0.50i)$. This provides us with

$$x_{T+h} = b_0 + b_1^{(T)} + b_2^{(T)}(0.45)^h$$
$$+b_3^{(T)}(0.15 + 0.50i)^h + b_4^{(T)}(0.15 - 0.50i)^h \text{ where}$$
$$b_0 = 0.0067 \sum_{j=T+1}^{T+h} \psi_{T+h-j}.$$

At time T, the b_i are constants for any horizon h. They depend on time T through the observed x_t, though.

A forecasted trend of the process is given by $b_0 + b_1^{(T)}$. This is due to the dependence of ψ and, hence, b_0 on h. Additionally, there are a geometrically declining term, $b_2^{(T)}(0.45)^h$ and damped sine waves provided by the two complex roots and their respective damping factors along with frequency and phase all being functions of the process parameters.

Chapter 4

Non-linearity, Trends and Volatility in financial data

4.1 Non-linear dynamics and chaos

In economic systems, the fluctuations have been assumed for a long time to be the result of exogenous shocks of stochastic but stationary nature. On the other hand, there are attempts noticeable to consider business cycles to be caused by an endogenous nonlinear deterministic structure.

This section is intended to deal with series generated by deterministic but non-linear movements. The uniqueness of series driven by processes following those laws is that they appear stochastic throughout statistical tests even though they are not. A suitable term is 'deterministic chaos'.[1] For an example from Mills (1997), we consider the difference equation

$$x_t = f(x_{t-1}),\ x_0 \in [0, 1]$$

[1] See Mills (1997).

such that

$$f(x) = \begin{cases} x/\alpha & x \in [0, \alpha] \\ \frac{1-x}{1-\alpha}, \ x \in [\alpha, 1], \ 0 < \alpha < 1 \end{cases} .$$

Samples generated by this process will most likely have ACFs similar to an AR(1) process $x_t = (2\alpha - 1)x_{t-1} + u_t$ rendering the process indiscernibly different from white noise for $\alpha = .5$. Keep in mind that our process is purely deterministic, though.

For considering models in the financial world which this type of process might be applicable to, we look at the dynamics of the dividend process that are being passed on to the equilibrium asset prices. The dynamics can be either of linear, non-linear or even chaotic nature.

The tests for nonlinearity to be presented in the following have no order with respect to domination of competitors. This depends on the fact that the different tests may perform better against some alternative models than others. Also, a clear definition of the null hypothesis is essential for comparison of the tests with respect to power.

4.2 Testing for non-linearity

Up to now, the empirical tests presented are designed to linear structures. But becausse of the increase in interest in deterministic, nonlinear chaotic systems, the need for suitable tests in this respect are called for, that is there ought to be tests for chaos and identification methods for the nonlinear deterministic system generator.

It is alleged in literature that it is often difficult to determine the nature of deviation from linearity. Many tests are just designed to reject or accept the hypothesis of linearity. An alternative for linearity is not offered.

Certain tests are apt to reliably reject the null hypothesis H_0 when the true model is of a particular non-linear moving average type but might pos-

sibly fail to reject H_0 if the true process obeys dynamics with conditional variance, for example.

4.2.1 Linear expansion test

A representative of such a test could be of the form of the expansion

$$x_t = \mu + \sum_{i=-\infty}^{\infty} \psi_i a_{t-i} + \sum_{i,j=-\infty}^{\infty} \psi_{ij} a_{t-i} a_{t-j} + \sum_{i,j,k=-\infty}^{\infty} \psi_{ijk} a_{t-i} a_{t-j} a_{t-k} + \dots$$

where x_t can be rejected as linear when any of the coefficients on the right side where the summands incorporate products of at least two error terms are non-zero. This test is based on cascading regressions of x_t including their cross-products. Several extensions and modifications of this test exist.

4.2.2 Correlation dimension test

The next test to be presented, here, applies the correlation integral defined as[2]

$$C_N = (l, T) = \frac{2}{T_N(T_N - 1)} \sum_{t<s} I_l \left(x_t^N, x_s^N \right). \tag{4.1}$$

where $T_N = T - N + 1$. The one-dimensional

$$x_t^N = (x_t, x_{t+1}, \dots, x_{t+N-1}),$$
$$x_s^N = (x_s, x_{s+1}, \dots, x_{s+N-1}),$$

are called *m-history* with *embedding dimension m.* [3] Further,

$$I_l(x_t^N, x_s^N) = \begin{cases} 1, & ||x_t^N - x_s^N|| < l \\ 0, & \text{else.} \end{cases}$$

[2] E.g., see Mills (1997).

[3] See, for example, Barnett and Serletis (2000).

where $||\cdot||$ denotes the supremum-norm.

In words, (4.1) gives an estimate of the probability that any two sets, $x_t^N and x_s^N$, of length N are within l of each other. Then, in case of existence of the limit, the so called *correlation dimension* of $\{x_t\}$ is defined as

$$v = \lim_{l \to 0} [\ln C_N(l,T) / \ln l].$$

Intuitively, if v is small, one can assess that the process is driven by a chaotic deterministic system in contrast to a stochastic one. However, there is no theoretical support of this notion as of now.

Now, letting $N \to \infty$, we want to know how v changes. If v continuously increases, then the system is stochastic.[4] In the opposite case, that is for a finite limit, then the series stems from a deterministic generating process.

Sampling properties of the correlation dimension are unknown, though. This impairs the testing power of this statistic when there are truely stochastic components in the system. The correlation dimension then becomes stochastic itself and, hence, one needs to know its distribution to use it effectively. As is argued in literature[5], the series has to have a correlation dimension below $2 \log_{10} T_N$ for one to be able to distinguish a chaotic system.

4.2.3 BDS test

In 1996, a test was presented which was devised by Brock, Dechert, and Scheinkman, henceforth BDS test. It tests the null hypothesis of *iid* x_t against an alternative of some sort. The test is nonparametric. Under the null, the BDS statistic

$$w_N(l,T) = \sqrt{T}[C_N(l,T) - C_1(l,T)^N] / \hat{\sigma}_N(l,T)$$

[4] See, for example,Barnett and Serletis (2000).

[5] For example, Barnett and Serletis (2000).

is asymptotically $N(0,1)$ distributed. The $\hat{\sigma}_N^2(l,T)$ is an estimate of the standard deviation of $C_N(l,T) - C_1(l,T)^N$.[6]

Motivation for the test results from the fact that $C_N(l,T)$ is an estimate of the probability that the two N-histories, x_t^N and x_s^N, are less than l apart. In case of *iid* x_t^N and X_s^N, this probability would be equal to $C_N(l,T) = C_1(l,T)^N$.

Under the null of whiteness, the asymptotic distribution of the BDS statistic is known. Thus, the statistic offers a test against any sort of dependence including nonwhite linear and nowhite nonlinear dependence. That is, the alternative could be a linear as well as a non-linear stochastic system or non-linear deterministic system. It, however, is not adequate for directly testing for chaos. This test is considered diagnostic in nature since the rejection of the H_0 that the series is *iid* can be a result of several cases of dependence structures. Whatever the specific cause for the rejection might be has to be determined in further testing.

4.2.4 The NEGM test

For this test, we introduce the dominant Lyapunov exponent, λ. It measures the average exponential divergence between two trajectories that, at some initial time, started from points with infinitesimally small distance between them. For the distance in a certain direction between the possibly multi-dimensional trajectories, the Lypunov exponent is defined as

$$\sigma = \lim_{t \to \infty} \frac{1}{t} \ln |d(t)|$$

where $d(t)$ is the average distance between the two trajectories at time t.

Operationally, chaotic behavior can be defined by a bounded system with positive Lyapunov exponent. An important characteristic of chaotic systems

[6]For a more detailed account of the properties of the estimator, we refer the reader to the reference list given in Mills (1997).

is the dependence on initial conditions.

We assume real valued $\{x_t\}$ generated by a nonlinear autoregressive model of the form

$$x_t = f(x_{t-L}, x_{t-2L}, \ldots, x_{t-NL}) + \epsilon_t,\ t \in \{1, \ldots, N\}.$$

N is the length of the autoregression and L is the lag parameter. f is a smooth unknown funtion and the ϵ are independent with zero mean and constant variance.

For the estimation of the Lyapunov exponent, we set up the state-space form of (4.2.4)

$$X_t = F(X_{t-L}) + E_t,\ F : \mathbb{R}^N \to \mathbb{R}^N.$$

Here, $X_t = (x_t, x_{t-L}, \ldots, x_{t-NL})',\ F(X_{t-L}) = (f(x_t, x_{t-L}, \ldots, x_{t-NL}), \ldots$ $\ldots, x_{t-L}, \ldots, x_{t-NL})'$, and $E_t = (\epsilon_t, 0, \ldots, 0)'$.

An intermediate step is to estimate the individual Jacobian matrices

$$J_t = \frac{\partial F(X_t)}{\partial X'}.$$

This is suggested to be done bythin plate splines or neural nets. The latter method involves a hidden layer with q units such that

$$f(X_{t-L}, \theta) = \beta_0 + \sum_{j=1}^{q} \beta_j \psi \left(\gamma_{0j} + \sum_{i=1}^{N} \gamma ij x_{t-iL} \right).$$

ψ is a known nonlinear function[7]. The parameter vector θ is fit to the data through nonlinear least squares, see Barnett and Serletis (2000), by computing

$$\min_{\theta} S(\theta) = \sum_{t=1}^{T} [x_t - f(X_{t-1}, \theta)]^2.$$

[7]Commonly, it is the logistic distribution function $\psi(u) = 1/(1 + \exp(-u))$.

The Bayesian Information Criterion (BIC) is suggested to minimize via the triple (L, N, q) to obtain estimate values for L, N, and q.

With the estimate $\hat{J}$, we can write $\hat{\Gamma} = \hat{J}_T \dots \hat{J}_1$ to obtain $\hat{v}_1(T)$ as the largest eigenvalue of the matrix $\hat{\Gamma}'_T \hat{\Gamma}_T$. The estimate of the dominat Lyapunov exponent is then

$$\hat{\lambda} = \frac{1}{2T} \ln |\hat{v}_1(T)|.$$

4.2.5 The White test

The motivation for the test is that the residuals produced by linearly filtering the process should be uncorrelated with any measurably function of the process' history.

The test uses a feed-forward neural network with one hidden-layer to fit the series.[8] The neural network is fitted to find a measurable function of the process' history. The test also fits an AR model to the series which is uncorrelated to the function under H_0. Any nonlinear structure in the residuals from the linear filter the test is designed to detect. The test statistic is asympotically χ^2 distributed.

We need to introduce the notion of *linearity in the mean*. This is the case for a process when the conditional mean is a function of elements of the information set[9] in a linear way. The test's null hypothesis is then linearity of the process in the mean.

4.2.6 The Kaplan test

This test is performed in the frequency domain to detect deterministic structures that are nt well pronounced in the time domain. The test statistic has a strictly positive lower bound only in the case of a stochastic process.

[8] See Barnett and Serletis (2000).

[9] Most likely these elements are past observations.

The test statistic is also produced for a sufficient number of linear stochastic processes mocking the data. If the test statistic of any artificial linear stochastic process resulting from fitting the data is larger than the value of the statistic obtained from the real data, then the hypothesis of linearity is rejected in favor of noisy nonlinear dynamics.

4.2.7 Application to financial data

As stated, for example, in Barnett and Serletis (2000), there is no generally accepted proof in favor of the existence of nonlinearity, not to mention chaos. Hence, dynamics of financial series commonly are assumed to have a stochastic source. The tests just presented, however, merely test linearity versus nonlinearity or chaos permitted to stem from an unrestricted variety of sources. So, even if the structure of the series is linear, the test should reject the null when only the shock terms are the results of nonlinear or chaotic dynamics.

Findings in favor of the nonlinear, chaotic hypothesis should seem reasonable since it appears improbale that shocks in financial or economic data stem from some merely random stochastic source. The strong hypothesis of linearity should be easy to overthrow.

In the table given in Barnett and Serletis (2000), results of nonlinearity and chaos testing reveal that seven East Euopean black-market exchange rates were shown to not be iid by means of the BDS test. The NEGM test for the same data revealed some hint in the direction of chaos. Another BDS test run on real-time rturns on four stock-market indices also showed evidence against *iid* data. Here, though, the NEGM test results gave no reason for suspecting chaos. The FTSE 100 with 60,000 observations provided they same test results as in the previous case. A BDS tests of S&P 500 and CRSP data rejected *iid* observations. A correlation dimension test of gold

and silver rates of return resulted in a value of v between 6 and 7. Finally, a BDS test daily CRSP value weighted returns brought forth evidence in favor of nonlinearity.

Technically, one can apply the tests in a cascading way. If the White test accepts linearity in mean, then either the Kaplan test or the BDS test can be performed to check for possible linearity. On the other hand, if the White test rejects its null of linearity in mean, then the NEGM test could look for possible signs of chaos. Finally, if full linearity is rejcted but the White test accepts its null hypothesis, the volatility of the process could be stochastic which is to presented in the sequel.

4.2.8 Summary

As mentioned in Barnett and Serletis (2000) for an unknown dynamical system, the correlation integral as well as the Lyapunov exponents react to dynamic noise. It might actually be impossible to distinguish between stochasticity and high-dimensional chaos. Also, testing for linearity in contrast to nonlinearity of a stochastic process becomes extremely difficult with increasing order of the linear filter. In the attempt to detect nonlinearity, it is therefore essential to keep the order of the linear filter as low as possible when the data is assumed to be linear. Also, chaos should be of low order when it the goal to test for nonlinearity caused by chaos.

It is impossible with the existing tests, however, to determine the source of chaos or nonlinearity. Chaos might come either from external chaotic shocks as well as being of endogenous origin. Hence, only the detection of chaos in fincancial data and economic systems is feasible. But the souce will have to remain hidden until further tests permit a more thorough analysis of chaos inherent in structures.

4.3 ARCH process

4.3.1 General introduction

In this section, we consider stochastic processes with time-changing variance, more precisely, conditional volatility. Now, the variance at time $t+1$ of the process $\{x_t\}$ conditional on time t is a function of past information up to and including time t.

A simple example may include a constant and the lagged observations of the previous period, i.e. $\sigma^2_{|\mathcal{F}_t} = \sigma^2_{|x_{t-1}} = c + a_1(x_{t-1} - \mu)^2$. c and a_1 are both positive. The process may be conditionally white noise, i.e. $x_t|\mathcal{F}_t \sim NID(\mu, \sigma^2_{|x_{t-1}})$, with unconditional mean μ and unconditional constant variance. In case of $a_1 < 1$, this variance process is stationary. The unconditional variance of the process $\{x_t\}$ is then $\sigma^2 = c/(1-a_1)$. The kurtosis

$$\mathcal{K} = 3\frac{1-a_1^2}{1-3a_1^2} \tag{4.2}$$

exists for $3a_1^2 < 1$. It is obvious from (4.2), that the unconditonal process $\{x_t\}$ is leptokurtic, i.e. $\mathcal{K} > 3$.

Introduced by Engle (1982), this model was first introduced. It is known to be the *first-order autoregressive conditional heteroscedastic* or ARCH(1) process. Usually, the notation is $\epsilon_t = x_t - \mu = \sigma_{|\mathcal{F}_t} \cdot \eta_t$ where η has mean zero and a constant variance of 1 and h_t instead of $\sigma_{|\mathcal{F}_t}$. Consequently,

$$\epsilon_t|x_{t-1}, \ldots \sim NID(0, h_t)$$
$$h_t = c + a_1\epsilon^2_{t-1}.$$

Just as with AR processes, the extension from ARCH(1) to ARCH(p) seems to be obvious. Thus,

$$h_t = \sigma^2_{|\mathcal{F}_t} = c + \sum_{i=1}^{p} a_i\epsilon^2_{t-1}.$$

Here again, to assure the variance to be positive, all lag coefficients $a_i,\ 1 \leq i \leq p$ have to be nonnegative, and $c > 0$. As before, in the AR case, all roots of the characteristic polynomial solving $c + \sum_{i=1}^{p} Z^i = 0$ have to be outside the unit circle. That is the case if $\sum_{i=1}^{p} a_i < 1$. Hence, the process' unconditional variance turns into $\sigma^2 = c/(1 - \sum_{i=1}^{p} a_i)$. This corresponds to the conditional variance of the form

$$h_t = c + \sum_{i=1} a_i \epsilon_t^2.$$

As, for example, Mills (1997) point out, a problem arises from estimating the parameters for increasing p. Here, the estimates might possibly become negative. As a consequence, the obtained conditional variance may no longer be positive which obviously violates the definition of the variance.

Also, in analogy to combining AR and MA processes to ARMA processes, the ARCH structure of the conditional variance process can be generalized to yield the generalized ARCH or GARCH(p, q) process of the form

$$\begin{aligned} h_t &= c + \sum_{i=1}^{p} a_i \epsilon_{t-1}^2 + \sum_{i=1}^{q} b_i h_{t-1} \\ &= c + a(L) + b(L) h_t. \end{aligned} \tag{4.3}$$

with $q > 0$ and $b_i \geq 0,\ 1 \leq i \leq q$ It was first introduced by Bollerslev (1986). The process is stationary if $a(B)+b(B) < 1$. Then the unconditional variance is $\sigma_\epsilon = c/(1 - a(1) - a(1))$.

The form can be changed by redefining $\nu_t = \epsilon_t^2 - h_t$. Hence, (4.3) is now of the ARMA(m,q) form

$$\epsilon_t^2 = c + \sum_{i=1}^{m} (a_i + b_i) \epsilon_{t-i}^2 + \nu_t - \sum_{i=1}^{q} b_i \nu_{t-1} \tag{4.4}$$

with $m = \max\{p, q\}$. The innovations ν_t are serially uncorrelated but heteroskedastic.

The variance for s periods into the future can be seen from (4.3) to be

$$h_{t+s} = a + \sum_{i=1}^{n}(a_i\epsilon_{t+s-i}^2 + b_i h_{t+s-i}) + \sum_{i=s}^{m}(a_i\epsilon_{t+s-i}^2 + b_i h_{t+s-i})$$

where $n = \min\{m, s-1\}$. This yields

$$E[h_{t+s}|\epsilon_t, \ldots] = c + \sum_{i=1}^{n}(a_i + b_i)E[h_{t+s-i}|\epsilon_t, \ldots] \tag{4.5}$$
$$+ \sum_{i=s}^{m}(a_i\epsilon_{t+s-i}^2 + bh_{t+s-i}).$$

For the unconditional variance, we obtain

$$\sigma_\epsilon^2 = \frac{c}{1 - \sum_{i=1}^{m}(a_i + b_i)}.$$

Then, (4.5) turns into

$$E[h_{t+s}|\epsilon_t, \ldots] = \sigma_\epsilon^2 + \sum_{i=1}^{n}(a_i + b_i)(E[h_{t+s-i}|\epsilon_t, \ldots] - \sigma_\epsilon^2)$$
$$+ \sum_{i=s}^{m}(a_i(\epsilon_{t+s-i}^2 - \sigma_\epsilon^2) + b(h_{t+s-i} - \sigma_\epsilon^2)).$$

For $s \to \infty$, this tends to σ_ϵ^2.

The most illustrative example is given by an GARCH(1,1) process

$$h_t = c + a_1\epsilon_{t-1}^2 + b_1 h_{t-1},\ c > 0, a_1 \geq 0, b \geq 0$$

which yields a stationarity of $\{\epsilon_t\}$ when $a_1 + b_1 < 1$. For $s \geq 2$, $E[h_{t+s}|\epsilon_t, \ldots] = c + (a_1 + b_1)E[h_{t+s-1}|\epsilon_t, \ldots]$.

For $c = 0$ and $a_1 + b_1 = 1$ the model becomes *integrated* GARCH or IGARCH(1,1). That is

$$h_t = a_1\epsilon t - 1^2 + (1 - a_1)h_{t-1}$$

with

$$E[h_{t+s}|\epsilon_t, \ldots] = E[h_{t+s-1}|\epsilon_t, \ldots] = \ldots$$
$$= E[h_{t+1}|\epsilon_t, \ldots] = h_{t+1}.$$

I.e., for all $s > 1$, the s step ahead conditional variance equals the one step ahead conditional variance. In other words, h_t has permanent influence on all future conditional variances. Shocks of the past shape the volatility in the future. Here, in contrast to the random walk model, the conditional variance has a unit root. With positive c, the conditional variance also has a trend.

Alternatives as to the conditional normal distribution of the ϵ is commonly provided by the student's t distribution with f degrees of freedom. It is heavier tailed than the Gaussian distribution. For further alternatives, the reader is referred to the vast amount of literature on this topic.

Up to now, h_t has been a function on quadratic lagged innovations ϵ_{t-1}. This restrictin can be relaxed. One possible form is

$$h_t = c + b_1 h_{t-1} + a_1 |\epsilon|^\lambda$$

where for $\lambda = 2$ the well known GARCH(1,1) model is retrieved, and another alternative is

$$h_t = c + b_1 h_{t-1} + a_1 (2\Phi(\epsilon_{t-1}^2/\gamma) - 1).$$

Φ is the standard normal cumulative distribution function. For very small γ, the conditional variance process approaches a linear regression on the lag one, h_{t-1}.

Another alternative is the well known *exponential* GARCH or EGARCH. The EGARCH(1,1) is of the form

$$\ln(h_t) = c + b_1 \ln(h_{t-1}) + a_1 (\theta \frac{\epsilon_{t-1}}{h_{t-1}} + (|\frac{\epsilon_{t-1}}{h_{t-1}}| - \sqrt{\frac{2}{\pi}})).$$

4.3.2 Estimation of the parameters of ARMA-ARCH models

Up to now, the error terms, $\epsilon_t = x_t - \mu$ with possibly zero μ, have been assumed to be serially uncorrelated. More flexibility is given when the zero mean x_t are considered following an ARMA(p, q) process with ARCH innovations. The model is then

$$\phi(B)(x_t) = \theta(B)\epsilon_t$$

$$h_t = E[\epsilon_t^2|\epsilon_{t-1}, \ldots] = c + \sum_{i=1}^{r} a_i \epsilon_{t-i}^2 + \sum_{i=1}^{s} b_i h_{t-i}$$

where the

$$h_t = z_t^{'}\omega = z_{1t}^{'}\omega_1 + z_{2t}^{'}\omega_2$$

$$z_t^{'} = (z_{1t}^{'}; z_{2t}^{'}) = (1, \epsilon_{t-1}^2, \ldots, \epsilon_{t-r}^2; h_{t-1}, \ldots, h_{t-s})$$

$$\omega^{'} = (\omega_1^{'}; \omega_2^{'}) = (c, a_1, \ldots, a_r; b_1, \ldots, b_s)$$

In this context, we will provide two methods to estimate the parameters. That is a Maximum Likelihood (ML) approach and a Least squares (LS) regression.

The first approach, now, is to estimate these parameters with the ML method. We define Θ to be the parameter vector of the form $\Theta = (\omega^{'}; \phi^{'})$ with the ARMA vector $\phi^{'} = (\theta_1, \ldots, \phi_p, \theta_1, \ldots, \theta_q)$. The true parameters is supposed to be $\Theta_0 = (\omega_0^{'}; \phi_0^{'})$.

The log-likelihood for observation at time t is [10] given to be

$$l_t(\Theta) = -\frac{1}{2}(\ln(h_t) + \frac{\epsilon_t^2}{h_t})$$

such that over all T observations

$$L_T(\Theta) = \frac{1}{T}\sum_{t=1}^{T} l_t(\Theta).$$

[10] In the Gaussian case.

The BHHH[11] algorithm proceeds as following. Let $\Theta^{(i)}$ denote the parameter estimate at iteration i. The algorith, then, computes the next iteration's value through

$$\Theta^{(i+1)} = \Theta^{(i)} + \lambda_i \left(\sum_{t=1}^{T} \frac{\partial l_t}{\partial \Theta} \frac{l_t}{\Theta'} \right) \sum_{t=1}^{T} \frac{\partial l_t}{\partial \Theta}.$$

$\frac{\partial l_t}{\partial \Theta}$ is evaluated at $\theta^{(i)}$. λ_i is a step size determined for each iteration i. The information matrix defined as $I = -E(\partial^2 l_t / \partial \Theta \partial \Theta')$ is block diagonal. This implies that the ω and ϕ can be estimated separately, i.e. each partameter set can be asymptotically efficiently estimated based on the consistent estimate of the other parameter set.

The strongly consistent estimate $\hat{\Theta}_{ML}$ for Θ_0 is asymptotically normally distributed with mean Θ_0 and covariance matrix I^{-1}. The later can be consistently estimated by $T(\sum_{t=1}^{T}(\partial l_t/\partial\Theta)(\partial l_t/\partial\Theta'))^{-1}$. The final BHHH iteration may serve to provide the values for the latter estimate.

The LS estimation is performed in the accustomed manner. That is, since the variance of the ϵ, $\sigma^2 = E(\epsilon_t^2)$, is constant and finite, the LS estimates of the ARMA parameters are found from $\hat{\sigma}^2(\phi) = T^{-1} \sum \epsilon_t^2 \rightarrow \min$. For finite $E(\epsilon_t^4)$, the estimates will be consistent and asymptotically normal. The covariance matrix of the $\hat{\phi}$ is given by

$$C = E\left[\frac{\partial \epsilon_t}{\partial \phi} \frac{\partial \epsilon_t}{\partial \phi'}\right]^{-1} E\left[\epsilon_t^2 \frac{\partial \epsilon_t}{\partial \phi} \frac{\partial \epsilon_t}{\partial \phi'}\right] E\left[\frac{\partial \epsilon_t}{\partial \phi} \frac{\partial \epsilon_t}{\partial \phi'}\right]^{-1}$$

evaluated at ϕ_0.

For computation of the LS ARCH parameters in equation (4.4), the squared innovations estimates, e_t, obtained from the ARMA model are used.

[11] For details, see, for example, Mills (1997).

4.3.3 Testing for the presence of ARCH errors

When the ARMA model is being estimated, the obtained residuals, e_t, have to be tested for ARCH effects since negligence of which can lead to significant model misspecification. The parameter standard errors are usually too high, then, when ARCH is not considered. This will most likely result in overparametrizatioon ot the ARMA structure.

The squared residuals inserted into (4.4) can be treated as ARMA which is shown, for example, in Bollerslev (1986). This will yield the order of the ARCH or GARCH model underlying the innovations. As shown in literature, the sample autocorrelations of the e_t^2 have asymptotic variance T^{-1}. If the ϵ_t^2 are independent, the according portemanteau statistics of the autocorrelations are asymptotically χ^2 distributed.

Engle (1982) provides a test for the null hypothesis of constant conditional variance against the hypothesis of a conditional variance given by an ARCH(r) model. The test is based on the Lagrange multiplier princile, hence, LM test. The e_t^2 are regressed on r lagged terms and a constant with the resulting test statistic $R^2 \cdot T$. R^2 is the squared correlation coefficient of the regression. The statistic is, thus, χ squared distributed with r degrees of freedom.

With GARCHr, s) as alternative to the null, a test in the above manner is infeasible. Moreover, for a null of GARCH(r, s), a test for a higher order structure is also not possible. With the same null, the LM statistics for GARCH(r, d) and ARCH($r + d$), unfortunately, are the same. This means that the test is indifferent between testing for additional d lagged innovations terms in an ARCH(r) model and adding $d - s$ lagged variance terms to yield a GARCH(r, d) model. A feasible test, however, is a null of ARCH(r) against GARCH(r, s), i.e. $H_0 : \omega_2 = 0$. The test statistic is $T \cdot R^2$ from the regression of $(\epsilon_t^2 h_t^{-1} - 1)$ on $(h_t z_t)$.

4.3.4 The use of GARCH models in applied econometrics

In econometrics, for the longest, dependence structures of variables have been estimated with models minimizing the squared error terms. The idea of GARCH models, however, is to provide some method to predict the variance of these error terms which is assumed to change during time. This is in contrast to the *homoscedastic* volatility of traditional regression models. Dynamic volatility is referred to as *heteroscedasticity*.

When heteroscedasticity is observed, the estimated regression parameters are still unbiased, see, for example, Engle (2001). But their respective standard errors and confidence intervals are too small. GARCH models, now, remedy that short coming and, moreover, predict the future variance based on the given information. Additionally, so called *robust* standard errors can be applied instead of the exact standard errors if the sample size is large. If the sample size should be small, a correction of the standard errors of the estimated coefficients with respect to heteroscedasticity might be called for. Precision of estimated regression parameters is of great importance for in fincancial issues.

When looking at financial time series, one can frequently observe periods of residuals large in absolute value and, on the other hand, periods with small abolute model residuals. This is called *volatility clustering*. Referring to these residuals as risk, one can, then, say that there is autocorrelation in the risk. This is the field where GARCH models prevail.

In econometrics, the prediction of mean and variance of financial returns is of prime interest. The main task is to use available information and incorporate it into predictions. In contrast to the mean, there have not been many sophisticated ways to estimate the variance. Instead of equally weighted observations of the squared residuals of a fixed length which has been common

practice, the weights are estimated in the GARCH models and allow for greater flexibility. Generally, GARCH weights of past squared residuals are declining but, because of the structure, never vanish. At each period , a new squared residual is added as the new information. In applications, this method of modeling conditional variances has performed successful, e.g. in Engle (2001).

The form of the GARCH model is such that the conditional variances are mean reverting. As shown before, the unconditional variance is constant and in the long run, the conditional variance approaches its unconditional counterpart.

To give an example, we look at the calculation of the *value at risk* of a certain given portfolio which is the value of the loss that is exceeded with a given probability. In other words, its is the 1 percent quantile of the return distribution.

The following example is given in Engle (2001). We assume to have a portfolio worth $ 1,000,000 which we want to estimate the 1 percent value at risk of. The portfolio consisting of indices and bonds is composed of 50 percent Nasdaq, 30 percent Dow Jones and 20 percent long bonds. The time of consideration is given to be the 23rd of March in 2000 when volatility was very high. Plots of the respective returns can be viewed in Figure 4.1.

ARCH effects are striking for all equities when comparing the beginning part to the end.

In Table 4.1, mean, standard deviation (std. dev.), skewness, and kurtosis are presented for the constituents as well as the portfolio, in column three. By looking at the respective std. dev.'s, one can notice the positive effect diversification has despite the high content of equity.

Table 4.2 shows the autocorrelations of the squared returns along with the Q-statistics and their corresponding p-values. For all 15 lags, autocorrelation is significant. More, they are all positive. All this is indicative of GARCH

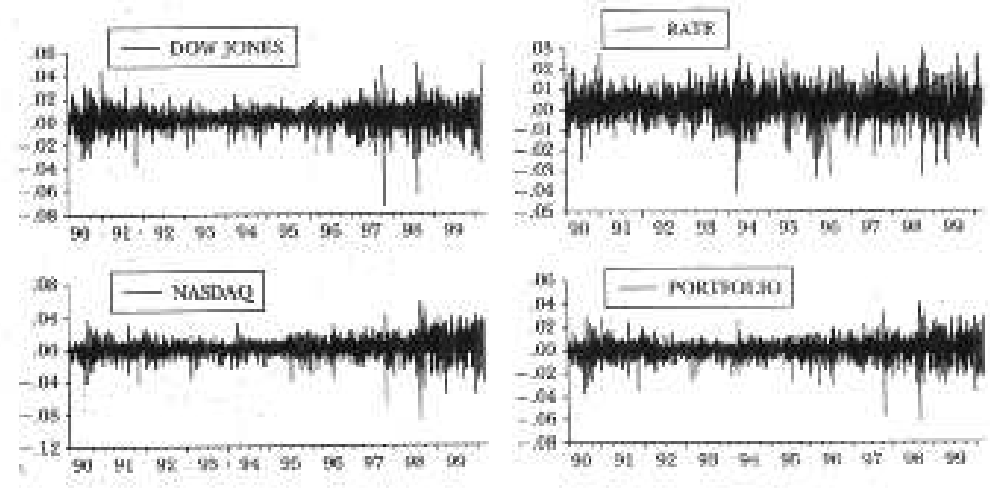


Figure 4.1:

effects.

Next, the std.dev.'s are predicted along with the portfolio's 1 percent quantile. This is done individually for three different intervals: March 1990 through March 23, 2000, March 1999 through March 23, 2000, and January 1, 2000 through March 23, 2000.

The historical 1 percent value at risk for the ten year period is at $ 22,477. The second interval yields a value at risk of $ 24,653, whereas the last interval's value at risk is at $ 35,159 which is definitely an increase over its two predecessors. For a short period into the future, these value at risk numbers are true if the distribution of the return does not change.

Table 4.3 lists the estimates of a GARCH(1,1) model fitted to the data. Standard errors of the estimated parameters are the result of a robust method.[12] The sum of the lag coefficients are very close to one indicating that the variance process reverts to its mean very slowly.

In 4.4, the values of the autocorrelations of the standardized residuals

[12]For further details, the reader is referred to Engle (2001).

Portfolio Data

NASDAQ				
		Dons Fones	Rate	Portfolio
M	0.0009	0.0005	0.0001	0.0007
STD.D	0.0115	0.0090	0.0075	0.0083
SKE	-0.5310	-0.5593	-0.2031	-0.4738
Kurto	7.4936	8.3288	4.9579	7.0026

Sampler: March 23,1990 to March 23,2000

Table 4.1:

are given. For all lags, these values seem to have been noticeably diminished compared to their counterparts resulting from the returns. When examining the respective values of the Q statistics, it is striking that any significance of autoregressive terms has vanished. Hence, there appear to be no ARCH effects in the residuals.

With the obtained parameter values, one forecasts a std. dev. for the next day of 0.0146. This is ca. twice the value of the value of the average std. dev. tabulated proviously. With this value, the normal distribution would yield a 1 percent quantile of the residuals of std.dev.$\times\Phi(0.01)$ resulting in a 1 percent value at risk of \$33,977.[13]

As stated, for example, in the original source, the normal distribution, however, is not a good fit for the empirical distribution of the standardized residuals. The empirical 1 percent quantile, in this case, is 2.844 which results in a value at risk of \$39,996. The empirical residuals appear to have much heavier tails in their distribution than the normal distribution permits. This is definitely reflecting the increased volatility in the returns since 2000.

It is left to notice that one need not go the way to compute the standard deviations. This might even be impossible to do if the tails of the underlying

[13]Here, Φ denotes the standard normal distribution function.

	AC	Q-Star	P
1	0.210	115.07	0.000
2	0.183	202.64	0.000
3	0.116	237.59	0.000
4	0.082	255.13	0.000
5	0.122	294.11	0.000
6	0.163	363.85	0.000
7	0.000	384.95	0.000
8	0.090	401.77	0.000
9	0.081	427.88	0.000
10	0.081	445.03	0.000
11	0.069	457.88	0.000
12	0.000	474.29	0.000
13	0.076	483.42	0.000
14	0.074	563.99	0.000
15	0.083	521.95	0.000

Table 4.2:

Variance equation

Variable	Coef	St.Err	Z-Stat	P-Value
C	1.40E-06	4.49E-07	5.1210	0.0018
ARCH(1)	0.0772	0.0179	4.3046	0.0000
GARCH(1)	0.9046	0.0196	46.1474	0.0000

Notes:Depart Variable,PORT

Sample :March 23,1990 to March 23,2000

Table 4.3:

distribution are too heavy to permit finite variance. Using the quantiles of the forecasts, instead, provides for a practical alternative.

The portfolio under consideration lost as much $67,000 on a single day, in the ensuing month through April 14, 2000. This is particularly a result of the decline in the Nasdaq in that period.

As can be seen in the plot in 4.2, the value at risk was trespassed only one percent of the time during the sample period which is to be expected since the parameters are based on this sample.

A plot of the forecasts for the period until the end of the second quarter of 2001 can be seen in Figure 4.3. Here, parameters are not reestimated, and the values at risk are merely updated using the latest inormation.

Volatilty models exist of various forms. But sill, the GARCH(1,1) is the simplest. Moreover, it is very robust.[14] A natural extension, though, appears to be a GARCH(p, q) model. Despite its higher parametrization, it might sometimes be a benefit to apply when the time series is very long. Generally, the GARCH(1,1) is very satisfactory, though.

One issue that is not touched by GARCH models is the direction of the returns. It has been found, however, that negative news affect volatility more intensely than positive news. One may want to look at EGARCH and

[14]See Engle (2001).

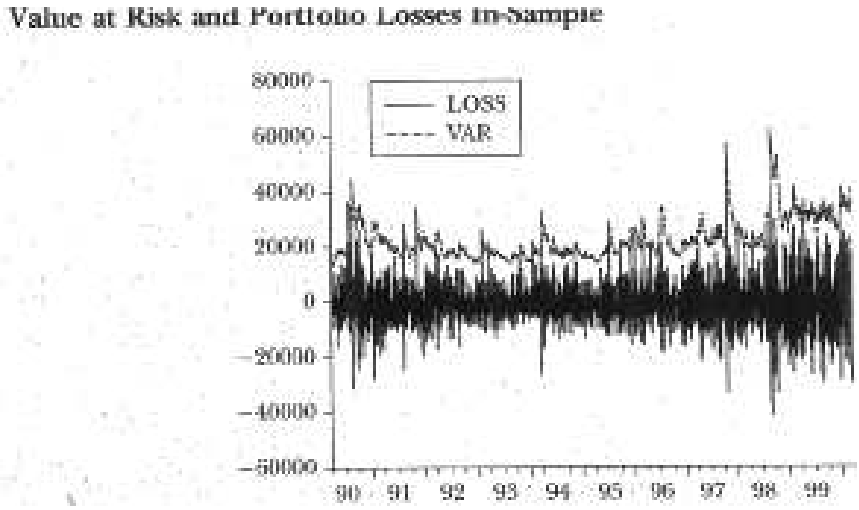


Figure 4.2:

TARCH, in this context.

Up to now, the origin of volatility has not been explained by the models. Investigatin into that direction have not, yet, led to illuminating results.[15] In general, it seems mots plausible that volatility is some function of news. Also, cross influences should be taken into consideration when examining the volatilities of several assets in one portfolio.

4.3.5 Some alternative parametric ARCH models: an empirical analysis

Based on tests performed in Loudon, Watt and Yadav (2000), we will present some empirical results of alternative extensions of the GARCH model with respect to capturing the volatility inherent in daily UK value-weighted stock index in the period between 1971-97. The period under investigation will be split up into three subperiods to provide for comparison between period

[15]See, for example, Engle (2001).

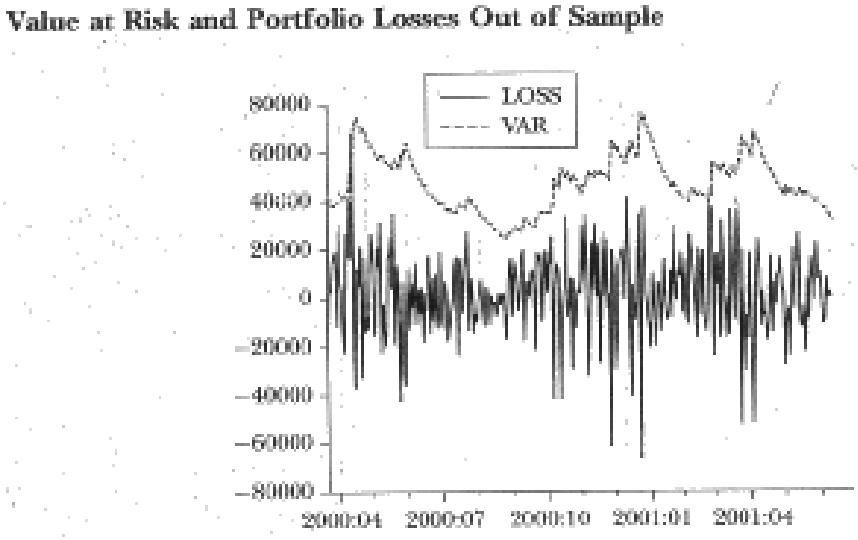


Figure 4.3:

related performance of the individual alternatives. Moreover, in addition to the in-sample performance, the predictive power with respect to out-of-sample performance will be tested for the individual models and compared wherever feasible.

As will be seen, outperformance of particular models is generally not consistent across different subperiods. The same is true for the ex-ante prediction power of the respective volatility. Additionally, we will see that ARCH is highly significant. Skewness and kurtosis of the series will be captured to some extend by the models. Despite the fact that not all volatility is captured by the models, except for a few models, the conditional variance is unbiased with respect to prediction of the individual model variances.

In general, the models perform quite similarly. An exception is the multiplicative GARCH model.

The testing includes eight model alternatives. All have in common that the stock index returns are regressed on five business weekday dummies and a holiday dummy. The remainder is itself regressed against a constant and

ten autoregressive lags. The final residuals, $\{\epsilon\}_t$, are assumed to conditinally Gaussian with zero mean and time dependent variance h_t. The general parameterization of the models, then, looks as follows

$$R_t = a_1 MON_t + a_2 TUE_t + a_3 WED_t + a_4 THU_t + a_5 FRI_t + a_6 HOL_t + \nu_t$$
$$v_t = b_0 + \sum_{i=1}^{10} b_i \nu_{t-i} + \epsilon_t.$$

The conditional variance of the $\{\epsilon\}_t$ is given by

$$h_t = \begin{cases} |\lambda\phi_t - \lambda + 1|^{1/\lambda}, & \lambda \neq 0 \\ \exp(\phi_t - 1), & \lambda = 0. \end{cases} \quad \phi_t = \alpha_0 + \phi_{t-1}\xi_{1,t-1} + \xi_{2,t-1}$$
$$\xi_{1,t-1} = \alpha_1 + \alpha_2|\epsilon - c|^\delta + \alpha_3 \max(0, c - \epsilon_t)^\delta$$
$$\epsilon_{2,t-1} = \alpha_4 \frac{|\epsilon_t - c|^\delta - 1}{\delta} + \alpha_5 \frac{(\max(0, c - \epsilon_t))^\delta - 1}{\delta}$$

The models are distinguisehd by the specifications below.

(1) LGARCH	$\lambda = 1, c = 0, \delta = 2, \alpha_3 = 0, \alpha_4 = 0, \alpha_5 = 0, \alpha_0 > 0, \alpha_1 \geq 0$, and $\alpha 2 \geq 0$
(2) MGARCH	$\lambda = 0, c = 0, \delta = 0, \alpha_2 = 0, \alpha_3 = 0, \alpha_4 = 2\alpha_1$, nd $\alpha_5 = 0$
(3) EGARCH	$\lambda = 0, c = 0, \delta = 1, \alpha_2 = 0$, and $\alpha_3 = 0$
(4) GJR-GARCH	$\lambda = 1, c = 0, \delta = 2, \alpha_5 = 0, \alpha_0 > 0, \alpha_1 \geq 0, \alpha_2 \geq 0$, and $\alpha 2 + \alpha 3 \geq 0$
(5) NGARCH	$\lambda = 1, \delta = 2, \alpha_3 = 0, \alpha_4 = 0, \alpha_5 = 0, \alpha_1 \geq 0$, and $\alpha_2 \geq 0$
(6) VGARCH	$\lambda = 1, \delta = 2, \alpha_2 = 0, \alpha_3 = 0, \alpha_5 = 0, \alpha_0 > 0, \alpha_1 \geq 0$, and $\alpha_4 \geq 0$
(7) TS-GARCH	$\lambda = .5, c = 0, \delta = 1, \alpha_3 = 0, \alpha_4 = \alpha_2, \alpha_5 = 0, \alpha_1 \geq 0, \alpha_2 \geq 0$, and $\alpha_0 > \alpha_1 + \alpha_2 + \alpha_3 - 1$
(8) TGARCH	$\lambda = .5, c = 0, \delta = 1, \alpha_4 = \alpha_2, \alpha_5 = \alpha_3, \alpha_1 \geq 0, \alpha_2 \geq 0, \alpha_2 + \alpha_3 \geq 0$, and $\alpha_0 > \alpha_1 + \alpha_2 - 1$.

Why are different parametrizations necessary. As mentioned by Loudon, Watt and Yadav (2000), the main reason might be the drawback of original GARCH(1,1) to use squared residuals. This fails to make use of the direction of past shocks. In reality, however, asymmetry in the magnitude as a

result of particular shocks is observed. That is, volatility reacts more pronouncedly to negative shocks than to positive shocks. Hence, models such as EGARCH, GJR-GARCH, NGARCH, VGARCH, and TS-GARCH are introduced to account for this asymetric relationship. The so called *News Impact Curve* provides a good presentation of the conditional volatilities' dependence on the directions of the immediately preceding innovations. Except for the LGARCH model, as is expected, the curve is asymmetricly centered about a particular value of ϵ_{t-1}.

Results of testing the 27 years worth of FT All Share Index data, performed in Loudon, Watt and Yadav (2000) are to presented in brief.[16] Parameters were obtained from conducting a MLE according to the BHHH algorithm mentioned in (4.3.2). The test incorporates in-sample as well as out-of-sample (*ex ante*) volatility estimation.

The following Table 4.5 reveals the respective estimated parameters along with the p-values. It can be seen that almost all parameters are highly significant. Generally, the previous value of the conditional variance has extreme influence on today's value. Hence, volatility forecasts will put a lot of weight on the level closest in history. Moreover, when models incorporated parameters for detection of asymmetry, these were highly significant, as well. In some cases, the levels of significance grew from period one through the final period. Also, despite the formal inapplicability of likelihood ratios, they were computed nevertheless and are listed in Table 4.5, as well.

The eight diagnostic checks conducted are listed in Table 4.5.[17] Some are based on the conditional variance forecast errors, $\epsilon_t^2 - h_t$, assuming that the true volatility in day t is ϵ_t^2. Standardized errors, $\epsilon_t/\sqrt{h_t}$ are used for computation of the skewness, kurtosis, and the LM statistics. Others refer

[16]For further details, the reader is referred to original source.

[17]For the complete listing of the tests, the reader is referred to Loudon, Watt and Yadav (2000).

Autocorrelations of Squared Standardized Residuals

k	AC	Q-Star	Prob
1	0.005	0.0589	0.808
2	0.039	4.0240	0.134
3	-0.011	4.3367	0.227
4	-0.017	5.0981	0.227
5	0.002	5.1046	0.403
6	0.009	5.3228	0.503
7	-0.015	5.8836	0.553
8	-0.013	6.3272	0.611
9	-0.024	7.8169	0.553
10	-0.006	7.9043	0.638
11	-0.023	9.3163	0.593
12	-0.013	9.7897	0.634
13	-0.003	9.8110	0.709
14	0.009	10.038	0.759
15	-0.012	10.444	0.791

Table 4.4:

to the squared inovations, ϵ_t^2. Now, Table 4.5 also reports the results from diagnostic checks for the ex ante forecasts of periods two and three. The parameters for the ex ante forecasts were obtained from the respective period preceding the one under investiation.

It should be mentioned that detection of skewness is model-dependent as well as period specific. In general, it is increasing. Leptokurtosis is always highly significant and prevailing for all periods.

After removing weekday effects, serial correlation is rejected for the raw returns, ϵ, but they definitely exhibit ARCH effects up to a lag order of ten. The news impact curve indicates pronounced dependence of the behavior of

the squared returns on the direction and magnitude of the previous returns. This gives reason to assume that models considering asymmetry might outperform the others. These findings agree with previous findings causing the invention of the respective models.

The root mean square errors (RMSE) are quite similar for all models throughout the three periods. Only for the MGARCH, higher RMSE errors are reported. Due to the RMSE, no clear preference can be determined since model orders change between the periods.

When analyzing the equality across all models of the respective conditional variance forecast errors, $\epsilon_t^2 - h_t$, reults are not quite obvious. When out-of-sample values are considered, equality is strongly rejected for both periods. In the case of in-sample predictions errors, the difference in the errors depends on the period For example, equality cannot be rejected in the second period whereas they are different in the remaining two.

Pairwise comparison of equality with respect to the LAGRCH model, reveals that the corresponding test statistics are not significant in almost all cases.

Skewness in the $\epsilon_t/\sqrt{h_t}$ is found particularly in the first two periods. In general, there is well pronounced skewness across all models throughout all three periods. As far as excess kurtosis is concerned, it is found for the standardized residuals in all models that they are significantly leptokurtotic. The level of kurtosis, however, is somewhat lower than for the raw returns. No ranking of the models is possible, in this respect.

The standardized residuals from the different models are apparently not significantly correlated according to LM tests based on ten lags. However, when the linear filtering parameters are not constantly adjusted to new information, the serial corrrelation becomes more noticeable. When the squared standardized residuals, ϵ_t^2/h_t^2, are tested, the LM tests reveal serial correlation in some models. This is evidence for insufficiency in the respective

models explanation of volatility. One should observe the differences between the ex-ante and ex-post results.

Impact of sign and magnitude of the previous value is substantially lower for the standardized residuals than for the raw filtered returns. For the first ones, there is some significance detected dependent on the sign when in-sample values are analyzed. Additional tests regress the residuals on the conditional variance.

All in all, it can be remarked that the appropriate model should be selected period-specifically.

Panel A:1971-80	LGARCH	MGARCH	EGARCH	GJRCARCH	NGARCH	VGARCH	TSGARCH	TGARCH
α_0	0.011	1.27	0.065	0.01	0.01	0.001	0.035	0.044
	$(0.002)^c$	$(0.017)^c$	$(0.008)^c$	$(0.002)^c$	$(0.002)^c$	$(0.003)^c$	$(0.003)^c$	$(0.005)^c$
α_1	0.922	0.162	0.991	0.926	0.926	0.992	0.916	0.923
	$(0.005)^c$	$(0.006)^c$	$(0.002)^c$	$(0.005)^c$	$(0.005)^c$	$(0.001)^c$	$(0.005)^c$	$(0.005)^c$
α_2	0.07			0.056	0.066	0.136	0.09	0.072
	$(0.007)^c$			$(0.01)^c$	$(0.007)^c$	$(0.009)^c$	$(0.007)^c$	$(0.008)^c$
α_3				0.021				0.025
				$(0.011)^a$				$(0.008)^c$
α_4			0.135					
			$(0.015)^c$					
α_5			0.042					
			$(0.015)^c$					
c					0.161	0.119		
					$(0.073)^b$	$(0.031)^c$		
Log-L	-3664.52	-3968.51	-3671.43	-3663.05^a	-3662.51^b	-3672.11	-3673.76	-3670.58^b
Panel B:1981-1990	LGARCH	MGARCH	EGARCH	GJRCARCH	NGARCH	VGARCH	TSGARCH	TGARCH
α_0	0.048	0.805	0.158	0.05	0.061	0.053	0.082	0.129
	$(0.007)^c$	$(0.013)^c$	$(0.009)^c$	$(0.007)^c$	$(0.008)^c$	$(0.008)^c$	$(0.007)^c$	$(0.009)^c$
α_1	0.826	0.128	0.928	0.828	0.782	0.917	0.814	0.817
	$(0.016)^c$	$(0.004)^c$	$(0.01)^c$	$(0.017)^c$	$(0.018)^c$	$(0.009)^c$	$(0.014)^c$	$(0.016)^c$
α_2	0.018			0.058	0.11	0.114	0.141	0.079
	$(0.011)^c$			$(0.014)^c$	$(0.013)^c$	$(0.011)^c$	$(0.007)^c$	$(0.013)^c$

Table continued over page

	LGARCH	MGARCH	EGARCH	GJRCARCH	NGARCH	VGARCH	TSGARCH	TGARCH
α_3				0.085				0.092
				$(0.012)^c$				$(0.009)^c$
α_4			0.153					
			$(0.023)^c$					
α_5			0.136					
			$(0.017)^c$					
c					0.46	0.338		
					$(0.07)^c$	$(0.059)^c$		
Log-L	-3125.86	-3316.19	-3109.99	-3115.43^c	-3108.62^c	-3123.19	-3132.64	-3113.92^c
Panel C:1991- October 1997								
α_0	0.014	0.258	0.074	0.01	0.01	0.003	0.022	0.044
	$(0.004)^c$	$(0.02)^c$	$(0.011)^c$	$(0.003)^c$	$(0.003)^c$	$(0.004)^c$	$(0.004)^c$	$(0.006)^c$
α_1	0.899	0.049	0.982	0.923	0.906	0.977	0.929	0.939
	$(0.017)^c$	$(0.007)^c$	$(0.005)^c$	$(0.015)^c$	$(0.015)^c$	$(0.006)^c$	$(0.012)^c$	$(0.011)^c$
α_2	0.071			0.027	0.052	0.029	0.067	0.031
	$(0.01)^c$			$(0.009)^c$	$(0.009)^c$	$(0.005)^c$	$(0.009)^c$	$(0.01)^c$
α_3				0.061				0.051
				$(0.014)^c$				$(0.01)^c$
α_4			0.061					
			$(0.02)^c$					
α_5			0.095					
			$(0.019)^c$					
c					0.659	0.703		
					$(0.136)^c$	$(0.165)^c$		
Log-L	-1742.34	-1812.08	-1729.95	-1735.17^c	-1730.84^c	-1741.97	-1738.34	-1729.41^c

Panel A:1971-80	Linearly filtered raw returns	LGARCH	MGARCH	EGARCH	GJRCARCH	NGARCH	VGARCH	TSGARCH	TGARCH
RMSE									
In sample		3.245	3.548	3.271	3.26	3.261	3.331	3.257	3.278
WC χ^2									
in sample		$(13.51)^c$	2.53	0.85	0.72	0.61	0.70	1.69	1.15
DM									
in sample			0.47	0.89	0.73	0.87	1.02	1.23	1.51
Skewness									
in sample	0.00	-0.27^c	-0.13^c	-0.28^c	-0.26^c	-0.26^c	-0.15^c	-0.30^c	-0.29^c
Kurtosis									
in sample	5.09^c	1.41^c	3.23^c	1.47^c	1.39^c	1.41^c	0.87^c	1.55^c	1.57^c
LM (levels)									
in sample	0.32	14.85	3.55	14.03	14.32	13.98	10.65	14.54	13.63
LM (squares)									
in sample	560.3^c	7.67	306.9^c	9.54	7.30	7.44	14.44	8.81	7.88
EN Sign bias									
in sample	3.81^c	-0.84	2.19^b	-0.73	-0.69	-0.76	-0.68	-0.75	-0.59
EN neg.size bias									
in sample	-8.39^c	-2.74	-0.99	-2.37^b	-2.43^b	-2.42^b	-2.17^b	-2.55^b	-2.08^b
EN pos.size bias									
in sample	16.13^c	0.14	4.57^c	0.44	0.48	0.56	1.07	-0.10	0.44

Panel A:1971-80 (continued)	Linearly filtered raw returns	LGARCH	MGARCH	EGARCH	GJRCARCH	NGARCH	VGARCH	TSGARCH	TGARCH
RMSE									
EN joint bias									
in sample	293.6^{c}	8.23^{b}	21.86^{c}	6.01	6.38^{a}	6.30^{a}	5.87	7.58^{a}	4.66
PS χ^2									
in sample		0.06	24.50^{c}	0.09	0.02	0.01	6.11^{b}	0.17	0.17
PS R^2									
in sample		0.199	0.019	0.186	0.192	0.191	0.165	1.193	1.183
PS LB									
in sample		68.00^{c}	103.6^{c}	86.31^{c}	76.81^{c}	78.97^{c}	162.2^{c}	79.02^{c}	92.41^{c}
PS(log) χ^2									
in sample		945.0^{c}	1057.6^{c}	949.2^{c}	944.1^{c}	944.3^{c}	945.3^{c}	944.6^{c}	941.8^{c}
PS(log) R^2									
in sample		0.092	0.026	0.09	0.092	0.092	0.088	0.09	0.09
PS(log) LB									
in sample		12.68	105.1^{c}	11.54	13.26	13.09	11.93	10.95	11.00

Panel B:1981-90									
RMSE									
in sample		3.175	3.47	3.112	3.124	3.151	3.276	3.293	3.246
out of sample		3.216	3.433	3.258	3.215	3.229	3.264	3.308	3.302
WC χ^2									
in sample		5.22	1.29	1.32	0.95	1.01	0.99	0.98	0.80
out of sample		63.05^{c}	1.42	0.43	0.01	1.64	1.09	0.61	0.60
DM									
in sample			0.03	1.16	-0.72	-0.26	1.13	0.79	1.06
out of sample			-3.25^{c}	1.43	-1.23	0.74	-2.03^{b}	1.19	1.43
Skewness									
in sample	-1.2^{c}	-0.94^{c}	-0.88^{c}	-0.64^{c}	-0.80^{c}	-0.69^{c}	-0.72^{c}	-0.80^{c}	-0.62^{c}
out of sample		-0.97^{c}	-1.22^{c}	-0.82^{c}	-0.93^{c}	-0.91^{c}	-0.79^{c}	-0.91^{c}	-0.84^{c}
Kurtosis									
in sample	17^{c}	8.65^{c}	8.14^{c}	4.82^{c}	6.93^{c}	5.35^{c}	6.05^{c}	46.32^{c}	4.36^{c}
out of sample		9.33^{c}	12.68^{c}	7.02^{c}	8.80^{c}	8.45^{c}	8.40^{c}	7.87^{c}	7.05^{c}
LM (levels)									
in sample	0.10	14.89	3.78	14.39	14.67	14.05	12.40	15.77	14.49
out of sample		17.49^{a}	31.27^{c}	17.78^{a}	17.82^{a}	17.65^{a}	13.65	16.33^{a}	16.91^{a}
LM (squares)									
in sample	1027.8^{c}	8.72	490.8^{c}	7.29	5.69	8.40	26.04^{c}	58.59^{c}	36.72^{c}
out of sample		20.16^{b}	539.5^{c}	51.72^{c}	18.17^{a}	23.12^{b}	9.95	140^{c}	123.5^{c}
EN Sign bias									
in sample	-6.32^{c}	-0.06	-3.71^{c}	0.23	0.39	-0.10	-0.38	-1.28	-0.71
out of sample		-0.26	-0.85	-0.82	-0.09	-0.28	-0.01	-2.13^{b}	-1.84^{a}

EN neg.size bias									
in sample	-23.19^{c}	-1.49	-11.42^{c}	-0.52	-0.72	-0.84	-2.24^{b}	-3.57^{c}	-2.11^{b}
out of sample		-2.74^{c}	-4.56^{c}	-3.90^{c}	-2.50^{b}	-2.82^{c}	-1.49	-6.60^{c}	-5.98^{c}
EN pos.size bias									
in sample	5.82^{c}	-0.63	0.97	0.01	0.01	-0.04	0.16	-0.98	0.04
out of sample		0.01	0.09	0.30	0.21	0.23	-0.10	0.00	0.28
EN joint bias									
in sample	473.4^{c}	5.47	129.4^{c}	1.96	0.90	1.04	6.30^{a}	16.73^{c}	4.81
out of sample		10.66^{b}	24.19^{c}	17.75^{c}	9.12^{b}	10.60^{b}	3.76	46.33^{c}	38.19^{c}
PS χ^2									
in sample		0.41	4.37	2.61	0.14	0.24	4.75^{a}	1.42	1.24
out of sample		1.06	85.94^{c}	0.46	1.26	0.83	13.01^{c}	0.23	0.09
PS R^2									
in sample		0.182	0.012	0.281	0.204	0.192	0.247	0.129	0.156
out of sample		0.135	0.03	0.111	0.137	0.129	0.112	0.084	0.087
PS LB									
in sample		668.7^{c}	933.2^{c}	470.2^{c}	631.0^{c}	627.6^{c}	606.8^{c}	704.2^{c}	650.1^{c}
out of sample		793.6^{c}	829.6^{c}	820.1^{c}	793.7^{c}	805.5^{c}	878.9^{c}	857.6^{c}	857.1^{c}
PS(log) χ^2									
in sample		972.2^{c}	1386.6^{c}	957.6^{c}	963.9^{c}	959.9^{c}	987.2^{c}	978.3^{c}	962.8^{c}
out of sample		1072.4^{c}	1947.2^{c}	1052.8^{c}	1070.0^{c}	1056.8^{c}	1156.3^{c}	1080.9^{c}	1058.8^{c}
PS(log) R^2									
in sample		0.044	0.01	0.043	0.046	0.046	0.044	0.04	0.043
out of sample		0.041	0.006	0.039	0.040	0.04	0.031	0.04	0.039
PS(log) LB									
in sample		9.02	47.57^{c}	9.31	9.01	9.02	8.79	11.09	9.63
out of sample		12.86	70.72^{c}	10.84	14.47	14.60	22.54^{b}	8.24	9.57

Panel C:1991- October 1997									
RMSE									
in sample		1.125	1.141	1.117	1.118	1.117	1.124	1.12	1.117
out of sample		1.118	1.166	1.107	1.109	1.11	1.118	1.112	1.108
WC χ^2									
in sample		33.49^{c}	1.30	1.85	2.21	4.16^{b}	0.04	1.05	1.99
out of sample		157.9^{c}	13.71^{c}	2.34	2.79^{a}	2.59	0.01	0.95	1.74
DM									
in sample			0.16	0.74	0.71	0.60	1.10	0.76	0.69
out of sample			-9.31^{c}	0.69	0.73	-1.68^{a}	-2.36^{b}	0.57	-0.21
Skewness									
in sample	0.34^{c}	0.04	0.26^{c}	0.06	0.06	0.05	0.09	0.06	0.06
out of sample		0.03	0.22^{c}	0.04	0.07	0.07	0.06	0.00	0.06
Kurtosis									
in sample	4.42^{c}	2.35^{c}	3.71^{c}	2.31^{c}	2.40^{c}	2.39^{c}	2.62^{c}	2.41^{c}	2.31^{c}
out of sample		1.79^{c}	4.33^{c}	1.92^{c}	2.01^{c}	2.02^{c}	2.03^{c}	1.77^{c}	2.01^{c}
LM (levels)									
in sample	0.32	1.83	0.64	2.59	2.29	2.59	1.88	2.02	2.66
out of sample		15.97	20.15^{b}	16.53^{a}	16.73^{a}	17.18^{a}	16.71^{a}	15.30	16.70^{a}
LM (squares)									
in sample	60.3^{c}	4.12	51.6^{c}	5.16	3.74	3.88	5.74	5.48	5.46
out of sample		6.21	36.1^{c}	7.22	6.61	7.78	5.11	7.84	9.09
EN Sign bias									
in sample	1.33	0.75	1.46	0.69	0.93	0.70	0.81	0.57	0.66
out of sample		1.20	0.79	1.28	1.52	1.26	1.35	1.01	1.24

EN neg.size bias									
in sample	-2.33^b	0.43	-0.75	0.80	0.92	0.97	0.31	0.16	0.80
out of sample		0.91	1.44	1.92^a	1.46	1.70^a	1.42	1.36	2.03^b
EN pos.size bias									
in sample	4.41^c	0.45	2.81^c	0.75	0.74	0.79	1.44	0.50	0.74
out of sample		0.66	-0.39	0.72	1.27	1.28	1.22	-0.13	0.84
EN joint bias									
in sample	24.8^c	0.57	8.79^b	1.25	1.39	1.73	2.23	0.39	1.27
out of sample		1.57	2.36	4.27	3.76	4.87	3.54	2.17	5.06
PS χ^2									
in sample		1.01	6.27^b	1.39	0.19	0.18	1.87	0.23	1.13
out of sample		14.45^c	104.3^c	23.29^c	20.35^c	24.00^c	29.44^c	18.57^c	27.19^c
PS R^2									
in sample		0.032	0.002	0.047	0.043	0.045	0.035	0.039	0.046
out of sample		0.036	0.005	0.053	0.05	0.049	0.037	0.043	0.052
PS LB									
in sample		17.0^a	1.15	16.6^a	15.4	16.3^a	18.9^b	18.3^a	17.0^a
out of sample		14.9	52.3^c	14.2	15.9	16.8^a	14.8	15.4	13.6
PS(log) χ^2									
in sample		679.6^c	1178.3^c	673.7^c	675.2^c	674.6^c	681.8^c	678.4^c	673.4^c
out of sample		870.3^c	1222.0^c	857.5^c	872.9^c	901.5^c	906.2^c	859.2^c	882.4^c
Panel C:1991- October 1997									
PS(log) R^2									
in sample		0.026	0.003	0.038	0.033	0.036	0.032	0.031	0.038
out of sample		0.027	0.005	0.035	0.031	0.032	0.028	0.03	0.034
PS(log) LB									
in sample		3.92	30.06^c	2.47	2.69	2.26	3.43	2.43	2.56
out of sample		8.55	33.32^c	9.45	8.24	8.12	8.87	9.18	9.85

Table 4.5:

Chapter 5

Cointegration and Error Correction Models

The idea of cointegration has, to a great extent, been conceived by Engle and Granger. The notion behind the theory is that variables may move together, in the long run, but seem to deviate from a common dynamic in the short-period movements. The theory of cointegration is economically valuable since it is observable in macroeconomics and finance that a universe of variables seem to be driven by a small number of common factors as is stated, for example, in Diebold (2004).

In this section, the joint behavior of two or more time series will be discussed. Let "x is $I(d)$" denote that the time series x has to be first-differenced d times to become stationary. So, by linearly combining two processes, x_t and y_t, in the fashion

$$u_t = y_t - ax_t,$$

we obtain the $I(d)$ process u_t. There is however the chance that we can find a coefficient a such that u_t might be intergrated of lower order than d, say, k with $0 < k < d$. This property can be described by saying that x_t and y_t

have *long run* components that can be cancelled out by aid of a such that u has no long run components. If this is the case, then x_t and y_t are said to be *cointegrated*.[1]

The previous results can be extended to high dimensions. Then, an n-dimensional vector $\mathbf{z}$ is said to be cointegrated of order $d-k$ ($\mathbf{z}$ is $CI(d, d-k)$) if there exists at least one n-dimensional vector $\mathbf{a}(\neq 0)$ such that the resulting $\mathbf{a}'\mathbf{z}$ is integrated of order k. There may be up to $m < n$ linearly independent vectors $\mathbf{a}_1, \ldots, \mathbf{a}_m$ with that property. Hence, m is the *cointegration rank* of the vector $\mathbf{z}_t$.

When considering cointegration of arbitrarily many processes, one can understand this as a *long-run equilibrium* between them, see Mills (1997). To demonstrate this, we restrict ourselves to the 2-dimensional case, for now, such that the relationship may be thought of as

$$y_t = ax_t.$$

The quantity u_t can be understood as extent of deviation from this equilibrium, hence, as "equilibrium error". When x_t and y_t are $I(1)$, the equilibrium error is, consequently, $I(0)$. Crossings of the zero line by u_t will be the rule, then, whereas large deviations from zero will be rare exception. To the contrary, when x_t and y_t are not cointegrated, then u_t will reveal the reverse patterns.

In the following, notation will be used that is described in Appendix A. Bring to mind condition (c) in the Appendix where non-singularity for the covariance-matrix, Σ, of the innovations, ξ_t, is required. If $|\Sigma| = \sigma_v^2\sigma_w^2 - \sigma_{vw} = 0$, the Σ is no longer non-singular and, hence, the asymptotic results of the Appendix do not hold. This is equivalent to $\sigma_{vw}/(\sigma_v^2\sigma_w^2) = 1$ so that, over a longer time horizon, the correlation coefficient of v and w is one. Thus, singularity of Σ evolves as a necessary condition of cointegration. The case

[1] See, for example, Mills (1997).

of zero correlation implies spurious regression of y_t on x_t. For values between zero and one, x_t and y_t are not cointegrated, however, not independent, either.

The bivariate cointegration of y_t and x_t is, next, extended to a multivariate regression on a vector $\mathbf{x}_t$. The model is

$$y_t = \alpha + \beta \mathbf{x}_t + u_t, \ t = 1, 2, \ldots .$$

Let $\mathbf{x}_t$ itself be modeled by

$$\mathbf{x}_t = \pi + \mathbf{x}_{t-1} + w_t, \ t = 1, 2, \ldots .$$

From the cointegration of y_t and $\mathbf{x}_t$ it follows that $\{u_t\}_1^\infty$ is $I(0)$ which causes $\{\epsilon_t\}_1^\infty = \{(u_t, w_t)\}_1^\infty$ to comply with conditions (a) through (d), in the Appendix to allow for some extent of dependence structure as well as simultaneity and non-stationarity. $E(\mathbf{x}_t u_t) \neq 0$ is permissible, too.

The parameter estimates obtained from the above model with the assumptions on $\{\epsilon_t\}$ are, now, consitent $\hat{\alpha}$ and $\hat{\beta}$ converge favorably at rate $O(1/T)$ compared to the conventional regression's case. They are not asymptotically normal, however, since the sample moments scaled in an appropriate manner converge in distribution to random matrices, see Mills (1997). One may inquire the Appendix for the functionals of Brownian motions obtained for the asymptotic distributions.

Let $\pi = 0$, for now. For the tests associated with the modified statistics, we employ $\{e_t\}$ to denote the estimates of $\{\epsilon_t\} = \{(u_t, w_t)\}$. For these, we obtain as a consistent estimate of the covariance matrix for a long time horizon of the form

$$\Omega_e = \Sigma_e + \Lambda_e + \Lambda_e^{'}$$

where Σ_e denotes the estimate of the contemporaneous covariance matrix of

$\{e_t\}$ and

$$\Lambda_e = \sum_{i=2}^{\infty} E(e_1 e_1^{'}).$$

Standard tests apply when $\{u_t\} \sim WN(0, \sigma^2)$.

For the case $\pi \neq 0$, the regressors, $\mathbf{x}_t$ are perfectly correlated in the limit as a result of the fact that the regressors will be dominated by the $O(T)$ trend πT. This leads to the singularity of the covariance matrix of the $\hat{\beta}$. Detrending of the variables is a solution to obtain the distribution from above.

As a specialization, one can consider the non-zero trend case, $\pi \neq 0$, along with a univariate x_t such that the $\sqrt{T}(\hat{\alpha} - \alpha)$ and $(\sqrt{T})^3(\hat{\beta} - \beta)$ are asymptotically normal with standard formulae found, for example, in Mills (1997), chapter 5.

As a concluding remark to the notion of cointegration, consider the case that the $\{w_t\}$ and $\{u_t\}$ are independent of each other. It can be shown that standardized $\hat{\alpha}$ and $\hat{\beta}$ are, again, asymptotically normal.

A note is left to be dropped on testing. The tests are usually of the sort of a unit root test performed on the regressions residuals obtained from

$$\hat{u}_t = y_t - \hat{\alpha}_t - \hat{\beta}\mathbf{x}_t.$$

The null hypothesis in those tests is commonly that of no cointegration. The Durbin-Watson test is often applied. But, one has to be aware of the problems arising with respect to the process-depending critical values of the dw statistic for various sample sizes. An alternative is a regression of $\nabla\hat{u}_t$ on $\hat{u}_{t-1}$ and lagged $\nabla\hat{u}_t$. The t-statistic of $\hat{u}_{t-1}$ will too often be too high so that the null might even be rejected falsely, i.e., a type one error would occur frequently. The literature provides a great variety of results of critical values from Monte Carlo simulations.

Chapter 6

The Capital Asset Pricing Model

6.1 Review of the CAPM

This section is intended to bring to mind the Capital Asset Pricing Model (CAPM) to those who already have, at least, some knowledge of it as well as provide those with a brief introduction to the model who have never heard of it. The facts delivered will suffice to serve as a basis for latter testing in terms of multi-factor modeling.

The general idea behind the model is to map the tradeoff between expected returns of portfolios and the risk expressed as variance. This was first introduced by Markowitz in 1959 in his Nobel Prize awarded work on investors' portfolio choice optimization. His conclusion was that an investor always selects the portfolio with given expected return that exposes him to the smallest risk, i.e. variance, possible. The converse would be that an investor picks a portfolio with the highest expected returns for a given level of risk that he is still willing to bera.[1] Later, Sharpe and Lintner refined the

[1]This model holds if then investor's utility function is quadratic or the asset returns

risk-return relationship by stating that the return of a portfolio is a linear function of the riskfree rate,[2] the excess return of a *market portfolio*, and an error term such that

$$R_{it} = R_f + \beta_{im}(E[R_m] - R_f) + \epsilon_t, \tag{6.1}$$

$$E[R_i] = R_f + \beta_{im}(E[R_m] - R_f). \tag{6.2}$$

where the $\{\epsilon\}$ are independent mean-zero with variance σ. The regression coefficient *beta* is

$$\beta_{im} = \frac{\text{Cov}[R_i, R_m]}{\text{Var}[R_m]}.$$

This model is referred to as the single-period CAPM. Empirical tests have been performed with respect to whether

(i) R_f is the true intercept,

(ii) the beta captures the entire *cross-sectional* variatin of expected returns, and

(iii) the *market risk premium is positive.*[3]

Often, the model is written in the form

$$E[Z_i] = \beta_{im} E[Z_m] \tag{6.3}$$

where Z_i denotes the excess return of portfolio i. The regression equation for OLS estimation is, then,

$$Z_{it} = a_{im} + \beta_{im} Z_{mt} + \epsilon_{it}.$$

---

are normally distributed.

[2]That is if a riskless asset exists.

[3]The term $E[R_i] - R_f$ is referred to as *risk premium.*

An alternative model, the so called *Black's* model, does not include the riskless asset. It is based on the fact that every[4] portfolio on the efficient set has a unique *zero-covariance portfolio*.[5] That means that every portfolio i has a unique portfolio $0i$ such that $\text{Cov}[R_i, R_{0i}] = 0$. In case of the portfolio with globally minimum variance, the so called *minimum variance portfolio*, has itself as the zero-covariance portfolio. The Black regression model is, then, written as

$$E[R_i] = E[R_{0m}] + \beta_{im}(E[R_m] - E[R_{0m}]) = \alpha_{im} + \beta_{im}E[R_m]. \tag{6.4}$$

The question of what the market portfolio actually is, usually, is answered in tests by using the largest index of the market. In the USA, this is the S&P 500 index.

A note on efficient set theory is dropped, here, to facilitate further understanding. Markowitz's optimization problem can be stated in the following way. Let ω denote the portfolio weights and Ω denote the variance-covariance matrix of the individual assets' returns, then, the optimal portfolios are determined by[6]

$$\begin{aligned} &\min_{\omega} \omega^{'}\Omega\omega \\ &\text{s.t.} \quad \omega^{'}\mu = \mu_{\mathbf{p}} \\ &\omega^{'}\mathbf{1} = \mathbf{1}. \end{aligned}$$

Solving the resulting Lagrangian, we obtain as a result for the optimal portfolio weights

$$\omega_p = \mathbf{g} + \mathbf{h}\mu_{\mathbf{p}}$$

[4]It will be shown that only the global minimum-variance portfolio has positive covariance with every ohther portfolio.

[5]The *efficient set* is the locus of all portfolios that yield a particular expected return with the least amount of risk.

[6]Note that, here, no riskfree asset is available.

where

$$\mathbf{g} = \frac{\mathbf{1}}{\mathbf{D}}[\mathbf{B}(\mathbf{\Omega^{-1}1}) - \mathbf{A}(\mathbf{\Omega^{-1}}\mu)]$$
$$\mathbf{h} = \frac{\mathbf{1}}{\mathbf{D}}[\mathbf{C}(\mathbf{\Omega^{-1}}\mu) - \mathbf{A}(\mathbf{\Omega^{-1}1})],$$

and

$$A = \mathbf{1}'\mathbf{\Omega^{-1}}\mu,$$
$$B = \mu'\mathbf{\Omega^{-1}}\mu,$$
$$C = \mathbf{1}'\mathbf{\Omega^{-1}1}, \text{ and}$$
$$D = BC - A^2.$$

Then, from any two distinct portfolios with minimum variance, the entire efficient frontier can be generated. Also, any linear combination of minimum-variance portfolios results in a portfolio that has minimum variance itself. Additionally, for any two portfolios p and r with minimum variance, the covariance can be written in the form

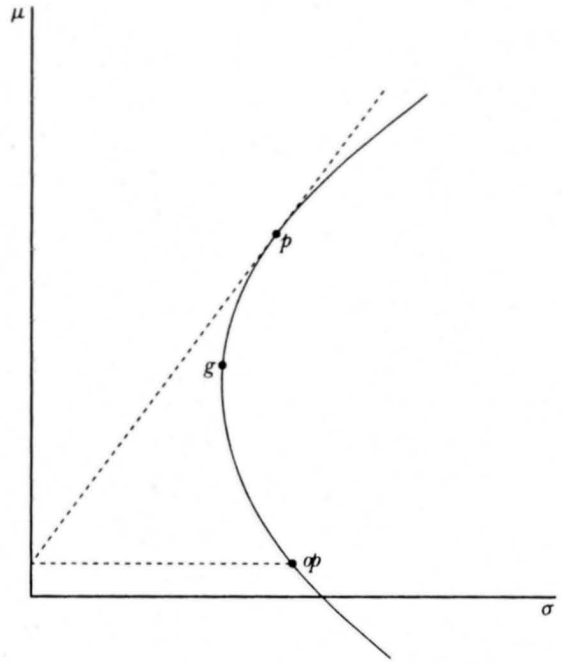
$$\text{Cov}[R_p, R_r] = \frac{C}{D}\left(\mu_p - \frac{A}{C}\right)\left(\mu_r - \frac{A}{C}\right) + \frac{1}{C}.$$

The global minimum-variance portfolio has mean A/C and variance $1/C$. Hence, it has, as an exception, no zero-covariance portfolio.

Any return on a portfolio q can be expressed as a regression on a minimum variance portfolio p and its zero-covariance portfolio $0p$ such that

$$R_q = \beta_0 + \beta_1 R_p + \beta_2 R_{0p} + \epsilon$$
$$E[\epsilon | R_p, R_{0p}] = 0.$$

Figure 6.1 illustrates this. As one can see, graphically, the zero-covariance portfolio $0p$ is determined by running the perpendicular to the y-axis through the intersection of y-axis and the tangent in point p. The intersection of the



Minimum-Variance Portfolios Without Riskfree Asset

perpendicular and the minimum-variance frontier is at the point $0p$. The global minimum variance portfolio is at g.

When a riskfree asset is added,[7] the optimization problem is altered to

$$\min_{\omega} \omega^{'} \Omega \omega$$
$$\text{s.t. } \omega^{'} \mu + (\mathbf{1} - \omega^{'} \mathbf{1}) \mathbf{R_f} = \mu_{\mathbf{p}}.$$

Again, after solving the Lagrangian for the optimal portfolio weight, one obtains for the weights invested in risky assets

$$\omega_p = \frac{(\mu_p - R_f)}{(\mu - R_f \mathbf{1})' \mathbf{\Omega^{-1}} (\mu - \mathbf{R_f} \mathbf{1})} \Omega^{-1} (\mu - R_f \mathbf{1}).$$

[7] E.g. a defaultfree governmant bond.

By setting

$$\bar{\omega} = \Omega^{-1}(\mu - R_f\mathbf{1})$$

and

$$c_p = \frac{(\mu_p - R_f)}{(\mu - R_f\mathbf{1})\Omega^{-1}(\mu - R_f\mathbf{1})}$$

we can express the optimal portfolio weights as proportional to the standardized weight vector

$$\omega_q = \frac{1}{\mathbf{1}'\Omega^{-1}(\mu - R_f\mathbf{1})}\Omega^{-1}(\mu - R_f\mathbf{1}) \tag{6.5}$$

The denominator is required for the vector to sum to one. The portfolio q in 6.5 represents the *tangency portfolio*. It contains zero investment in the riskless asset. Thus, any minimum-variance portfolio is composed of an investment in the riskfree asset and the tangency portfolio, respectively.

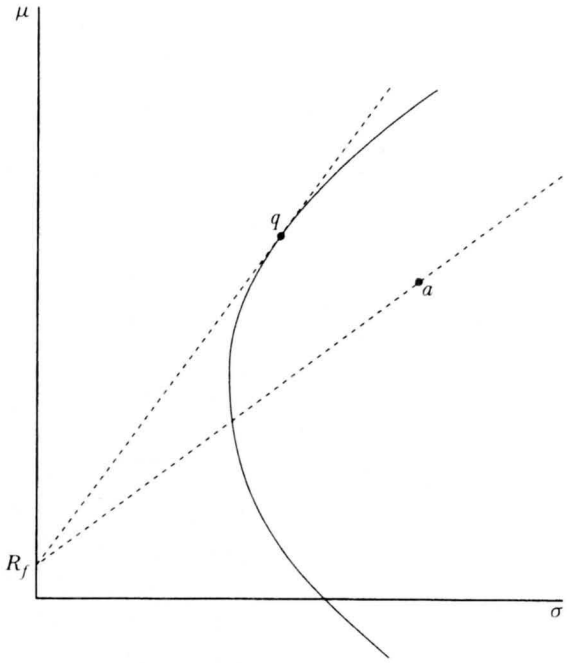
The tangency portfolio is characterized additionally by the fact that it has the biggest *Sharpe ratio*. The Sharpe ratio of the portfolio p with (μ_p, σ_p) is given by the relationship between excess return compensation per unit risk, i.e.

$$sr_p = \frac{\mu_p - R_f}{\sigma_p}.$$

This is demonstrated in Figure 6.1.

6.2 Estimating and Testing the CAPM

Let us take on the CAPM with unrestricted borrowing and lending, first. Let there be N assets with excess returns vector Z_t such that the model to test



Minimum-Variance Portfolios With Riskfree Asset

for is

$$Z_t = a + \beta Z_{mt} + \epsilon_t$$
$$E[\epsilon_t] = 0$$
$$E[\epsilon_t \epsilon_t^{'}] = \Sigma$$
$$E[Z_{mt} = \mu_m,\ E[(Z_{mt} - \mu_m)^2] = \sigma_m^2$$
$$\mathrm{Cov}[Z_{mt}, \epsilon_t] = 0.$$

Under the null hypothesis of the Sharpe-Lintner version, the vector a is zero as well as m, the market portfolio, is equal to the tangent portfolio.

Maximum likelihood tests are performed to retrieve the coefficients a and β.[8] Further, it is assumed that the returns follow a multinomial normal law

[8]They are the same as if estimated by OLS. Minimization is performed via the inverse

such that

$$f(Z_t|Z_{mt}) = \left(\frac{2}{\pi}\right)^{-\frac{N}{2}} |\Sigma|^{-\frac{1}{2}}$$
$$\times \exp\left(-\frac{1}{2}(Z_t - a - \beta Z_{mt})'\Sigma^{-1}(Z_t - a - \beta Z_{mt})\right).$$

Then, persuing the usual process of taking logarithms as well as differentiating with respect to a and β, one obtains the estimates

$$\hat{a} = \hat{\mu} - \hat{\beta}\hat{\mu}_m$$
$$\hat{\beta} = \frac{\sum_{t=1}^{T}(\mathbf{Z_t} - \hat{\mu})(Z_{mt} - \hat{\mu}_m)}{\sum_{t=1}^{T}(Z_{mt} - \hat{\mu})^2} \tag{6.6}$$
$$\hat{\Sigma} = \frac{1}{T}\sum_{t=1}^{T}(Z_t - \hat{a}\hat{\beta}Z_{mt})(Z_t - \hat{a}\hat{\beta}Z_{mt}). \tag{6.7}$$

The mean return parameter estimates are given by

$$\hat{\mu} = \frac{1}{T}\sum_{t=1}^{T}\mathbf{Z}_t$$
$$\hat{\mu}_m = \frac{1}{T}\sum_{t=1}^{T}Z_{mt}.$$

Conditional on the values of $Z_{m1}, \ldots, Z_{mT}$, the distributions of the estimators are[9]

$$\hat{\mathbf{a}} \sim \mathcal{N}\left(a,\ \frac{1}{T}\left[1 + \frac{\hat{\mu}_m^2}{\hat{\sigma}_m^2}\right]\Sigma\right)$$
$$\hat{\beta} \sim \mathcal{N}\left(\beta,\ \frac{1}{T}\left[\frac{1}{\hat{\sigma}_m^2}\Sigma\right]\right)$$
$$T\hat{\Sigma} \sim \mathcal{W}_N(T-2, \Sigma)$$

of the Fisher information matrix which is the negative expectation of the second order derivative of the log-likelihood function.

[9]See, for example, Campbell, Lo, and MacKinlay (1997).

where $\hat{\sigma}_m^2$ is the uncorrected sample variance.[10]

We will consider three statistics to test for the CAPM with a risklees asset, next. First, state the test properly. Let the alternatives be denoted by

$$H_0 : \mathbf{a} = 0 \text{ and}$$
$$H_1 : \mathbf{a} \neq 0.$$

The *Wald* test statistic is given by

$$\begin{aligned} J_0 &= \hat{\mathbf{a}}' [\text{Var}[\hat{\mathbf{a}}]]^{-1} \hat{\mathbf{a}} \\ &= T \left[1 + \frac{\hat{\mu}_m^2}{\hat{\sigma}_m^2} \right]^{-1} \hat{\mathbf{a}}' \Sigma^{-1} \hat{\mathbf{a}}. \end{aligned}$$

Under the null, J_0 will be chi-square distributed with N degrees of freedom. Here, the maximum likelihood estimator $\hat{\Sigma}$ serves as a consistent estimator for Σ so that the distributional null assumption will be guaranteed, asymptotically.

An alternative from the realm of finite-sample distribution by the following statistic

$$J_1 = \frac{T - N - 1}{N} \left[1 + \frac{\hat{\mu}_m^2}{\hat{\sigma}_m^2} \right]^{-1} \hat{\mathbf{a}}' \hat{\Sigma}^{-1} \hat{\mathbf{a}}.$$

Under the null, J_1 is unconditionally distributed central F with N degrees of freedom in the numerator and $(T - N - 1)$ degrees of freedom in the denominator.

As the third alternative, we consider the estimators $\hat{\beta}$ and $\hat{\Sigma}$ form the constrained model, i.e. the Sharpe-Lintner model with $\hat{a} = 0$.[11] We, then, look at the likelihood ratio of the unconstrained versus the constrained model

[10] $\mathcal{W}(\cdot)$ denotes the Wishart distribution with $T - 2$ degrees of freedom.

[11] These estimators β^* and Σ^* are derived by simply setting $\hat{\mu} = 0$ and $\hat{\mathbf{a}} = 0$ in (6.6) and (6.7), respectively,. The estimator β^* has asymptotic variance $1/T(1/\hat{\mu}_m^2 + \hat{\sigma}_m^2)\Sigma$ and Σ^* is distributed with one degre of freedom less.

which is defined by

$$\begin{aligned}\mathcal{LR} &= \mathcal{L}^* - \mathcal{L} \\ &= -\tfrac{T}{2}[log|\Sigma^*| - log|\hat{\Sigma}|].\end{aligned}$$

as the basis for the test with $\mathcal{L}^*$ being the likelihood function of the constrained model.[12] The test statistic is, now, given by

$$\begin{aligned}J_2 &= -2\mathcal{LR} \\ &= T\left[log|\Sigma^*| - log|\hat{\Sigma}|\right] \overset{asymp}{\sim} \chi^2(N).\end{aligned}$$

The statistic J_2 is a monotonic transformation of the finite-sample statistic J_1 and, hence, has a finite-sample distribution related to that of J_1. The relationship can be expressed by

$$\begin{aligned}\beta^* &= \hat{\beta} + \frac{\hat{\mu}_m}{\hat{\mu}_m^2 + \hat{\sigma}_m^2}\hat{\mathbf{a}} \text{ and} \\ \Sigma^* &= \hat{\Sigma} + \left(\frac{\hat{\sigma}_m^2}{\hat{\mu}_m^2 + \hat{\sigma}_m^2}\right)\hat{\mathbf{a}}\hat{\mathbf{a}}'.\end{aligned} \tag{6.8}$$

Developping the determinants of (6.2) and a few intermediate steps not carried out, here, one derives at

$$J_1 = \frac{T - N - 1}{N}\left(\exp\left(\frac{J_2}{T}\right) - 1\right).$$

indicating the likelihood ratio test characteristic of J_1. Because of differences between the large-sample and finite-sample distributions of J_2 under the null, an adjustment is made leading to

$$J_3 = J_2\frac{T - \frac{N}{2} - 2}{T} \overset{asymp.}{\sim} \chi^2(N).$$

Turning our attention, once more, to the Black version of the CAPM

$$E[R_t] = \mathbf{1}E[R_{0mt}] + \beta(E[R_{mt}] - E[R_{0mt}]),$$

[12]For further details on the derivation, the reader is referred Campbell, Lo, and MacKinlay (1997), for example.

the model for the N assets' real returns to be tested is

$$\begin{gathered}\mathbf{R}_t = \mathbf{a} + \beta R_{mt} + \epsilon_t \qquad (6.9)\\ E[\epsilon_t] = 0\\ E[\epsilon_t \epsilon_t^{'}] = \Sigma\\ E[R_{mt}] = \mu_y,\ E[(R_{mt} - \mu_m)^2] = \sigma_m^2\\ \mathrm{Cov}[R_{mt}, \epsilon_t] = 0.\end{gathered}$$

The above is the general unrestricted version of the real returns. If the Black model is true, then $\mathbf{a} = (\mathbf{1} - \beta)E[R_{0mt}]$ has to hold. This is complex because of its non-linear relationship of the regression vectors. Moreover, if the returns are assumed to be *i.d.d.* normally distributed, the Maximum Likelihood estimators of the unrestricted model (6.9) one obtains are

$$\begin{gathered}\hat{\mathbf{a}} = \hat{\mu} - \hat{\beta}\hat{\mu}_m\\ \hat{\beta} = \frac{\sum_{t=1}^{T}(\mathbf{R}_t - \hat{\mu})(R_{mt} - \hat{\mu}_m)}{\sum_{t=1}^{T}(R_{mt} - \hat{\mu}_m)^2}\\ \hat{\Sigma} = \frac{1}{T}\sum_{t=1}^{T}(\mathbf{R}_t\hat{\mathbf{a}}\hat{\beta}R_{mt})(\mathbf{R}_t\hat{\mathbf{a}} - \hat{\beta}R_{mt})^{'}\\ \hat{\mu} = \frac{1}{T}\sum_{t=1}^{T}\mathbf{R}_t,\ \hat{\mu}_m = \frac{1}{T}\sum_{t=1}^{T}R_{mt}.\end{gathered}$$

The distribution of the estimators conditional on the $(R_{m1}, \ldots, R_{mT})^{'}$ are

$$\begin{gathered}\hat{\mathbf{a}} \sim \mathcal{N}\left(\mathbf{a}, \frac{1}{T}\left[1 + \frac{\hat{\mu}_m^2}{\hat{\sigma}_m^2}\Sigma\right]\right)\\ \hat{\beta} \sim \mathcal{N}\left(\beta, \frac{1}{T}\left[\frac{1}{\sigma_m^2}\right]\right)\\ T\hat{\Sigma} \sim \mathcal{W}_N(T-2, \Sigma)\\ \hat{\sigma}_m^2 = \frac{1}{T}\sigma_{t=1}^{T}(R_{mt} - \hat{\mu}_m)^2.\end{gathered}$$

The relationship between $\hat{\mathbf{a}}$ and $\hat{\beta}$ can be expressed by their covariance

$$\mathrm{Cov}[\hat{\mathbf{a}}, \hat{\beta}] = -\left[\frac{\hat{\mu}_m}{\hat{\sigma}_m^2}\right]\Sigma.$$

Now, with threstriction from the Black's model, the log-likelihood function is

$$\begin{aligned}\mathcal{L}(E[R_{0mt}], \beta, \Sigma) &= -\tfrac{NT}{2}\log(2\pi) - \tfrac{T}{2}\log|\Sigma| \\ &\quad -\tfrac{1}{2}\textstyle\sum_{t=1}^{T}(\mathbf{R}_t - E[R_{0mt}](\mathbf{1}-\beta) - \beta R_{mt})'\Sigma^{-1} \\ &\quad \times(\mathbf{R}_t - E[R_{omt}](1-\beta) - \beta R_{mt}).\end{aligned}$$

The ususal differentiation with respect to the unknown parameters[13] yields the estimates

$$\begin{aligned}\hat{E}^*[R_{0mt}] &= \frac{(\mathbf{1}\hat{\beta}^*)'\hat{\Sigma}^{*-1}(\hat{\mu} - \hat{\mu} - \hat{\beta})\hat{\mu}_m}{(\mathbf{1}-\hat{\beta}^*)'\hat{\Sigma}^{*-1}(\mathbf{1}-\hat{\beta}^*)} \\ \hat{\beta}^* &= \frac{\sum_{t=1}^{T}(\mathbf{R}_t - \hat{E}^*[R_{0mt}]\mathbf{1})(R_{mt} - \hat{E}^*[R_{0mt}])}{\sum_{t=1}^{T}(R_{mt} - \hat{E}^*[R_{0mt}])^2} \\ \hat{\Sigma}^* &= \tfrac{1}{T}\textstyle\sum_{t=1}^{T}(\mathbf{R}_t - \hat{E}^*[R_{0mt}](\mathbf{1}-\hat{\beta}^*) - \hat{\beta}^* R_{mt}) \\ &\quad \times(\mathbf{R}_t - \hat{E}^*[R_{0mt}](\mathbf{1}-\hat{\beta}*) - \hat{\beta}^* R_{mt}).\end{aligned}$$

Usually, one takes the unconstrained estimates as initial values since the constrained estimates cannot be solved explicitly for.

For comparison of the two models, the constrained as well as the unconstrained, a likelihho ratio test can be performed using the unconstrained and constrained Maximum Likelihood estimates. The null hypothesis is, then, stated as

$$\begin{aligned}H_0 &: \mathbf{a} = (\mathbf{1}-\beta)E^*[R_{0mt}] \\ H_1 &: \mathbf{a} \neq (\mathbf{1}-\beta)E^*[R_{0mt}].\end{aligned}$$

The corresponding statistic is given by

$$J_4 = T[\log|\hat{\Sigma}^*| - \log|\hat{\Sigma}|] \overset{asymp.}{\sim} \chi^2(N-1).$$

[13] Omitted, here.

Notice that, here, a dgree of freedom is lost compared to the Sharpe-Lintner version because it is used up for the estimation of the zero-beta expected return. For the residual covariance matrix, $N(N-1)/2$ parameters have to be estimated in addition to $2N$ parameters for $\mathbf{a}$ and β. In case of the constrained model, there are the same number of parameters for the covariance matrix and parameter β but only one estimate for $E^*[R_{0mt}]$ yielding a surplus of $N-1$ free parameters over the unconstrained version.

It is argued in literature that the true distribution of J_4 in finite samples does not match the χ^2 very well. Just as in the Sharpe-Lintner version, the finite-sample properties of the likelihood ratio statistic can be improved by an alteration resulting in

$$J_5 = \left(T - \frac{N}{2} - 2\right)\left[\log|\hat{\Sigma}^*| - \log|\hat{\Sigma}|\right] \overset{asymp.}{\sim} \chi^2(N-1).$$

Conducting the estimation via iteration of the first-order conditions as well as relying on asymptotic behavior can be very unsatisfactory as argued in Campbell, Lo, and MacKinlay (1997). A solution is given by the following alternative approach.[14] Regress on the relationship of the excess returns, i.e.

$$\mathbf{R}_t - E[R_{0mt}] = \mathbf{a} + \beta(R_{mt} - E[R_{0mt}]) + \epsilon_t \tag{6.10}$$

with known $E[R_{0mt}]$. Then, Maximum Likelihood estimation yields the estimators

$$\hat{\mathbf{a}}(E[R_{0mt}]) = \hat{\mu} - E[R_{0mt}]\mathbf{1} - \hat{\beta}(\hat{\mu}_m - E[R_{0mt}]),$$
$$\hat{\beta} = \frac{\sum_{t=1}^{T}(\mathbf{R}_t - \hat{\mu})(R_{mt} - \hat{\mu})}{\sum_{t=1}^{T}(R_{mt} - \hat{\mu})^2},$$
$$\hat{\Sigma} = \frac{1}{T}\sum_{t=1}^{T}[\mathbf{R}_t - \hat{\mu} - \hat{\beta}(R_{mt}\hat{\mu})][\mathbf{R}_t - \hat{\mu} - \hat{\beta}(R_{mt}\hat{\mu})]'.$$

[14] The procedure presented follows Campbell, Lo, and MacKinlay (1997), pp. 200 closely.

The unconstrained log-likelihood ratio independent of $E[R_{0mt}]$ is given to be

$$\mathcal{L} = -\frac{NT}{2}\log(2\pi) - \frac{T}{2}\log|\hat{\Sigma}| - \frac{NT}{2}. \tag{6.11}$$

The constrained estimates obtained by setting $\mathbf{a} = (\mathbf{1} - \beta)E[R_{0mt}]$ are

$$\begin{aligned}
\hat{\beta}^* &= \frac{\sum_{t=1}^{T}(\mathbf{R}_t - E[R_{0mt}]\mathbf{1})(R_{mt} - E[R_{0mt}])}{\sum_{t=1}^{T}(R_{mt} - E[R_{0mt}])^2}, \\
\hat{\Sigma}^* &= \frac{1}{T}\sum_{t=1}^{T}\left[\mathbf{R}_t - E[R_{0mt}](\mathbf{1} - \hat{\beta}^*) - \hat{\beta}^* R_{mt}\right] \\
&\quad \times \left[\mathbf{R}_t - E[R_{0mt}](\mathbf{1} - \hat{\beta}^*) - \hat{\beta}^* R_{mt}\right]'
\end{aligned}$$

turning (6.11) into

$$\mathcal{L}^*(E[R_{0mt}]) = -\frac{NT}{2}\log(2\pi) - \frac{T}{2}\log|\hat{\Sigma}^*(E[R_{0mt}])| - \frac{NT}{2}.$$

The according log-likelihood ratio is, then,

$$\mathcal{LR}(E[R_{0mt}]) = -\frac{T}{2}\left[\log|\hat{\Sigma}^*(E[R_{0mt}])| - \log|\hat{\Sigma}|\right]. \tag{6.12}$$

Without presenting the individual steps, the optimal $E[R_{0mt}]$ is retrieved by minimizing (6.12).[15] Generally, the optimal estimator, $\hat{E}^*[R_{0mt}]$, is the solution to a quadratic equation. The benefit of this procedure is that we arrive at the estimators analyticaly instead of iteratively.

The variance of the estimator is

$$\mathrm{Var}(\hat{E}^*[R_{0mt}]) = \frac{1}{T}\left(1 + \frac{(\mu_m - E[R_{0mt}])^2}{\sigma_m^2}\right)[(\mathbf{1} - \beta)\Sigma^{-1}(\mathbf{1} - \beta)]^{-1}$$

which is necessary for inferences concerning the true value of the parameter. Theoretically, as can be seen, the true values of β and Σ have to be inserted. But evaluation at the Maximum Likelihood estimates yield confidence bounds based on the asymptotic normality of $\hat{E}^*[R_{0mt}]$.

[15] For the reader interested in the details, Campbell, Lo, and MacKinlay (1997) is strongly recommended.

Now, testing for a zero intercept in (6.10), one can resort to the statistic

$$J_6(E[R_{0mt}]) = \frac{T-N-1}{N}\left[1+\frac{(\hat{\mu}_m - E[R_{0mt}])^2}{\hat{\sigma}_m^2}\right]\hat{\mathbf{a}}(E[R_{0mt}])\hat{\Sigma}^{-1}\hat{\mathbf{a}}(E[R_{0mt}])$$

which, under the null hypothesis of a zero intercept, is distributed $F(N, T-N-1)$. Usually, $E[R_{0mt}]$ is not known, however. Still the test can be performed aproximately using $\hat{E}^*[R_{0mt}]$. However, it will be harder to reject the null compared to the true value of $E[R_{0mt}]$ due to the fact that $\hat{E}^*[R_{0mt}]$ minimizes J_6.

A problem arises when infering if one falsely relies on either normal or *i.i.d.* returns. Theoretically, the CAPM does not exclusively hold for normal returns, though, this assumption is sufficient. It looks different in the case of temporal dependence of returns. But, still, empirical testing of the performance may be enlightening.

Given the ergodicity, finite fourth moments, and stationarity, a Generalized Method of Moments (GMM) is used, here, to conduct a robust test of the Sahrpe-Lintner version of the CAPM.[16] Let the notation be as follows,

$$\begin{gathered}\mathbf{h}_t^{'} = [1 Z_{mt}],\\ \epsilon_t = \mathbf{Z_t} - \mathbf{a} - \beta Z_{mt}, \text{ and}\\ \theta^{'} = [\mathbf{a}^{'}\ \beta^{'}].\end{gathered}$$

Define

$$\mathbf{f}_t(\theta) = \mathbf{h}_t \otimes \epsilon$$

with $x \otimes y$ denoting the $2 \times N$ products of each component of x with each component of y. For the true parameter vector, θ_0, $E[\mathbf{f}_t(\theta_0)]$ has to hold.

[16]For a detailed account of the method, the reader is referred to Campbell, Lo, and MacKinlay (1997).

This moment condition is approached by the sample average of the

$$\mathbf{g}_T(\theta) = \frac{1}{T}\sum_{t=1}^{T}\mathbf{f}_t(\theta).$$

The GMM estimator $\hat{\theta}$ has to satisfy

$$\min_{\theta}\mathbf{g}_T(\theta)^{'}\mathbf{W}\mathbf{g}_T(\theta) \tag{6.13}$$

with positive definite $2N \times 2N$ weighting matrix $\mathbf{W}$. Now, pick $\hat{\theta}$ as the estimate setting the sample average of the $\mathbf{g}_T$ equal to zero. This is legitimate since the system of $2N$ conditions is exactly identified. The GMM estimator is independent of the matrix $\mathbf{W}$ in (6.13). The estimators are

$$\hat{\mathbf{a}} = \hat{\mu} - \hat{\beta}\hat{\mu}_m,$$
$$\hat{\beta} = \frac{\sum_{t=1}^{T}(\mathbf{Z}_t - \hat{\mu})(Z_{mt} - \hat{\mu}_m)}{\sum_{t=1}^{T}(Z_{mt} - \hat{\mu})^2}.$$

Setting

$$\mathbf{D}_0 = E\left[\frac{\partial \mathbf{g}_T(\theta)}{\partial \theta'}\right], \quad \text{and}$$
$$\mathbf{S}_0 = \sum_{l=-\infty}^{+\infty} E[\mathbf{f}_t(\theta)\mathbf{f}_t t - l(\theta)^{'}],$$

the distribution of the estimator is given by

$$\hat{\theta} \overset{asymp.}{\sim} \mathcal{N}\left(\theta, \frac{1}{T}[\mathbf{D}_0^{'}\mathbf{S}_0^{-1}\mathbf{D}_0]^{-1}\right).$$

Consistent estimators of $\mathbf{D}_0$ are found by inserting the Maximum Likelihood estimates of μ_m and σ_m^2. For details of a consistent estimator, $\mathbf{S}_T$, of $\mathbf{S}_0$, the reader is asked to consult Campbell, Lo, and MacKinlay (1997). Skipping further details, a test statistic fo rthe Sharpe-Linter CAPM is, then, provided by

$$J_7 = T\hat{\mathbf{a}}^{'}\left[\mathbf{R}[\mathbf{D}_T^{'}\mathbf{S}_T^{-1}\mathbf{D}_T]\mathbf{R}^{'}\right]^{-1}\hat{\mathbf{a}}$$

which under the null hypothesis of $\mathbf{a} = 0$ is asymptotically distributed $\chi^2(N)$.

6.3 Multifactor Models

Motivation

The opinion of the validity as well as usefulness is controversial. This is despite the fact that the CAPM is still commonly used in practice. Some say, however, that the notion of the market portfolio serving as the single factor is ambiguous. Moreover, one factor alone is considered insufficient given the vast varitey of economic quantities with the potential of having influence on prices of traded assets. Hence, multifactor models have been introduced. Basically, two main approaches exist. One is the Arbitrage Pricing Theory (APT) by Ross and the other is the Intertemporal Capital Asset Pricing Model (ICAPM) by Merton.

The APT admits multiple risk factors, here K. In contrast to the CAPM, there is no need for a market portfolio. The assumptions made are that markets are competitive and that they are frictionless. The model looks like

$$\mathbf{R} = \mathbf{a} + \mathbf{B}\mathbf{f} + \epsilon$$
$$E[\epsilon|\mathbf{f}] = 0$$
$$E[\epsilon\epsilon^{'}|\mathbf{f}] = \Sigma$$

where $\mathbf{R}$ is the n-dimensional return vector of the n assets, $\mathbf{a} \in \mathbb{R}^n$, $\mathbf{B}$ is the $n \times K$ matrix of factor loadings, and ϵ is the n-dimensional vector of approximately uncorrelated, asset-specific noise terms. The factors are assumed to be the source of the common variation of the returns. Hence, the overall disturbance vanishes in a well-diversified portfolio. The name is intuitive since the theory of the exact model prohibits the creation of any arbitrage possibilities. Hence, a portfolio that is created of an ever growing number of assets with zero initial investment and bearing no risk should be expected to yield zero return. The unsystematic risk is diversified away by $n \to \infty$. Bearing no systematic risk can be achieved by composing a

portfolio such that the accumulated factor sensitivities add to unity. The expected value of the portfolio, then, is a constant required to be zero.

When there is arbitrage, however, it is shown that the APT is no longer exact such that

$$\mu \approx \mathbf{1}\lambda_0 + \mathbf{B}\lambda_{\mathbf{K}}.$$

Here, λ_0 is a zero-beta parameter possibly equal to the riskless asset if exists and λ_K is a $(K \times 1)$ vector of factor risk premia.

In contrast to the APT, the ICAPM is an exact factor pricing model. Also an unshared feature is that the ICAPM requires certain assumption with respect to the conditional distribution of the returns. Just like the APT, it is a multi-factor model, however, one factor is the market portfolio and the remaining are state variables.

In the succeeding analysis, exact factor pricing will be under scrutiny, only. One factor is a market portfolio proxy such as an index and the additional factors may be returns of real traded portfolios or, to the contrary, any other quantities that are not traded, at all.

Estimation and Testing

Before we approach the different models, a few prerequisites have to be mentioned. Throughout time, returns conditional on the factor realizations are understood to be serially *i.i.d.*. Jointly, they are multivariate normal. The factors will neither be determined in number nor in nature. All estimation will be performed by means of the Maximum Likelihood method. For all four subsequent models, likelihood ration tests will be appropriate to test for the significance of the respective constraints. The statistic will be of the familiar form

$$J = -\left(T - \frac{N}{2} - K - 1\right)\left(\log|\hat{\Sigma}| - \log|\hat{\Sigma}^*|\right)$$

where the estimates of the residual covariance matrices of the unconstrained models are $\hat{\Sigma}$, respectively, and of the constrained models $\hat{\Sigma}^*$, respectively. T indicates the length of the time series of observations and N indicates the number of included portfolios. There will be K factors throughout. The term in front of the difference of the logarithms is a scale factor to improve convergence of the finite-sample null distribution to its large sample counterpart which is χ^2 with as many degrees of freedom as there are restrictions stated by the null hypothesis.

The first of the four alternative models employs portfolios as factors in addition to the existence of a riskfree asset. Let the unconstrained model of the excess returns $\mathbf{Z}_t$ be formulated as

$$\begin{gathered}
\mathbf{Z}_t = \mathbf{a} + \mathbf{B}\mathbf{Z}_{Kt} + \epsilon_t \\
E[\epsilon_t] = 0 \\
E[\epsilon\epsilon^{'}] = \Sigma \\
E[\mathbf{Z}_{Kt}] = \mu_K,\ E[(\mathbf{Z}_{Kt} - \mu_K)(\mathbf{Z}_{Kt} - \mu_K)^{'}] = \Omega_K \\
\mathrm{Cov}[\mathbf{Z}_{Kt}, \epsilon_t^{'}] = \mathbf{0}
\end{gathered}$$

where the $\mathbf{Z}_{Kt}$ are the factor excess returns, $\mathbf{B}$ is the $(N \times K)$ matrix of factor sensitivities, $\mathbf{a}$ is the vector of asset returns intercepts, and the ϵ_t are disturbances. The covariance matrices of the disturbances and the factor portfolio excess returns, respectively, are indicated accordingly.

The estimators of the unconstrained model are

$$\begin{gathered}
\hat{\mathbf{a}} = \hat{\mu} - \hat{\mathbf{B}}\hat{\mu}_K \\
\hat{\mathbf{B}} = \left[\sum_{t=1}^{T}(\mathbf{Z}_t - \hat{\mu})(\mathbf{Z}_{Kt} - \hat{\mu}_K)^{'}\right]\left[\sum_{t=1}^{T}(\mathbf{Z}_{Kt} - \hat{\mu}_K)(\mathbf{Z}_{Kt} - \hat{\mu}_K)^{'}\right]^{-1} \\
\hat{\Sigma} = \frac{1}{T}\sum_{t=1}^{T}(\mathbf{Z}_t - \hat{\mathbf{a}} - \hat{\mathbf{B}}\mathbf{Z}_{Kt})(\mathbf{Z}_t - \hat{\mathbf{a}} - \hat{\mathbf{B}}\mathbf{Z}_{Kt})^{'}
\end{gathered}$$

where

$$\hat{\mu} = \frac{1}{T}\sum_{t=1}^{T}\mathbf{Z}_t \text{ and } \hat{\mu}_K = \frac{1}{T}\sum_{t=1}^{T}\mathbf{Z}_{Kt}.$$

When the constraint of $\mathbf{a} = 0$ holds, the estimators are

$$\hat{\mathbf{B}}^* = \left[\sum_{t=1}^{T}\mathbf{Z}_t\mathbf{Z}_{Kt}^{'}\right]\left[\sum_{t=1}^{T}\mathbf{Z}_t\mathbf{Z}_{Kt}^{'}\right]^{-1}$$

$$\hat{\Sigma}^* = \frac{1}{T}\sum_{t=1}^{T}(\mathbf{Z}_t - \hat{\mathbf{B}}^*\mathbf{Z}_{Kt})(\mathbf{Z}_t - \hat{\mathbf{B}}^*\mathbf{Z}_{Kt})^{'}.$$

The likelihood ratio test statistic J, under the null, is distributed χ^2 with N degrees of freedom due to the number of restrictions being N. An exact F-test which is possible, here, can be performed using the statistic

$$J_1 = \frac{T-N-K}{N}[1 + \hat{\mu}_K^{'}\hat{\Omega}_k^{-1}\hat{\mu}_K]^{-1}\hat{\mathbf{a}}^{'}\hat{\Sigma}^{-1}\hat{\mathbf{a}}$$

with

$$\hat{\Omega}_K = \frac{1}{T}\sum_{t=1}^{T}(\mathbf{Z}_{Kt} - \hat{\mu}_K)(\mathbf{Z}_{Kt} - \hat{\mu}_K)^{'}.$$

Under the null hypothesis of the restricted model, J_1 is unconditionally distributed central F with N degrees of freedom in the numerator and $(T - N - K)$ degrees of freedom in the denominator. The benefit, in this case, is the exact distribution of the statistic circumventing the problems arising with asymptotic distribution theory.[17]

The second model still uses traded portfolios as factors. In contrast to the model just presented, however, we drop the assumption of the existence of a riskless asset. Now, a multifactor zero-beta equivalent to the CAPM is created. The zero-beta portfolio is conceived as having no sensitivity to

[17]See Campbell, Lo, and MacKinlay (1997).

any other factor. Let $\mathbf{R}_t$ denote the N-dimensional vector of real returns for the N assets. The factor sensitivites are expressed by the $(N \times K)$ matrix $\mathbf{B}$. The N-dimensional vectors $\mathbf{a}$ and ϵ_t are the asset return intercept and disturbenace, respectively. The unconstrained K-factor linear model, then, is given by

$$\begin{gathered}
\mathbf{R}_t = \mathbf{a} + \mathbf{B}\mathbf{R}_{Kt} + \epsilon_t \\
E[\epsilon_t] = 0 \\
E[\epsilon_t \epsilon_t^{'}] = \Sigma \\
E[\mathbf{R}_{Kt}] = \mu_K, \; E[(\mathbf{R}_{Kt} - \mu_K)(\mathbf{R}_{Kt} - \mu_K)^{'}] = \Omega_K \\
\mathrm{Cov}[\mathbf{R}_{Kt}, \epsilon_t^{'}] = \mathbf{0}.
\end{gathered}$$

The Maximum Likelihood estimators are obtained to be

$$\begin{gathered}
\hat{\mathbf{a}} = \hat{\mu} - \hat{\mathbf{B}}\hat{\mu}_K \\
\hat{\mathbf{B}} = \left[\sum_{t=1}^{T}(\mathbf{R}_t - \hat{\mu})(\mathbf{R}_{Kt} - \hat{\mu}_K)^{'}\right]\left[\sum_{t=1}^{T}(\mathbf{R}_{Kt} - \hat{\mu}_K)(\mathbf{R}_{Kt} - \hat{\mu}_K)^{'}\right]^{-1} \\
\hat{\Sigma} = \frac{1}{T}\sum_{t=1}^{T}(\mathbf{R}_t - \hat{\mathbf{a}} - \hat{\mathbf{B}}\mathbf{R}_{Kt})(\mathbf{R}_t - \hat{\mathbf{a}} - \hat{\mathbf{B}}\mathbf{R}_{Kt})^{'}
\end{gathered}$$

with

$$\hat{\mu} = \frac{1}{T}\sum_{t=1}^{T}\mathbf{R}_t \text{ and } \hat{\mu} = \frac{1}{T}\sum_{t=1}^{T}\mathbf{R}_{Kt}.$$

In the unconstrained model, the asset returns enter into a linear relationship with the factor senistivity matrix. The factors are assumed to be portfolio returns in excess of the zero-beta portfolio in the form of the intercept.

$$\mathbf{R}_t = \mathbf{1}\gamma_0 + \mathbf{B}(\mathbf{R}_{Kt} - \mathbf{1}\gamma_0) + \epsilon_t = (\mathbf{1} - \mathbf{B1})\gamma_0 + \mathbf{B}\mathbf{R}_{Kt} + \epsilon_t.$$

The estimators of the constrained model are given by

$$\begin{aligned}
\hat{\mathbf{B}}^* &= \left[\sum_{t=1}^{T}(\mathbf{R}_t - \mathbf{1}\hat{\gamma}_0)(\mathbf{R}_{Kt} - \mathbf{1}\hat{\gamma}_0)'\right] \\
&\times \left[\sum_{t=1}^{T}(\mathbf{R}_{Kt} - \mathbf{1}\hat{\gamma}_0)(\mathbf{R}_{Kt} - \mathbf{1}\hat{\gamma}_0)'\right]^{-1} \\
\hat{\Sigma}^* &= \frac{1}{T}\sum_{t=1}^{T}[\mathbf{R}_t - \mathbf{1}\hat{\gamma}_0 - \hat{\mathbf{B}}^*(\mathbf{R}_{Kt} - \mathbf{1}\hat{\gamma}_0)] \\
&\times[\mathbf{R}_t - \mathbf{1}\hat{\gamma}_0 - \hat{\mathbf{B}}^*(\mathbf{R}_{Kt} - \mathbf{1}\hat{\gamma}_0)]' \\
\hat{\gamma}_0 &= [(\mathbf{1} - \hat{\mathbf{B}}^*\mathbf{1})'\hat{\Sigma}^{*-1}(\mathbf{1} - \hat{\mathbf{B}}^*\mathbf{1})]^{-1} \\
&\times[(\mathbf{1} - \hat{\mathbf{B}}^*\mathbf{1})'\hat{\Sigma}^{*-1}(\hat{\mu} - \hat{\mathbf{B}}^*\mathbf{1})]. \qquad (6.14)
\end{aligned}$$

The asmptotic variance of $\hat{\gamma}_0$ is

$$\begin{aligned}
\mathrm{Var}[\hat{\gamma}_0] &= \frac{1}{T}\left(1 + (\hat{\mu}_K - \hat{\gamma}_0\mathbf{1})'\Omega_K^{-1}(\hat{\mu}_k - \hat{\gamma}_0\mathbf{1})\right) \\
&\times[(\mathbf{1} - \hat{\mathbf{B}}^*\mathbf{1})'\hat{\Sigma}^{*-1}(\mathbf{1} - \hat{\mathbf{B}}^*\mathbf{1})]^{-1}.
\end{aligned}$$

The estimators are found iteratively by entering the unconstrained estimates for $\mathbf{B}$ and Σ into (6.3). The restriction is

$$\mathbf{a} = (\mathbf{1} - \mathbf{B}\mathbf{1})\gamma_0.$$

The log-likelihood ration test statistic J has $N-1$ degrees of freedom, under the null hypothesis. This is one degree less compared to the unconstrained model since, in the constrained model, the return on the zero-beta portfolio, γ_0 has to estimated.

The third model assumes that the factors are non-traded portfolios of, for example, macroeconomic quantities such as changes in the GNP, interest rates, bond yields, and unemployment etc. We will test exact models. With the notation from the previous settings, we only introduce as additional variables the $(K \times 1)$ vector of factor realizations, $\mathbf{f}_{Kt}$. Then, we have for the

unconstrained model

$$\mathbf{R}_t = \mathbf{a} + \mathbf{B}\mathbf{f}_{Kt} + \epsilon_t$$
$$E[\epsilon_t] = 0$$
$$E[\epsilon_t \epsilon_t^{'}] = \Sigma$$
$$E[\mathbf{f}_{Kt}] = \mu_{fK},\ E[(\mathbf{f}_{Kt} - \mu_{fK})(\mathbf{f}_{Kt} - \mu_{fK})^{'}] = \Omega_K$$
$$\mathrm{Cov}[\mathbf{f}_{Kt}, \epsilon_t^{'}] = \mathbf{0}.$$

The Maximum Likelihood estimators are

$$\hat{\mathbf{a}} = \hat{\mu} - \hat{\mathbf{B}}\hat{\mu}_{fK}$$
$$\hat{\mathbf{B}} = \left[\sum_{t=1}^{T}(\mathbf{R}_t - \hat{\mu})(\mathbf{f}_{Kt} - \hat{\mu}_{fK})^{'}\right]\left[\sum_{t=1}^{T}(\mathbf{f}_{Kt} - \hat{\mu}_{fK})(\mathbf{f}_{Kt} - \hat{\mu}_{fK})^{'}\right]^{-1}$$
$$\hat{\Sigma} = \frac{1}{T}\sum_{t=1}^{T}(\mathbf{R}_t - \hat{\mathbf{a}} - \hat{\mathbf{B}}\mathbf{f}_{Kt})(\mathbf{R}_t - \hat{\mathbf{a}} - \hat{\mathbf{B}}\mathbf{f}_{Kt})^{'}$$

Again,

$$\hat{\mu} = \frac{1}{T}\sum_{t=1}^{T}\mathbf{R}_t \text{ and } \hat{\mu}_{fK} = \frac{1}{T}\sum_{t=1}^{T}\mathbf{f}_{Kt}.$$

The constrained model is found by setting equal the unconditional expectations

$$\mu = \mathbf{a} + \mathbf{B}\mu_{fK} \stackrel{!}{=} \mathbf{1}\lambda_0 + \mathbf{B}\lambda_K$$

to yield the restriction

$$\begin{aligned}\mathbf{a} &= \mathbf{1}\lambda_0 + \mathbf{B}(\lambda_K - \mu_{fK}) \\ &\equiv \mathbf{1}\lambda_0 + \mathbf{B}\gamma_1\end{aligned}$$

with the K-dimensional vector of the factor risk premia λ_K. The constrained regression is, now,

$$\mathbf{R}_t = \mathbf{1}\gamma_0 + \mathbf{B}\gamma_1 + \mathbf{B}\mathbf{f}_{Kt} + \epsilon_t.$$

where $\gamma_0 \equiv E[\lambda_0]$. Then, the estimators are

$$\hat{\mathbf{B}}^* = \left[\sum_{t=1}^{T}(\mathbf{R}_t - \mathbf{1}\hat{\gamma}_0)(\mathbf{f}_{Kt} + \gamma_1)'\right]\left[\sum_{t=1}^{T}(\mathbf{f}_{Kt} + \gamma_1)(\mathbf{f}_{Kt} + \gamma_1)'\right]^{-1}$$

$$\hat{\Sigma}^* = \frac{1}{T}\sum_{t=1}^{T}[(\mathbf{R}_t - \mathbf{1}\hat{\gamma}_0) - \hat{\mathbf{B}}^*(\mathbf{f}_{Kt} + \hat{\gamma}_1)]$$
$$\times[(\mathbf{R}_t - \mathbf{1}\hat{\gamma}_0) - \hat{\mathbf{B}}^*(\mathbf{f}_{Kt} + \hat{\gamma}_1)]'$$

$$\hat{\gamma} = [\mathbf{X}'\hat{\Sigma}^*\mathbf{X}]^{-1}[\mathbf{X}'\hat{\Sigma}^{*-1}(\hat{\mu} - \hat{\mathbf{B}}^*\hat{\mu}_{Kt})]$$

with $\gamma \equiv [\gamma_0, \gamma_1']'$ and $\mathbf{X} \equiv [\mathbf{1}\hat{\mathbf{B}}^*]$. The asymptotic variance of $\hat{\gamma}$ is given by

$$\text{Var}[\hat{\gamma}] = \frac{1}{T}\left(1 + (\hat{\gamma}_1 + \hat{\mu}_{fK})'\hat{\Omega}_K^{-1}(\hat{\gamma}_1 + \hat{\mu}_{fK})\right)$$
$$\times[\mathbf{X}'\hat{\Sigma}^{*-1}\mathbf{X}]^{-1}.$$

The asymptotic variances of the respective components of $\hat{\gamma}$ can be found in many text books on this subject such as Campbell, Lo, and MacKinlay (1997).

The fourth model incorporates portfolios spanning the mean-variance efficient frontier. Hence, the intercept λ_0 is zero. The portfolios are well-diversified in the APT sense. With this particularity, the setting is left unchanged with respect to the previious three alternatives. The unconstrained model, then, is exactly that of the unconstrained model of the second case. The constraints are as follows,

$$\mathbf{a} = 0 \text{ and } \mathbf{B}_1 = \mathbf{1}.$$

This can be made more accessible, intuitively, by referring to Black's version of the CAPM. Let it be written as an unconstrained model in the form

$$\mathbf{R}_t = \mathbf{a} + \beta_{0m}\mathbf{R}_{0t} + \beta_m\mathbf{R}_{mt} + \epsilon_t$$

where $\mathbf{R}_{mt}$ and $\mathbf{R}_{0t}$ are the returns on the market portfolio and its zero-beta portfolio, respectively. In that instance, $\mathbf{a} = 0$ and $\beta_{0m} + \beta_m = \mathbf{1}$ which is equivalent to the restrictions from above. The unconstrained estimators are

$$\begin{aligned}
\hat{\mathbf{a}} &= \hat{\mu} + \hat{\mathbf{B}}\hat{\mu}_K \\
\hat{\mathbf{B}} &= \left[\sum_{t=1}^{T}(\mathbf{R}_t - \hat{\mu})(\mathbf{R}_{Kt} - \hat{\mu}_K)'\right] \\
&\quad \times \left[(\mathbf{R}_{Kt} - \hat{\mu}_K)(\mathbf{R}_{Kt} - \hat{\mu}_K)'\right] \\
\hat{\Sigma} &= \frac{1}{T}\sum_{t=1}^{T}(\mathbf{R}_t - \hat{\mathbf{a}} - \hat{\mathbf{B}}\mathbf{R}_{Kt})(\mathbf{R}_t - \hat{\mathbf{a}} - \hat{\mathbf{B}}\mathbf{R}_{Kt})'
\end{aligned}$$

with

$$\hat{\mu} = \frac{1}{T}\sum_{t=1}^{T}\mathbf{R}_t \text{ and } \hat{\mu} = \frac{1}{T}\sum_{t=1}^{T}\mathbf{R}_{Kt}.$$

The constrained model, basically, has the same shape as the unconstrained counterpart. We just have to partition $\mathbf{B}$ into a $(N \times 1)$ column vector $\mathbf{b}_1$ and the remaining $(N \times (K-1))$ matrix $\mathbf{B}_1$. The factor returns are split into the first row $\mathbf{R}_{1t}$ and the remaining $(K-1)$ rows $\bar{\mathbf{R}}_{Kt}$. The constraint, then, is $\mathbf{b}_1 + \mathbf{B}_1\mathbf{1} = \mathbf{1}$. The constrained model is of the form

$$\mathbf{R}_t - \mathbf{1}\mathbf{R}_{1t} = \mathbf{B}_1(\bar{\mathbf{R}}_{Kt} - \mathbf{1}\mathbf{R}_{1t}) + \epsilon_t.$$

The resulting Maximum Likelihood estimators are

$$\begin{aligned}
\hat{\mathbf{B}}_1^* &= \left[(\mathbf{R}_t - \mathbf{1}\mathbf{R}_{1t})(\bar{\mathbf{R}}_{Kt} - \mathbf{1}\mathbf{R}_{1t})'\right] \\
&\times \left[(\bar{\mathbf{R}}_{Kt} - \mathbf{1}\mathbf{R}_{1t})(\bar{\mathbf{R}}_{Kt} - \mathbf{1}\mathbf{R}_{1t})'\right]^{-1}. \\
\hat{\mathbf{b}}^* &= \mathbf{1} - \hat{\mathbf{B}}_1^*\mathbf{1} \\
\hat{\Sigma}^* &= \frac{1}{T}\sum_{t=1}^{T}(\mathbf{R}_1 - \hat{\mathbf{B}}_{Kt}^*)(\mathbf{R}_1 - \hat{\mathbf{B}}_{Kt}^*)'
\end{aligned}$$

The test statistic J is distributed with $2N$ degrees of freedom being equal to the number of restrictions, under the null. An exact version of a test statistic is given by Campbell, Lo, and MacKinlay (1997) as

$$J_2 = \frac{T-N-K}{N}\left[\frac{|\hat{\Sigma}^*|}{|\hat{\Sigma}|} - 1\right].$$

This is central F distributed with $2N$ degrees of freedom in the numerator and $2(T-N-K)$ degrees of freedom in the denominator.

Factor selection

So far, the number of factors have been assumed knwon as well their identities. Here, neither is assumed, so that attention is given to their retrieval. For the specification of the factors, basiclally, two methods exist. The first is of statistical nature which is inherent in the APT. The second is of more theoretical nature in that the factors are selected based on their respective ability to capture the risks of markets in an economic sense.

We will begin with the first approach, the statistical one. Here, usually, te set of returns to use for the retrieval is much larger than in the estimation and testing case of the pervious paragraphs. The general model is

$$\begin{gathered}\mathbf{R}_t = \mathbf{a} + \mathbf{B}\mathbf{f}_t + \epsilon_t \\ E[\epsilon_t \epsilon_t^{'}] = \Sigma, \\ \mathbf{R}_t \in \mathbb{R}^N,\ \mathbf{f}_t \in \mathbb{R}^K,\ \epsilon_t \in \mathbb{R}^N.\end{gathered}$$

The statistical analysis itself is subdivided into two groups of which the first is the factor analysis and the second is the principal component analysis. We will focus on the first, for now.

The factor analysis is basically a two-step procedure. The first step estimates the sensitivity matrix, $\mathbf{B}$, and the $(N \times N)$disturbance covariance

matrix, Σ. The second step is intended to construct measures of factor realizations.

In the first step, a strict factor structure of, say, K factors is determined such that the $(N \times N)$ matrix Σ is diagonal since the K factors account for all the cross covariation. The covariance matrix of the N returns, in the strict factor structure case, can be decomposed in the way

$$\Omega = \mathbf{B}\Omega_K\mathbf{B}^{'} + \Sigma$$

with $E[\mathbf{f}_t\mathbf{f}_t^{'}] = \Omega_K$.[18] Here, $\mathbf{B}$ is not unique. If the factors are orthogonal to each other, $\mathbf{B}$ is unique up to rotational transformation. Because of the orthogonality of the factors,

$$\Omega = \mathbf{B}\mathbf{B}^{'} + \mathbf{D}.$$

With the joint *i.i.d.* normal assumption of the returns, $\mathbf{B}$ and $\mathbf{D}$ can be found through Maximum Likelihood estimation, though maybe, in a slow and difficult process.

In the second step, we estimate the factors. The will be assumed to have mean zero, even though, results will not change much if this assumption is dropped. Let μ denote the vector of expected values of the N returns. So, we can write

$$(\mathbf{R}_t - \mu) = \mathbf{B}\mathbf{f}_t + \epsilon_t.$$

We, now, find the factors through Generalized Least Squares (GLS) method, i.e.

$$\hat{\mathbf{f}}_t = (\hat{\mathbf{B}}^{'}\hat{\mathbf{D}}^{-1}\hat{\mathbf{B}})^{-1}\hat{\mathbf{B}}^{'}\hat{\mathbf{D}}^{-1}(\mathbf{R}_t - \mu). \tag{6.15}$$

The factors are linear combinations of the returns and, hence, there can be constructed portfolios that are perfectly correlated with the factors. From

[18] Σ is a diagonal matrix.

(6.15), we obtain for the factor return vector

$$\hat{\mathbf{R}}_{Kt} = \mathbf{AWR}_t$$

with

$$\mathbf{W} = (\hat{\mathbf{B}}'\hat{\mathbf{D}}^{-1}\hat{\mathbf{B}})^{-1}\hat{\mathbf{B}}'\hat{\mathbf{D}}^{-1}$$

and a matrix of norming entries, $\mathbf{A}$ such that the weights in the factor portfolios add to unity.

The weights are such that the factors' contributions to the residual variance are minimal, respectively. Formally,

$$\begin{aligned} &\min_{\omega_j} \omega_j^{'}\hat{\Sigma}\omega_j, \text{ s.t.} \\ &\omega_j^{'}\hat{\mathbf{b}}_k = 0, \ \forall k \neq j \\ &\omega_j^{'}\hat{\mathbf{b}}_k = 1, \ \forall k = j. \end{aligned}$$

If the population values of $\mathbf{B}$ and Σ are known, then the factor estimates will have the maximum correlation with the population fators, when returns are normal. In practice, things, however, are a little different.[19]

We will, now, turn to principal component analysis. It is a technique to reduce the number of variables under analysis. This has to be done, though, in a way that will not cause loss of too much information in the covariance matrix. Here, the reduction will be from $N \rightarrow K$. The principal components (PC) serve as factors. The first PC is a linear combination of asset returns with maximum variance. The second PC is as the first a linear combincation with maximum variance, however, of all combinations orthogonal to the first. The will be repeated until the Kth factor is found. Formally, the first PC is $x_1^*\mathbf{R}_t$ where x_1^* is the $(N \times 1)$ solution to

$$\begin{aligned} &\max_{x} x^{'}\hat{\Omega}x, \text{ s.t.} \\ &x^{'}x = 1. \end{aligned}$$

[19] Any thorough monograph on multivariate analysis will cover this.

From algebra, we know that x_1^* is the eigenvector associated with the largest eigenvalue of $\hat{\Sigma}$. The first factor, then, is $\omega_1 \mathbf{R}_t \equiv x_1^*/(\sum x_{1,i}^*)\mathbf{R}_t$. x_2^* is found analogously to x_1^* but with the orthogonality requirement $x_2^{*'} x_1^* = 0$. x_2^* is the eigenvector, then, associated with second largest eigenvalue of $\hat{\Omega}$. Rescaling of x_2^* to unity yields the weights vector of the second factor. This will be repeated up to the Kth factor.

We will discuss briefly the issue of selecting the number of factors, K. The importance is that K should be sufficiently small. Basically, two approaches exist. The first is a repetition of estimation and testing for various K. Subsequently, the estimators are tested for their, respective sensitivity to varying sizes of K. The second approach tests explicitly. This is done by an asymptotic likelihood ratio test with statistic

$$J_5 = -\left(T - 1 - \frac{1}{6}(2N + 5) - \frac{2}{3}K\right) \times [\ln|\hat{\Omega}| - \ln|\hat{\mathbf{B}}\hat{\mathbf{B}}^{'} + \hat{\mathbf{D}}|]$$

where all estimators in the above equation are obtained from Maximum Likelihood estimation. Asymptotically, this statistic is χ^2 distributed with $1/2[(N-K)^2 - N - K]$ degrees of freedom.

In contrast to the statistical approaches we just discussed, there are also theoretical methods as promised. These can be further subdivided into two groups. The first specifies factors from macroeconomic and financial quantities. They are supposed to adequately capture the systematic risks inherent in the market. Examples include yield spreads, expected as well as unexpected inflation, and industrial production growth. The second specifies characteristics of firms that assumed to be comprised in the systematic risk. The factors are portfolios of stocks based on those characteristics. Examples include market value of equity, price-to-earnigs ratio, and book-to-market value of equity ratio.

Deviations from exact factor pricing

In empirical studies of multifactor models, deviations from the exact form have been detected. Since additional factors are only of temporary usefulness to guarantee better fit, two basic sources have to be analyzed. Deviations result from either risk-based or non-risk-based misspecifications. To testfor the possible origin, the maximum squared Sharpe ratio will aid since it behaves differently in the two distinct cases. For risk-based deviations, it has an upper bound whereas this is not the case when deviations are of non-risk-based nature.

In order to illuminate the possibilities and the effect of introducing an additional factor portfolio when the existing K factor portfolios fail to explain the exact linear excess return relationship, we consider, as before, the equations system

$$\mathbf{Z}_t = \mathbf{a} + \mathbf{B}\mathbf{Z}_{Kt} + \epsilon_t.$$

Notation is as in the accustomed way from the previous sections. The covariance matrix Ω has full rank. In case of exact factor pricing, $\mathbf{a} = \mathbf{0}$. Then, a linear combination of the factor portfolios will assemble the tangency portfolio with excess returns Z_{qt} and $(N \times 1)$ weights vector

$$\omega_{\mathbf{q}} = (\mathbf{1}'\Omega^{-1}\mu)^{-1}\Omega^{-1}\mu.$$

If factor pricing is not exact, the tangency portfolio cannot be formed from the K factor portfolios. The intercept $\mathbf{a}$, now, is no longer equal to zero. In order to fill the gap, a so called *optimal orthogonal portfolio*[20] has to be found which is orthogonal to the linear subspace spanned by K factor portfolios. Formally,

Definition 6.3.1. Let the K factor portfolios fail to generate the tangency portfolio. The optimal orthogonal portfolio weights ω_h with respect to the

[20]See Campbell, Lo, and MacKinlay (1997).

K factor portfolios satisfy

$$\omega_q = \mathbf{W}_p\omega + \omega_h(1 - \mathbf{1}^{'}\omega)$$

and

$$\omega_h^{'}\Omega\mathbf{W}_p = 0$$

where ω_q are the weights of the tangency portfolio and $\mathbf{W}$ is the matrix of the weights of the K factor portfolios.

Adding this portfolio will render $\mathbf{a} = 0$ and, due to its orthogonality to the K factor portfolios, $\mathbf{B}$ will remain unchanged. Thus,

$$\begin{gathered}
\mathbf{Z}_t = \mathbf{B}\mathbf{Z}_{Kt} + \beta_h Z_{ht} + \mathbf{u}_t \\
E[\mathbf{u}_t] = 0 \\
E[\mathbf{u}_t\mathbf{u}_t^{'}] = \mathbf{\Phi} \\
E[Z_{ht}] = \mu_h,\ E[(z_{ht} - \mu_h)^2] = \sigma_h^2 \\
\mathrm{Cov}[\mathbf{Z}_{Kt}, \mathbf{u}_t] = 0 \\
\mathrm{Cov}[Z_{ht}, \mathbf{u}_t] = 0.
\end{gathered}$$

Setting

$$\mathbf{a} = \beta_h\mu_h,$$

the residual variance Σ can be written as

$$\Sigma = \beta_h\beta_h^{'}\sigma_h^2 + \mathbf{\Phi} = \mathbf{a}\mathbf{a}^{'}\frac{\sigma_h^2}{\mu_h^2} + \mathbf{\Phi}.$$

Hence, it becomes obvious that there has to be the common component in the residual variance. Otherwise, a portfolio could be formed with positive deviation and vanishing residual variance achieved through naive diversifiaction.[21]

[21]Diversification is referred to as naive when, simply, the number of securities with approximately same variances is increased.

We will turn our attention to the squared Sharpe ratio, now. We know that this measure is maximal for the tangency portfolio q. It is, then,

$$s_q^2 = \mu^{'}\Omega^{-1}\mu.$$

We also know from before that, in case the K factor portfolios should not suffice, the tangency portflio can be constructed by aid of an additional optminal orthogonal factor portfolio. Due to its orthogonality to the K factor portfolios, the tangency portfolios Sharpe ratio can be split into the two parts

$$s_q^2 = s_h^2 + s_K^2.$$

Campbell, Lo, and MacKinlay (1997) provide the relationships $s_h^2 = \mu_h^2/\sigma_h^2$ and $S_K^2 = \mu_K^{'}\Omega_K^{-1}\mu_K$. Interestingly, similar results hold when, as is the case in empirical tests, merely a subset of the N assets is analyzed. Here, to express the tangency portfolio via the maximum squared Sharpe ratio, the excess returns vector has to be composed of the subset returns and the K factor returns yielding $[\mathbf{Z}_t^{'}\mathbf{Z}_{Kt}^{'}]$ with means $\mu_s^{*'}$ and covariance matrix Ω_s^*. Hence, the squared Sharpe ratio of the subset's tangency portfolio is

$$s_{q_s}^2 = \mu_s^{*'}\Omega_s^{*-1}\mu_s^*.^{22}$$

In terms of the subset's intercept $\mathbf{a}_s$ and residual covariance matrix Σ_s, the squared Sharpe ratio of the subset's tangency portfolio satisfies the following relationship,

$$s_{q_s}^2 - s_K^2 = \mathbf{a}_s^{'}\Sigma_s^{-1}\mathbf{a}_s$$

such that it becomes evident that

$$s_{h_s}^2 = \mathbf{a}_s^{'}\Sigma_s^{-1}\mathbf{a}_s \leq s_h^2$$

and

$$s_{q_s}^2 = s_{h_s}^2 + S_K^2.$$

In the following, alternative theories explaining the deviations of exact multifactor models are discussed. Basically, as mentioned previously, two directions of reasoning exist, i.e., on the one hand, additional factor portfolios do improve the model and, on the other hand, additional factors only work in-sample since they lack any theoretical motivation. Both competing alternative hypotheses are tested against the null hypothesis of $\mathbf{a} = 0$. The generaliezed F-test statistic

$$J_1 = \frac{T - N - K}{N}[1 + \hat{\mu}_K^{'}\hat{\Omega}_K^{-1}\hat{\mu}_K]^{-1}\hat{\mathbf{a}}^{'}\hat{\Sigma}^{-1}\hat{\mathbf{a}}$$

will be applied. It is distributed $F_{N,T-N-K}(\delta)$. The non-centrality parameter is

$$\delta = T[1 + \hat{\mu}_K^{'}\hat{\Omega}_K^{-1}\hat{\mu}_K]^{-1}\hat{\mathbf{a}}^{'}\hat{\Sigma}^{-1}\hat{\mathbf{a}}.$$

Under the assumption that the deviations are risk-based, i.e. factors are missing,

$$\delta = T[1 + \hat{\mu}_K^{'}\hat{\Omega}_K^{-1}\hat{\mu}_K]^{-1}s_{h_s}^2 < Ts_h^2 \leq Ts_q^2.$$

Hence, there is an upper bound for the non-centrality parameter. Consequently, the diffrenece in distributions under the null hypothesis and the alternative hypothesis, respectively, is bounded, no matter if all asset prices are mispecified. However, when the assumption excludes a missing factor, i.e. it is non-risk-based, the squared Sharpe ratio becomes meaningless and posssibly may rise beyond all bounds. Thus, everything left equal, when comparing the two competing alternatives, the test statistic should be larger in the non-risk-based case since no boundaries exist.

We will consider, here, an example of real data presented in Campbell, Lo, and MacKinlay (1997) to demonstrate the behavior of δ under the two competing alternative hypotheses. The model is given to be a one-factor asset pricing model using excess returns. The time series consists of 342 monthly observations. The dependent variables are 32 portfolios whereas the one independent factor is the market. Hence, if the null hypothesis holds, $\mathbf{a} = 0$ and we obtain the CAPM. Under the null, the test statistic J_1 is distributed F with 32 degrees of freeedom in the numerator and 309 degrees of freedom in the denominator.

Now, we first consider the risk-based case. Of interest is the distribution of the upper bound for the noncentrality parameter. From tests, we obtian estimates for s_q^2 of $\approx .031$ per month and s_h^2 of $\approx .021$ per month. We, then, receive for the distribution of J_1 a noncentral $F_{32,309}(7.1)$ distribution. Alternatively in the non-risk based case, we assume that the components of the intercept $\mathbf{a}$ are normally distributed with zero mean and standard deviation of, first, $\sigma_a^{(1)} = .0007$ and, second, of $\sigma_a^{(2)} = .001$. The parameter Σ is obtained from portfolios sorted by market capitalization during the sample period between 1963-91. For the first case, we have an expected value of for δ of 39.4. In the second instance, we have 80.3. The noncentral F distributions change accordingly. The four cases just discussed are illustrated in Figure The plot reveals that the distribution under the null hypothesis and the risk-based case almost coincide. The two non-risk-based alternative distributions are further to the right. This a result of the boundedness of the noncentrality parameter in the risk-based case whereas the mode of the distribution increases along with increasing standard deviations for the two non-risk-based alternatives. A test statistic of 1.91 found in literature suggests that the risk-based alternative fails, as can be seen. This value seems to correspond more to the non-risk-based alternative with lower standard deviation. However, one and the same parametrization of the non-risk-based

distribution can be obtained from different reasonings. Thus, there is no uniqueness. It is left to be said that, unfortunately, misspecification of multifactor models is not easily solved and requires more thorough analysis.

6.4 Different alterations of the CAPM

Conditional CAPM

Still, as in the historical CAPM, we assume that investors have common subjective expectations with respect to returns and covariances. However, along with time, these expectations vary so that the moments turn into random variables, themselves. Some consistency remains, though. Based on time $t-1$, all market participants share identical expectations, again, for the ensuing time t. This calls for the need to incorporate an information process or filtration into the model yielding the Conditional CAPM or CCAPM.

Let $\mathcal{F}_t$ denote the filtration so that

$$E[R_{it}|\mathcal{F}_{t-1}] = E[R_{ft}|\mathcal{F}_{t-1}] + \beta_{imt}(E[R_{mt}|\mathcal{F}_{t-1}] - E[R_{ft}|\mathcal{F}_{t-1}])$$

where i denotes the asset, m denotes the market portfolio, and the riskfree asset is indicated by f. The beta parameter is conditional on time, as well, hence,

$$\beta_{imt} = \frac{\text{Cov}[R_{it}, R_{mt}|\mathcal{F}_{t-1}]}{\text{Var}[R_{mt}|\mathcal{F}_{t-1}]}.$$

Since, at time $t-1$, the riskfree return is known for the ensuing period, we might as well drop the condition.

It has been shown that if the model cannot be rejected when it is conditioned on a subset I_t of the filtration process, then, it also must hold for the entire $\mathcal{F}$.[23]

[23] See Javed (2000).

In testing, observing the return of the market portfolio is circumvented by assuming the market price of risk constant,

$$\lambda = \frac{(E[R_{mt}|I_{t-1}] - R_{ft})}{\text{Var}[R_{mt}|I_{t-1}]}.$$

In turn,

$$R_{mt} = R_{ft} + \lambda \text{Var}[R_{mt}|I_t] + u_{mt}$$

with

$$E[R_{mt}|I_t] = R_{ft} + \lambda \text{Var}[R_{mt}|I_t]$$

finite conditional variance of the innovation u_{mt}. For the individual asset i, we have

$$R_{it} = R_{ft} + \lambda \text{Cov}[R_{it}, R_{mt}|I_t] + u_{it}.$$

Both, u_{mt} and u_{it} are orthogonal to the information set I_{t-1}. Hence, the relationship between asset i and the market portfolio m in terms of their common covariance can be expressed as the covariance of their respective innovations. That is,

$$\text{Cov}[R_{mt}, R_{it}|I_t] = E[u_{mt} \cdot u_{it}|I_t].$$

The entire cross-sectional return relationship between assets i and the market portfolio can be formulated in terms of the innovations such that

$$R_{it} = R_{ft} + \lambda \text{Cov}[u_{it}, u_{mt}|I_t] + u_{it}.$$

Conditional variance processes

A particular version of the forgone paragraphs is presented, here. As mentioned in previous chapters, changing moments, particularly the variance,

can be well described as clustering. Moments of the returns incoporate past shocks in a linear or non-linear way. The latter, as we well know, is the case with the ARCH processes. In this context, let the excess asset returns be written as

$$\begin{aligned} Z_{imt} &= a + \beta Z_{mt} + \epsilon_t \\ \epsilon_t | \mathcal{F}_t &\sim N(0, \sigma_t^2) \\ \sigma_t^2 &= \omega + \phi_1 \epsilon_{t-1}^2 + \ldots + \phi_p \epsilon_{t-p} + \theta_1 \sigma_{t-1}^2 + \ldots + \theta_q \epsilon_{t-q}. \end{aligned}$$

However, the first moment, the mean, can also be conditional on past innovations. This is commonly encountered by adding a function f of the conditional variance such that

$$Z_{imt} = a + \beta_{im} Z_{mt} + f(\sigma_t^2) + \epsilon_t$$

with everything else equal. The function f can be interpreted as extra compensation for higher. In case, the variance increases, the function which may monotone in σ_t^2 increases such that the investor is compensated with a premium for the additional risk.

CAPM with higher order co-moments

As mentioned before and as is widespread understanding, asset returns are mostly not normally distributed. Also, quadratic utility functions are found to be of questionable usefulness in reality. These concerns are of essential importance with respect to basic prerequisites of the CAPM. Particularly the first issue concerning the return distribution has been given a lot of attention. Hence, it is only natural that effort has been taken to incorporate findings of non-normality into the factor models. Since some of the indicator statistics for rejection of the normal distribution are the skewness and kurtosis, it is intuitive to use these statistics to explain deviations of the cross-sectional

expected returns from theory. The problem with skewness and kurtosis is that these two statistics cannot be diversified away by an increasing number of assets.

Introduction of higher-order co-moments can be found in some models, as well as conditional skewness and systematic co-skewness. It was detected that in many cases, the use of skewness could well capture the asymmetry in the returns. After correcting for it, the CAPM performed better. Concludingly, it can be stated that as long as it is not guaranteed that the returns are normal, testing for the traditional single-factor as well as multifactor versions of the CAPM does not make sense.

For an account of several alternatives, the reference list will give initial information.

Bibliography

[1] Campbell, J.Y., Lo, W., and Mac Kinlac, A.C. (1997), *The Econometrics of Financial Markets*, Princeton University Press, Princeton.

[2] Barnett, W.A. and Serletis, A. (2000), *Martingales, nonlinearity, and chaos*, Journal of Economic Dynamics & Control, 24, 703-24.

[3] Bollerslev, T. (1986), *Generalized Autoregressive Conditinal Heteroscedasticity*, Journal of Econometrics, 31, 307-27.

[4] Bos, T. and Newbold, P. (1984), *An empirical investigation of the possibility of stochastic systematic risk in the market model*, Journal of Business, 57, 35-41.

[5] Chan,L.K.C. and Lakonishok,J. (1993), *Are the reports of beta's death premature?*, Journal of Portfolio Management, 19, 51-62.

[6] Chung, Y.P., Johnson, H. and Schill, M.J. (2001), *Asset pricing when returns are nonnormal: Fama-French factors vs higher-order systematic co-moments*, Working Paper, A. Gary Anderson Graduate School of Management, University of California, Riverside.

[7] Davis, J. (1994), *The cross-section of realised stock returns: the pre-COMPUSTAT evidence*, Journal of Finance, 49, 1579-1593.

[8] Diebold, F.X. (2004), *The Nobel Memorial Prize for Robert Engle*, The Rodney L. White Center for Financial Research, Wharton School.

[9] Engle, R.F. (2001), *GARCH 101: The Use of ARCH/GARCH Models in Applied Econometrics*, Journal of Economic Perspectives, 15, Fall 2001, 157-68.

[10] Engle, R.F. (1982), *Autoregressive Conditional Heteroscedasticity with Estimates of the Variance of U.K. Inflation*, Econometrica, 50, 987-1008.

[11] Faff, R. and Brooks, R.D. (1998), *Time-varying beta risk for Australian industry portfolios: and exploratory analysis*, Journal of Business Finance and Accounting, 25, 721-745.

[12] Faff, R., Ho, Y.K. and Zhang, L. (1998), *A generalised method of moments test of the three-moment capital asset pricing model in the Australian equity market,* Asia Pacific Journal of Finance, 1, 45-60.

[13] Faff, R., Lee, J. and Fry, T. (1992), *Time stationarity of systematic risk: some Australian evidence,* Journal of Business Finance and Accounting, 19, 253-70.

[14] Fama, E. and French, K. (1995), *Size and book-to-market factors in earnings and returns,* Journal of Finance, 50, 131-156

[15] Ferson, W.E. and Harvey, C.R. (1999), *Conditioning variables and the cross-section of stock returns,* Journal of Finance, 54, 1325-1360.

[16] He, J. and Ng, L.K. (1994), *Economic forces, fundamental variables and equity returns,* Journal of Business, 67, 5999-639

[17] Jagannathan, R. and Wang, Z. (1996), *The conditional CAPM and teh cross-section of expected returns,* Journal of Finance, 51, 3-53

[18] Javed, A.Y. (2000), *Alternative Capital Asset Pricing Models: A Review of Theory and Evidence*, PIDE.

[19] Jones, P. (1991), 3rd edition, John Wiley and Sons: New York, *Investments Analysis and Management.*

[20] Kandel, S. and Stambaugh, R.F. (1995), *Portfolio inefficiency and the cross-section of expected returns,* Journal of finance, 50, 157-184.

[21] Kothari, S.P., Shanken, J. and Sloan, R.G. (1995), *Another look at the cross-section of expected stock returns,* Journal of Finance, 50, 185-224.

[22] Kraus, A. and Litzenberger, R.H. (1976), *Skewness preference and the valuation of risk assets,* Journal of Finance, 31,1085-1100.

[23] Loudon, G.F., Watt, W.H., and Yadav, P.K. (2000), *An Empirical Analysis of Alternative Parametric ARCH Models*, Journal of Applied Econometrics, 15, 117-36.

[24] Miles, D. and Timmermann, A. (1996), *Variation in expected stock returns: evidence on the pricing of equities from a cross-section of UK companies,* Economica, 63, 369-382

[25] Mills, T.C. (1997), *The Econometrics of Financial Time Series*, Cambridge University Press.

[26] Sears, R.S. and Wei, K.C.J. (1985), *Asset pricing, higher moments, and the market risk premiums: a note,* Journal of Finance, 40, 1251-1253

Appendix A

Notation

A.1 Brownian Motion

Let $W(t)$ and $V(t)$ denote two independent standard Brownian motions on $C[0,1]$, such that their respective increments $(W_1 - W_0, \ldots, W_t - W_{t-1})$ and $(V_1 - V_0, \ldots, V_t - V_{t-1})$ are multivariate normally distributed with zero means and covariance, and variances equal to $t_i - t_{i-1}$. The differences are written as $w_j \equiv W_j - W_{j-1}$ and $v_j \equiv V_j - V_{j-1}$, respectively. Additionally, we denote

$$Q_t = \sum_{j=1}^{t} w_j P_t = \sum_{j=1}^{t} v_j.$$

We obtain the following covariance matrix for P_T and Q_T as $T \to \infty$

$$\Sigma = \begin{bmatrix} \sigma_v^2 & \sigma_{vw} \\ \sigma_{vw} & \sigma_w^2 \end{bmatrix}$$

where

$$\sigma_{vw} \equiv \lim_{T \to \infty} \frac{1}{T} E P_T Q_T.$$

A.2 Asymptotic behavior

[1] With w_t and v_t from A.1 we define

$$\xi_t \equiv (v_t, w_t)'$$
$$S_t \equiv \sum_{j=1}^{t} \xi_j, \ S_0 \equiv 0$$

Let the estimated linear regression be

$$y_t = \hat{\alpha} + \hat{\beta} x_t + u_t.$$

Conditions on $\{\xi_t\}_1^\infty$ for permitting y_t and x_t to be correlated $I(1)$ processes with differences that are "weakly dependent"[2] and permit heterogeneous innovations

(a) $E\xi_t = 0, \ \forall t$

(b) $\sup_{i,t} E|\xi_{it}|^{\beta+\epsilon} < \infty, \ \beta > 2, \ \epsilon > 0, \ \xi_{1t} = v_t, \ \xi_{2t} = w_t$

(c) $\exists \ \Sigma \equiv \lim_{T\to\infty} \frac{1}{T} E S_T S_T\prime$ (pos. definite).

(d) $\{\xi_t\}_1^\infty$ is *strong mixing*. That means that temporal dependence as well as heterogeneity is permitted to some extend.

Further, define dw to be the Durbin-Watson statistic. Let

$$\mu_{vw} \equiv \int_0^1 v(r)w(r)dr - \int_0^1 v(r)dr \int_0^1 w(r)dr.$$

With V and W from (A.1), we set

$$\phi \equiv \int_0^1 V(r)dr - \frac{\mu_{vw}}{\mu_{ww}} \int_0^1 W(r)dr$$

[1] For further interpretation of the respective statements, consult Mills (1997).
[2] See Mills (1997).

Asymptotic properties are given for the estimated regression parameters and corresponding statistics by the list below.

(i) $\hat{\beta} \Rightarrow \frac{\sigma_v}{\sigma_w} \frac{\mu_{vw}}{\mu_{ww}}$

(ii) $\frac{1}{\sqrt{T}}\hat{\alpha} \Rightarrow \sigma_v \phi$

(iii) $\frac{1}{t_\beta} \Rightarrow \frac{\mu_{vw}}{\sqrt{v}}$

(iv) $\frac{1}{t_\alpha} \Rightarrow \phi \frac{\mu_{ww}}{\sqrt{v \int_0^1 W(r)^2 dr}}$

(v) $R^2 \Rightarrow \frac{m u_{vw}^2}{\mu_{vv}\mu_{ww}}$

(vi) $dw \xrightarrow{P} 0.$