Estimation of Nonlinear Stochastic Processes

Jon Kierkegaard

IMM

Abstract

This thesis concerns estimation of stochastic volatility models in finance, discrete time models as well as continuous time models.

Returns of financial time series are investigated for characteristics that may be used for volatility modelling. The well-known GARCH models as well as some related models are investigated, and their abilities to capture the characteristics of financial returns are evaluated in a large estimation study where the traditional Maximum Likelihood estimation method as well as an alternative recursive estimation method are used.

Parameters of bivariate continuous time stochastic volatility models are traditionally very difficult to estimate. This thesis investigates an indirect estimation procedure, were the parameters of continuous time models are obtained from discrete time model parameter estimates. A new discrete time stochastic volatility model, SSSH-GARCH, is introduced.

Keywords

Preface

This thesis on estimation of discrete and continuous time stochastic volatility models constitutes the last part of the requirements for the Master's Degree in Engineering at The Technical University of Denmark.

During the course of the writing of this thesis I have been helped and assisted by a number of people whom I wish to thank. In particular, I would like to thank my supervisors Professor Henrik Madsen and Ph.D. student Jan Nygaard Nielsen for their help and guidance. Without their insights and suggestions, the quality of this thesis would be on quite another, lower, level. I would also like to thank Associate Professor Poul Thyregod for a lot of help with statistical problems and Lars Jensen for inspiring discussions.

Lyngby, Denmark, April the 3rd, 2000

Jon Lee Kierkegaard

5

Contents

Chapter 1

Introduction

1.1 Introduction

The concept of volatility is probably the single most investigated phenomenon in modern day mathematical finance. The vast interest in volatility is motivated by two related trends. The growing number of companies utilizing risk management (one) and the huge amount of derivatives traded on todays exchange markets (two).

The fair price of an option is determined by a number of factors including the volatility of the underlying asset. All these factors except the volatility are directly observable from the market or specified by the option contract.

When a company through risk management wants to establish its exposure to financial risk, it has to know the volatility of each of the assets it possess. The volatility is generally perceived to be *the* measure of risk.

Motivated by these facts, it is the purpose of this thesis to investigate a number of models which can be used to quantify and forecast the volatility of stocks. The focus of the thesis is on the particular volatility models known as ARCH and GARCH models, originally proposed by (Engle 1982) and (Bollerslev 1986), respectively. A large part of the thesis is devoted to the estimation of the parameters of these models using historical data.

For quite some time, research in mathematical finance has been divided by a gap between discrete and continuous time. The vast majority of theories of mathematical finance has been formulated in continuous time, mainly because continuous

time is much easier to deal with theoretically than discrete time, but also, because
real life finance evolves in continuous time. In contrast, all data series appear in
discrete time, a fact that makes discrete time models the most adequate in modeling
real life data and parameters of continuous time models difficult to estimate.

Following the work of (Duan 1997) the thesis investigates an approach where pa-
rameters of continuous time stochastic volatility models are estimated using pa-
rameter estimates of discrete time models.

Throughout the thesis the estimation methods introduced and the properties de-
scribed are illustrated using real life data as well as simulated data. Four series of
stock closing prices have been selected as representatives of real life stocks.

The thesis may be seen as a continuation of the GARCH investigations initiated by
(Bisgaard 1998). It is, however, my hope that this thesis may stand alone as a study
of the estimation of parameters of discrete and continuous time stochastic volatility
models.

1.2 Scope

Financial time series comes in many forms and from different sources. Daily
quotes of the price of Hewlett-Packard stocks, annual observations of the US 30-
year treasure bond rate and semi-annual quotes of the price of gold are all examples
of financial time series.

Because the term "financial time series" encompasses time series from a vast num-
ber of different sources, financial time series have only very few common char-
acteristics. Because of this, no usable model fits every financial time series and
thus, only a subclass of financial time series is considered here. Stocks have been
chosen as stocks are known to possess a number of common characteristics, stocks
are volatile and options on stocks are common. Furthermore, this thesis considers
ARCH and GARCH models, and these models are traditionally applied to stocks
prices, though they have also proven to be applicable to other financial time series.
For a discussion of ARCH model applicability, see (Bera & Higgins 1993).

For practical purposes, four stock price series have been selected for study. These
are daily closing prices of the stocks of the Hewlett-Packard, Sony, Mobil and
Pepsi companies from October 20, 1992 through October 20, 1997, a total num-
ber of 1265 observations per share. Four series is of course too small a number
to constitute a representative selection of stock series. This means that any con-

clusion obtained using the series should not be generalized unless the generality is accounted for specifically.

Though stocks are the focus of attention, some of the characteristics described and some of results obtained apply to other types of financial time series. Typically, properties of stocks are also retained by stock indices and exchange rates but readers merely interested in interest rates or commodities will not find much of interest in this report.

Though volatility models are traditionally used for risk management or option pricing, none of these applications will be investigated here. The focus of this thesis is on estimation of parameters of stochastic volatility models, not on applied option pricing and risk management.

1.3   Notation

Throughout the thesis I have attempted to keep the number of abbreviations at a minimum, an in the few places where abbreviations have been used, they are properly introduced. Thus, there should be no need to provide a list of abbreviations here.

The mathematical notation used is in line with the notation used in mathematical finance. Vectors are column vectors unless otherwise stated, $X^\top$ is the row vector whose elements are identical to those of the column vector X.

1.4   Software

A large part of this thesis is devoted to investigations of the properties of various methods used to estimate parameters of stochastic volatility models like ARCH and GARCH. When performing these investigations, I have been faced with the choice whether to write the necessary computer programs my self or use standard software like SAS or S-Plus that offers packages for estimation of GARCH parameters.

Because the standard software packages hide a lot of the elements of estimation of stochastic volatility models to the user and because I have wanted to investigate the influence of these elements on the obtained parameter estimates, I have chosen to write the software myself as Matlab M-files. This means that I have been in full control of the estimation procedures, but also that I have spent a lot of time writing

the computer code. Appendix A provides a list of all the computer programs I have written as part of my work on this thesis.

1.5   Outline of the thesis

The thesis is structured in the following way:

Chapter 2 reviews the statistical properties of stock price series and return series. The statistical properties investigated includes the mean, variance and correlation structure. A basic model for describing financial return series is discussed, and the concept of volatility is formally defined.

Chapter 3 introduces the discrete time ARCH and GARCH models as well as models derived from these ones. The Augmented GARCH process is presented as a model class that encompasses the standard ARCH and GARCH models as well as derived models.

Chapter 4 is devoted to the estimation of ARCH and GARCH model parameters. The traditional estimation methods such as Maximum Likelihood and Quasi Maximum Likelihood are evaluated, and the traditional estimation problems are emphasized. An alternative recursive estimation method is proposed and evaluated in comparison with the traditional methods.

In Chapter 5 the continuous time Augmented GARCH process is derived from the discrete time process. Special attention is put on the transition from discrete to continuous time and two well-known continuous time stochastic volatility models are estimated using parameter estimates of discrete time models.

Chapter 6 contains a summary of the results obtained in the thesis, proposes areas for future research and concludes the thesis.

Chapter 2

Financial Time Series and Volatility

The purpose of this chapter is to look at the characteristics of financial time series relevant to volatility modeling and to introduce and define a number of concepts essential to the subject of the thesis.

Section 2.1 deals with missing observations. Section 2.2 introduces returns and return series and look at some common characteristics of returns. Section 2.3 formally defines the concept of volatility.

2.1 Financial Time Series

2.1.1 Daily quotes

As mentioned earlier, four series of daily stock closing prices are used for illustration purposes. It should be noted that these "daily" closing prices are not literally daily. Stock markets are closed on weekends and on holidays meaning that there will be non trading days, i.e. days without official closing prices.

In this thesis non trading days will be disregarded in the sense that each of the four series will be treated as series of equidistant observations. This must be kept in mind when obtaining statistical conclusions using the four series.

17

The non trading days have been disregarded as is not obvious how to treat these "missing" observations. Studies (Burghardt & Hanweck Jr. 1993) indicate that during weekends and holidays the volatility is different from the volatility of other days, but it has not been possible to quantify this volatility in a way that allows it to be incorporated into the generally used volatility models.

2.2   Return series

In general it is not very interesting to look at just the price of an investment, a stock in this case. Seen from an investors point of view, the return of the investment is much more interesting. Mainly because the investor has the relative gain of his investment in focus, not the nominal price of the investment, but also because the return interpreted as the relative price change allow for comparisons across companies, stock markets and currencies[1]. In addition, the returns are stationary, a property not possessed by the actual stock prices. Thus, for the rest of the thesis, return series are the objects of consideration.

In the world of finance, the concept of return is not unambigiously defined. Let the stock price (of some stock) at time t be denoted by S_t. The return at time t may be defined by

DEFINITION 2.1 (THE ARITHMETIC RATE OF RETURN).

$$R_{1,t} = \frac{S_t \Leftrightarrow S_{t-1}}{S_{t-1}} \tag{2.2.1}$$

▲

Another definition is

DEFINITION 2.2 (GEOMETRIC RATE OF RETURN).

$$R_{2,t} = \ln \frac{S_t}{S_{t-1}}. \tag{2.2.2}$$

▲

---

[1]Although, when dealing with different currencies, the exchange rate has to be taken into account.

The geometric rate of return is also known as the *Compound Return*.

Throughout the thesis, the geometric return will be used. It will be denoted by R_t and it will be referred to as the *return*. The geometric return is chosen because it is the most widely used, thus allowing for easy benchmarking of obtained results. Also, as will become clear in a later chapter, the geometric rate of return makes it possible to relate discrete and continuous time models.

In Figure 2.1 the geometric returns of Hewlett-Packard, Sony, Mobil and Pepsi are plotted.
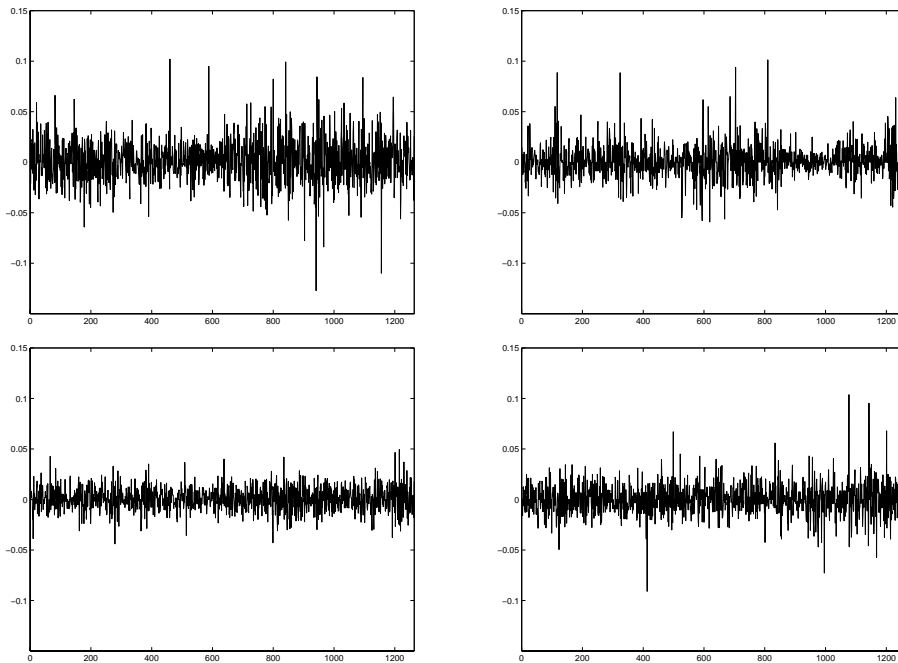
Figure 2.1: Returns of Hewlett-Packard (upper left), Sony (upper right), Mobil (lower left) and Pepsi (lower right).

2.2.1   Statistical properties of return series

By looking at various statistical properties of return series such as summary statistics, autocorrelations and extreme observations (outliers) it often possible to have a reasonable guess at the process driving the series.

In the early days of mathematical finance, most theoretical models of return series assumed that the returns could be considered as Gaussian White Noise[2].

Today, it is generally known that this assumption is not correct. Return series are known to possess a number of characteristics different from those of Gaussian White Noise. The most prominent of these are

1. **Excess kurtosis**. Return series have excess kurtosis compared to Gaussian White Noise. That is, return series are heavy-tailed.

2. **Heteroskedasticity**. The variance of the returns changes over time.

3. **Autocorrelation**. The returns exhibit only insignificant autocorrelation. However, the *squared* returns are autocorrelated at a significant level.

These characteristics are generally known as the *Stylized Facts*. Stylized fact number 1 and 2 were described as early as 1963 by (Mandelbrot 1963)[3], stylized fact number 3 by (Engle 1982) if not before. A more in-depth description of fact number 3 is provided by (Taylor 1994), claiming that the autocorrelations of the squared returns are positive, statistically significant, generally smaller as time-lag increases and noticably larger than the respective autocorrelations of the ordinary returns.

Researchers in the field of mathematical finance mention a number of additional characteristics. One of these is the so-called *leverage*[4] *effect* claiming that the future volatility is negatively correlated with the current return, an effect expressed by a tendency of the volatility to fall in response to an increase and rise in response to a descrease of the stock price with the magnitude of the change in volatility depending on the size *and* the sign of the stock price change. The leverage effect was described as early as (Black 1976). A lot of researchers claim that return series exhibit a positive mean though insignificant and some evidence of skewness (nonsymmetry) in the unconditional distribution of returns have been reported.

The deviations from Gaussian White Noise are interesting, not at least if, informally, Gaussian White Noise is regarded as data containing no information, that is, noise. In this view, any deviation from Gaussian White Noise is an indication of "presence" of information, information that may be used for modelling.

---

[2]Recall (Madsen 1995, Chapter 4) that a process $\{\varepsilon_t\}$ is Gaussian White Noise if $\{\varepsilon_t\}$ is a sequence of uncorrelated Gaussian distributed variables with zero mean and constant variance.

[3]In his paper, Mandelbrot do not use the term *heteroskedasticity*. He just mentions that large returns have a tendency to be followed by large returns (of either sign) and small returns have a tendency to be followed by small returns (of either sign), a phenomenon known under the term *volatility persistence*.

[4]Leverage (here): debt/equity ratio of a company.

Stylized fact number 3 is of particular interest when it comes to volatility modelling. According to this fact the past variance can be used to predict the future variance. The fact also helps explaining why the focus is on the volatility, not on the mean. The past does not reveal information about the future mean.

The tendency of the conditional variance of the returns to be non-constant has together with the positive autocorrelation of the squared returns motivated the introduction of the property *volatility persistence*. The volatility persistence may be heuristically defined as the length of time a shock persists in the conditional variance of subsequent returns. If the volatility clusters involve many returns, the volatility persistence is said to be high. If the volatility clusters tend to be of smaller size, the volatility persistence is low.

2.2.2   Exemplifying the statistical properties

The return series characteristics may be verified using visual as well as formal statistical methods.

2.2.2.1   Moments

In Figure 2.2 the returns of Hewlett-Packard is displayed next to Gaussian White Noise. The Gaussian White Noise series has by construction the same variance as the return series.
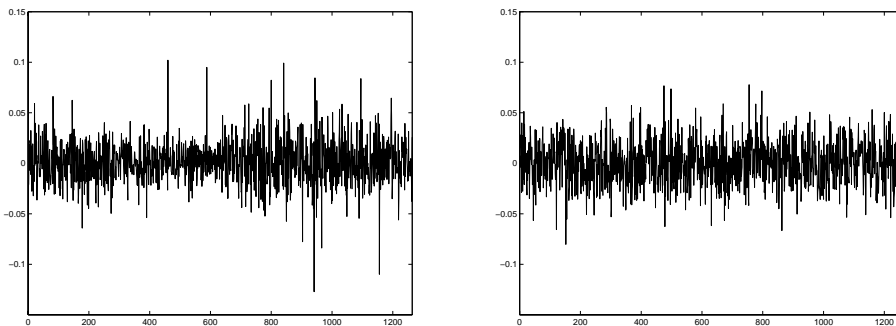
Figure 2.2: Returns of Hewlett-Packard (left) and Gaussian White Noise (right).

From the two plots it is obvious that Gaussian White Noise is not a suitable de-

scription of the return series. The return series has a substantial number of outliers compared to the Gaussian White Noise series, and in addition one may notice the non-constant variance. For instance, the standard deviation of the subseries consisting of the Hewlett-Packard returns number 500 through 550 is considerably less than the standard deviation of the whole series. Similarly, the subseries consisting of returns number 950 through 1100 has a relatively large standard deviation.

The mean of the return series appear to be close to zero.

From Figure 2.2 it appears like the return series of Hewlett-Packard contains a great number of extreme observations, outliers, compared to Gaussian White Noise, and thus it seems like the unconditional distribution of the Hewlett-Packard returns has fat tails compared to the normal distribution. Because of this, it may be illustrative to compute the sample values of the higher order moments of the return series distribution.

NOTE 1 (SAMPLE MOMENTS).
Given T observations u_t, $t = 1, 2, \ldots, T$ of a stochastic variable U, the sample skewness is computed by

$$\frac{\mathrm{E}\left[(U - \mathrm{E}(U))^3\right]}{\mathrm{E}\left[(U - \mathrm{E}(U))^2\right]^{3/2}} = \frac{\frac{1}{T}\sum_{t=1}^{T}(u_t - \hat{\mu})^3}{\left(\frac{1}{T}\sum_{t=1}^{T}(u_t - \hat{\mu})^2\right)^{3/2}} \tag{2.2.3}$$

and the sample kurtosis by

$$\frac{\mathrm{E}\left[(U - \mathrm{E}(U))^4\right]}{\mathrm{E}\left[(U - \mathrm{E}(U))^2\right]^2} = \frac{\frac{1}{T}\sum_{t=1}^{T}(u_t - \hat{\mu})^4}{\left(\frac{1}{T}\sum_{t=1}T(u_t - \hat{\mu})^2\right)^2} \tag{2.2.4}$$

where

$$\hat{\mu} = \frac{1}{T}\sum_{t=1}^{T} u_t. \tag{2.2.5}$$

The skewness is the third central moment divided by the standard deviation raised to the power of three. The kurtosis is the fourth central moment divided by the standard deviation raised to the power of four.

Computing the sample mean, standard deviation, skewness and kurtosis of our four
return series yields the values displayed in Table 2.1.

Series	HP	Sony	Mobil	Pepsi
Mean $(\times 10^4)$	0.122	8.168	6.900	5.861
Standard deviation $(\times 10^2)$	2.112	1.606	1.147	1.529
Skewness	0.048	0.664	0.032	0.218
Kurtosis	6.089	7.725	4.040	7.601

Table 2.1: The computed values of the sample mean, variance, skewness
and kurtosis of the four return series.

The skewness and kurtosis of Gaussian White Noise is 0 and 3, respectively. The
table supports the assumption that distributions of return series exhibit excess kur-
tosis and thus that they have fatter tails than the normal distribution. The skewness
estimates conform to an assumption of non-symmetry, but an investigation of the
relevant literature reveals that the estimates obtained of Sony and Pepsi are unusu-
ally high, see (Bisgaard 1998), (Blair 1995, chapter 2) or (Engle & Ng 1993). It is
also just a coincidence that all the skewness estimates are positive.

The findings of Table 2.1 are perhaps best illustrated by visualizing the empir-
ical distributions of the return series. This has been done in Figure 2.3 where
also curves of normal distributions with means and standard deviations identical to
those of Table 2.1 are shown.

Looking at just the four series it may be tempting to conclude that the means of
stock return series are different from zero, but as suggested by various researchers,
the mean may not be *significantly* different from zero. Assessing the significance
is difficult as is it not at this point reasonable to assume that the returns follows a
particular distribution. Furthermore, the skewness estimates indicate that the return
series distributions may not be symmetric thus prohibiting the use of the Wilcoxon
signed rank statistic (Conradsen 1995) as would otherwise be the natural choice.
Instead the less powerful sign statistic may be computed to test the hypothesis of
zero mean.

NOTE 2 (THE SIGN TEST).
The binomial sign test (Conradsen 1995, Chapter 3) may be used to test the hy-
pothesis that the mean μ of a stochastic variable is identical to a particular value
μ_0.

The sign test utilizes the fact that if observations equal to μ_0 are disregarded, the
number of observations less that μ_0 is binomial distributed, and the probability that
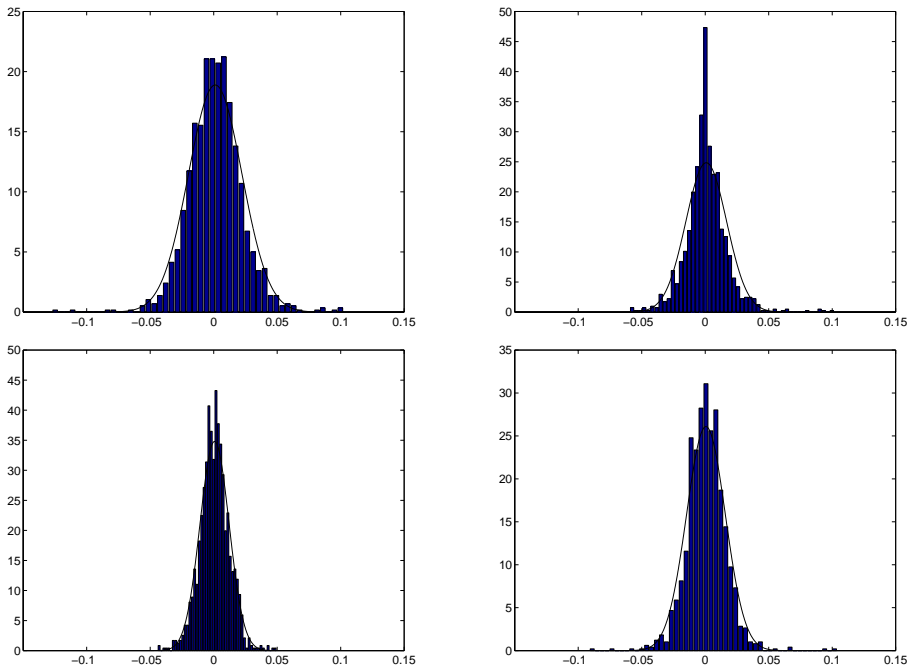
Figure 2.3: Empirical distributions of Hewlett-Packard (upper left), Sony (upper right), Mobil (lower left) and Pepsi (lower right). Notice the "outliers" corresponding to the heavy tails.

a randomly chosen observation is less than μ_0 is $\frac{1}{2}$. That is, the test statistic is

$$M = x \Leftrightarrow n/2 \qquad (2.2.6)$$

where x is the number of observations greater than μ_0 and n is the number of observations equal to μ_0. The probability of a greater absolute value of M under the null hypothesis is

$$2|M| \sum_{j=0}^{\min(x,n-x)} \binom{n}{j} 0.5^n \qquad (2.2.7)$$

The test results are displayed in Table 2.2.

Series	HP	Sony	Mobil	Pepsi
Test statistic	$18\frac{1}{2}$	17	$32\frac{1}{2}$	8
Probability	0.304	0.3224	0.0643	0.131

Table 2.2: The test statistics of the sign test.

As seen from Table 2.2 the hypothesis of zero mean is accepted for all series at all levels below 6.4%. Thus it is assumed that the return series exhibit no trend.

The existence or non-existence of heavy tails may also be assessed formally using statistical methods. The Jarque-Bera test uses sample values of the skewness and the kurtosis to test for Gaussianity.

NOTE 3 (JARQUE-BERA).
The Jarque-Bera test (Jarque & Bera 1980) uses sample estimates of skewness and kurtosis to test for Gaussianity.

The test statistic for a series with T observations u_t is

$$T_T = \left(\frac{T}{6}b_1^2 + \frac{T}{24}(b_2 \Leftrightarrow 3)^2 \right) \tag{2.2.8}$$

where b_1 and b_2 are the sample skewness and kurtosis computed by (2.2.3) and (2.2.4), respectively.

T_T is approximative $\chi^2(2)$-distributed under the (null) hypothesis of the observations u_t originating from a Gaussian distribution.

The test statistics obtained when applying the Jarque-Bera test to the four return series are displayed in Table 2.3.

Series	HP	Sony	Mobil	Pepsi
Test statistic	5.029×10^2	1.269×10^3	5.714×10^1	1.125×10^3

Table 2.3: The Jarque-Bera test statistic.

Under the null the distribution of the test statistic is $\chi^2(2)$. For comparison the $\chi^2(2)_{99\%}$ fractile is 5.99. As seen from Table 2.3 the hypothesis of Gaussianity is strongly rejected in all four cases.

2.2.2.2 Heteroskedasticity

When looking for heteroskedasticity[5] it is illustrative to divide the return series
into subseries and compare the standard deviation of the different subseries to one
another. Table 2.4 shows the sample standard deviation of each of 32 subseries.

Series	HP	Sony	Mobil	Pepsi
1 - 158	1.997	1.660	1.056	1.383
159 - 316	1.840	1.409	1.180	1.297
317 - 474	1.720	1.618	1.019	1.453
475 - 632	1.497	1.730	0.925	1.383
633 - 790	2.250	1.919	1.058	1.343
791 - 948	2.772	1.440	1.397	1.405
949 - 1106	2.418	1.074	1.055	1.899
1107 - 1264	2.157	1.837	1.409	1.924

Table 2.4: Standard deviation $(\times 100)$ for 8 subseries per series. The numbers in the first column indicate which returns are included in the subseries.

Looking at the table it seems evident that none of the series has a constant variance,
all the return series exhibit heteroskedasticity over the sample period. The Hewlett-
Packard series is by far the most illustrative at this point. The subseries consisting
of returns number 791 through 948 has a standard deviation almost twice the size
of the standard deviation of the subseries 475 through 632.

It is difficult to formally test for heteroskedasticity when no distribution of the
returns is assumed. Had the returns followed a normal distribution, Bartlett's test
had been a possibility, but with the results of Table 2.3 in mind, Gaussianity is a
highly incorrect assumption. Furthermore, test results obtained using Barlett's test
would be difficult to interpret in cases where the hypothesis of homoskedasticity is
accepted. One would never know if a different choice of subclasses would lead to
a rejection of the hypothesis.

Because of the above mentioned difficulties, a formal test for heteroskedasticity is
deferred to a later chapter.

---

[5]In the literature the term *heteroskedasticity* seems to be ambigiously defined. In this thesis, a
time series is said to be heteroskedastic if it exhibits changing conditional variance over time.

2.2.2.3 Autocorrelation

Autocorrelation is illustrated using plots of the empirical autocorrelation functions as seen from Figure 2.4. The 95% level of significance is $\pm 2/\sqrt{1264} = 0.0563$.

Focusing on the autocorrelations of Mobil, it is readily seen that the autocorrelation of the squared returns is significant differently from zero at lags 2, 5, 8, 10, 11, 13, 14, 15 and 25 whereas the autocorrelation function of the ordinary returns is significantly different form zero at lags 2, 4 and 6.

Formal methods for addressing the presence of autocorrelation exist. A Ljung-Box test uses the empirical values of the autocorrelation function to test for autocorrelation in the returns, squared or not. If the squared returns are used the test is denoted a *Portmanteau Q-test*.

NOTE 4 (LJUNG-BOX AND PORTMANTEAU).
Given observations u_t, $t = 1, 2, \ldots, T$ the Ljung-Box test (McLeod & Li 1983) uses the empirical estimates of the autocorrelation function to test for autocorrelations in the observations. If the squared observations are used, the test is denoted a *Portmanteau Q-test* and the test is then used to test for autocorrelation in the squared returns. The test assumes that the observations u_t are independent and identically distributed with finite variance.

The Ljung-Box statistic is

$$Q(q) = N(N+2) \sum_{k=1}^{q} \frac{\hat{\rho}^2(k)}{N - k} \tag{2.2.9}$$

where the estimate of the autocorrelation function in the case of ordinary returns is

$$\hat{\rho}(k) = \frac{\sum_{t=k+1}^{T}(u_t - \hat{\mu})(u_{t-k} - \hat{\mu})}{\sum_{t=1}^{T}(u_t - \hat{\mu})^2} \tag{2.2.10}$$

where $\hat{\mu}$ is the sample mean,

$$\hat{\mu} = \frac{1}{N}\sum_{t=1}^{T} u_t. \tag{2.2.11}$$

In the case of squared returns, u_t is replaced by u_t^2 yielding

$$\hat{\rho}(k) = \frac{\sum_{t=k+1}^{T}(u_t^2 - \hat{\sigma}^2)(u_{t-k}^2 - \hat{\sigma}^2)}{\sum_{t=1}^{T}(u_t^2 - \hat{\sigma}^2)^2} \tag{2.2.12}$$
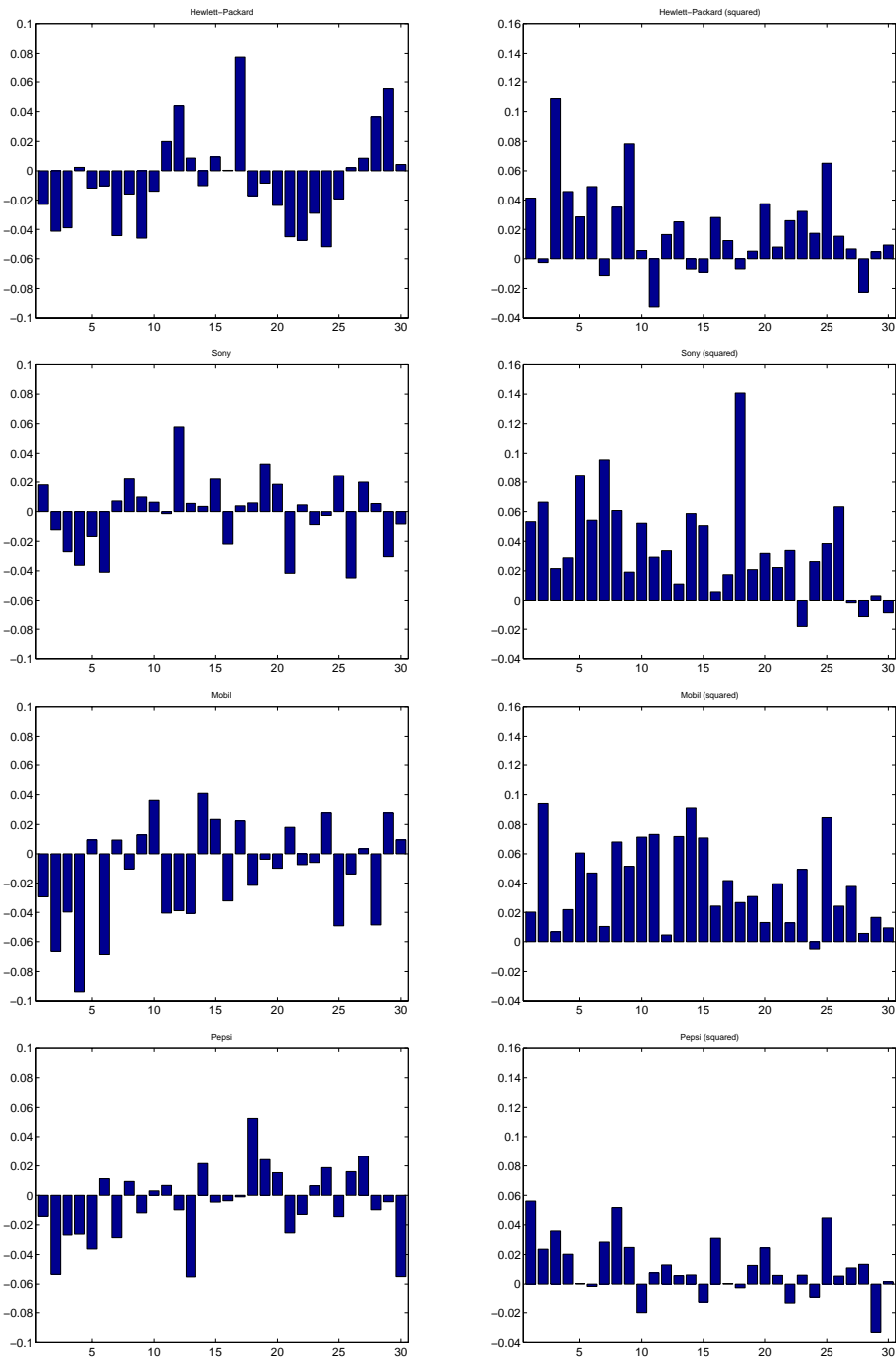
Figure 2.4: Autocorrelations of ordinary returns (left column) and squared returns (right column). The plots in the first row are from the Hewlett-Packard return series, the second row is from Sony, the third row is Mobil and the last row is Pepsi.

where $\hat{\sigma}^2$ is the sample second order moment

$$\hat{\sigma}^2 = \frac{1}{N} \sum_{t=1}^{T} u_t^2. \tag{2.2.13}$$

Under the hypothesis of no autocorrelations, the test statistic is asymptotically $\chi^2(q)$ distributed.

The test indicates whether the observations exhibits autocorrelation for lags less than or equal to q.

The results of carrying out the Ljung-Box and the Portmanteau Q-test on the four return series are as follows: For all series, the squared returns exhibits autocorrelation at the 95% level. Mobil is the only series exhibiting autocorrelation in the ordinary returns.

These findings corresponds to the description of stylized fact number 3 earlier in this section, though it is a bit surprising that the autocorrelation function of the ordinary returns of Mobil are significant. In line with the literature, this autocorrelation is disregarded.

In Table 2.5 the test results are presented in their entirety.

A warning

One should be careful when using autocorrelation function estimates on data that exhibits non-linearity and heavy tails. Studies by (Davis & Resnick 1996) have revealed that for this kind of data, autocorrelation function estimates based on (2.2.10) or (2.2.12) have unknown asymptotical properties, i.e. the estimator of the autocorrelation function may not converge to the true autocorrelation function. Because of this, the validity of the autocorrelation function plots and the Ljung-Box test applied to the return series is, strictly speaking, unknown. The reason for using the autocorrelation function estimates anyway is that they seem to be widely used in financial applications and there does not seem to exist a suitable alternative.

2.2.2.4 Leverage effect

The obvious way to look for indications of a leverage effect in a return series would be to construct a series of the squared returns and then look for negative crosscor-

Series	HP		Sony		Mobil		Pepsi	
Return type	Ord	Sq	Ord	Sq	Ord	Sq	Ord	Sq
1	41.4	14.1	51.8	5.8	29.5	47.3	61.3	**4.6**
2	24.4	33.6	73.7	**1.0**	**3.5**	**0.3**	14.5	9.6
3	19.1	**0.1**	67.3	**2.1**	**3.3**	**0.8**	18.9	9.7
4	31.3	**0.1**	52.4	**2.9**	**0.1**	**1.5**	22.7	14.4
5	42.4	**0.1**	61.4	**0.1**	**0.1**	**0.4**	19.8	23.2
6	53.5	**0.1**	45.8	**0.1**	**0.0**	**0.3**	27.9	33.5
7	37.2	**0.1**	56.8	**0.0**	**0.0**	**0.6**	28.9	34.3
8	44.4	**0.1**	60.3	**0.0**	**0.1**	**0.1**	37.4	18.6
9	30.5	**0.0**	68.7	**0.0**	**0.2**	**0.1**	45.5	21.0
10	37.0	**0.0**	76.5	**0.0**	**0.2**	**0.0**	54.9	24.8
11	41.5	**0.0**	83.3	**0.0**	**0.1**	**0.0**	63.3	31.7
12	31.2	**0.0**	54.2	**0.0**	**0.1**	**0.0**	70.2	37.9
13	38.0	**0.1**	62.0	**0.0**	**0.1**	**0.0**	45.6	45.5
14	44.6	**0.1**	69.4	**0.0**	**0.1**	**0.0**	48.8	53.0
15	51.3	**0.2**	71.4	**0.0**	**0.1**	**0.0**	56.2	59.0
16	58.6	**0.2**	73.4	**0.0**	**0.1**	**0.0**	63.3	56.9
17	19.0	**0.3**	79.0	**0.0**	**0.2**	**0.0**	70.0	63.9
18	22.1	**0.4**	83.6	**0.0**	**0.2**	**0.0**	51.8	70.2
19	26.8	**0.7**	80.8	**0.0**	**0.4**	**0.0**	53.3	74.7
20	28.6	**0.6**	82.9	**0.0**	**0.5**	**0.0**	57.8	75.4
21	21.9	**0.9**	75.5	**0.0**	**0.7**	**0.0**	58.7	80.1
22	15.7	**1.0**	80.1	**0.0**	**1.0**	**0.0**	63.4	83.2
23	15.9	**1.0**	83.8	**0.0**	**1.4**	**0.0**	68.7	86.7
24	10.1	**1.3**	87.3	**0.0**	**1.5**	**0.0**	71.4	89.4
25	11.7	**0.4**	87.5	**0.0**	**1.0**	**0.0**	74.8	82.4
26	14.5	**0.6**	80.2	**0.0**	**1.3**	**0.0**	77.7	85.9
27	17.5	**0.8**	81.8	**0.0**	**1.7**	**0.0**	77.6	88.4
28	15.8	**0.9**	85.2	**0.0**	**1.1**	**0.0**	81.1	90.3
29	9.4	**1.3**	83.9	**0.0**	**1.2**	**0.0**	84.6	88.4
30	11.6	**1.7**	86.7	**0.0**	**1.6**	**0.0**	71.1	90.9

Table 2.5: The Ljung-Box and Portmanteau test results. The table does not display the test statistics, only the probability (in percent) that a $\chi^2(n)$-distribution exceeds the test statistic, where n is the lag number. Column headers of "Ord" indicate ordinary returns, "Sq" indicates squared returns. Values significant at the 95% level are printed in **bold**.

relations for positive lags between the series of ordinary returns and the series of squared returns.

However, doing so has not revealed any leverage effect in any of the four stocks. The series *do* (not surprisingly) exhibit crosscorrelations, but these are positive as well as negative with no predominant sign. For instance, the estimates crosscorrelation function for ordinary and squared Hewlett-Packard returns is as displayed in Figure 2.5.
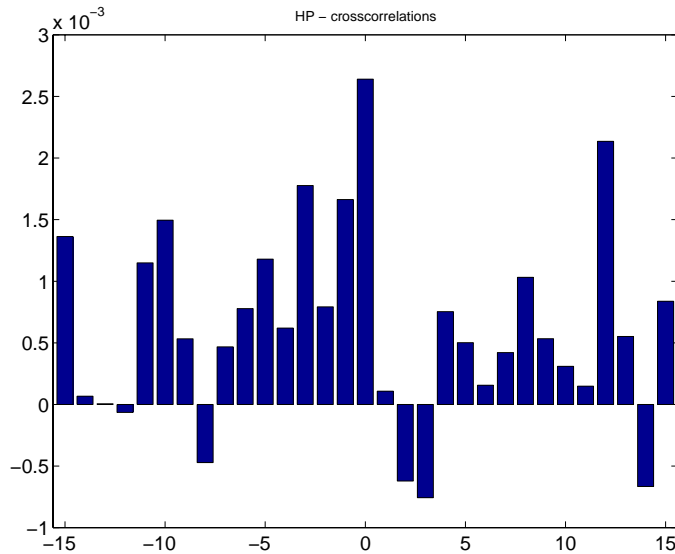
Figure 2.5: Empirical crosscorrelation function of ordinary and squared Hewlett-Packard returns, lags and leaps from 0 to 15.

Another way to look for the leverage effect is to estimate the conditional variance using non-parametric methods, c.f. (Madsen & Holst 1996, Chapter 2). Estimates of the conditional variance $V(R_t|R_{t-1})$ of the four return series computed using a Gaussian kernel are plotted in Figure 2.6.

From the plot the leverage effect of Hewlett-Packard is evident. Comparing the conditional variance of returns equal in size but with different signs yields that the conditional variance of the negative return is larger. The other three return series do not seem to exhibit the same degree of leverage effect thus suggesting that the leverage is not present in all return series or that the effect is hidden by other more dominant characteristics.
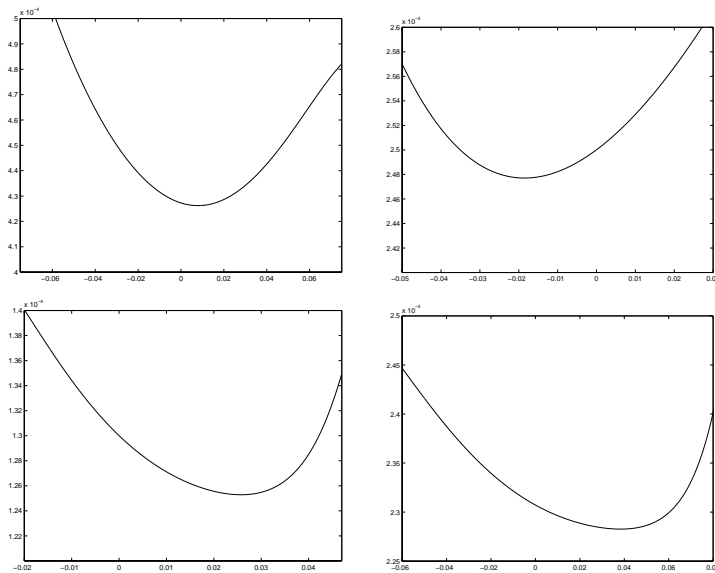
Figure 2.6: Conditional variances estimated using a non-parametric method. The estimates are from the series Hewlett-Packard (upper left), Sony (upper right), Mobil (lower left) and Pepsi (lower right). All the variances are conditional on the observation in lag one, and all have been estimated using a Gaussian kernel. The kernel bandwiths are 0.175, in the cases of Hewlett-Packard and Mobil and 0.185 for Sony and Pepsi.

2.2.2.5   Summary

In this section a number of return series characteristics have been presented. These include excess kurtosis, heteroskedasticity, autocorrelation in the squared returns as well as the leverage effect. It has been shown how to verify the presence of these characteristics by means of visual inspection and by the use of statistical tests. The techniques have been illustrated on four return series and the series have been proven to exhibit all of the characteristics except the leverage effect.

In the next chapter models will be introduced that capture the aforementioned characteristics. But before proceeding to model the return series characteristics, the concept of volatility must be defined.

2.3 Volatility

It is not obvious how to define the volatility of a return series. Traditionally, the standard deviation of the returns has been used as a measure of volatility. This measure is easy to estimate and to interpret as long as the volatility is constant. Unfortunately, as previously revealed, standard deviations of stock returns are not constant. Todays level of volatility may be different from yesterdays level, a fact that suggests that what is needed is a measure of the instantaneous volatility.

Though it may be possible to talk about the standard deviation of a single return, it is not at all obvious how to estimate this standard deviation. However, if it is assumed in line with (Shephard 1996) that the returns R_t may be described by a simple model, then concept of volatility may be defined.

DEFINITION 2.3 (DISCRETE TIME STOCHASTIC VOLATILITY).
Assuming a return series R_t can be represented by the model

$$R_t = \mu_t + \sigma_t \varepsilon_t \tag{2.3.1}$$

where μ_t describes the mean level of the return at time t, $\sigma_t \varepsilon_t$ describes the variation about the mean and ε_t is a zero mean stationary identically independently distributed random variable, the volatility of R_t is defined by σ_t.   ▲

For the purpose of this thesis, the model (2.3.1) is perhaps too unrestrictive to facilitate an easy-to-use definition. The main problem is that μ_t is allowed to be stochastic thus leaving no clear indication of where to draw the line between μ_t and $\sigma_t \varepsilon_t$ of (2.3.1). Thus, in the rest of this thesis it will be assumed that μ_t is stochastic but that $\mu_t | \psi_{t-1}$ is deterministic, where ψ_{t-1} is the information set available at time $t \Leftrightarrow 1$.

It is of course a limitation that the definition of volatility depends on a model. Intuitively, one would perceive the concept of volatility to be a property of the underlying process generating the returns and thus independent of the choice of describing model. Unfortunately, there is no workaround as the instantanious volatility is not observable. Looking at the bright side, the model (2.3.1) is very general.

Chapter 3

Discrete time GARCH models

The previous chapter presented a lot of return series characteristics, characteristics that are mostly related to the volatility of the returns.

This chapter presents a number of models that all can be used to model the return series volatility. A lot of different ARCH and GARCH models are described with emphasis on the abilities of the models to incorporate the return series characteristics. Descriptions of the origins of the different models are included.

The chapter is sectionized as follows: Section 3.1 establishes the general framework for ARCH and GARCH models. Section 3.2 deals with the ARCH process by (Engle 1982). Section 3.3 introduces the GARCH process by (Bollerslev 1986) and Section 3.4 is devoted to the EGARCH process proposed in (Nelson 1991). Section 3.5 handles various processes derived from ARCH and GARCH and finally, Section 3.6 deals with the Augmented GARCH process by (Duan 1997).

3.1   The general framework

As mentioned above, a large number of different ARCH and GARCH models exist. To keep things clear, this section establishes the framework common to all the models to be described later.

DEFINITION 3.1 (THE ARCH PROCESS).
Consider the stochastic process $\{\varepsilon_t\}$ and the information set ψ_{t-1} where

$$\varepsilon_t = z_t \sigma_t \qquad\qquad (3.1.1)$$

with σ_t positive and measurable by ψ_{t-1} and the z_t's identically independently distributed with

$$\mathrm{E}(z_t) = 0, V(z_t) = 1 \qquad\qquad (3.1.2)$$

and z_t independent of σ_t for all t.

An *Arch* process is defined as a process $\{\varepsilon_t\}$ that can be written on the form (3.1.1) and (3.1.2). ▲

In line with (Bera & Higgins 1993) and (Bollerslev, Chou & Kroner 1992) (3.1.1) and (3.1.2) will be referred to by Arch (capital "A", lower case "rch"), whereas specific models under this framework are given denotations in all capitals. Note that what remains to be specified for the Arch model to be employable is the process driving $\{\sigma_t\}$ and the distribution of z_t. In most of the literature this distribution is chosen to be normal, but alternative distributions are not rarely seen.

Often, $\{\varepsilon_t\}$ are innovations in the mean for some other stochastic process $\{y_t\}$ where

$$y_t = g(x_t, b) + \varepsilon_t \qquad\qquad (3.1.3)$$

and x_t is in the information set ψ_{t-1} and b is a parameter vector. In most of the cases discussed in this thesis, though, $\{\varepsilon_t\}$ will be directly observable.

3.2   ARCH

3.2.1   Definition

The ARCH process was proposed by (Engle 1982) as the first successful attempt to model conditional heteroskedasticity. In its most general form[1], the ARCH(q) process is defined by

---

[1]In his original paper, Engle assumed $g(x_t, b) = x_t^\top b$ and thus considered $y_t = x_t^\top b + \varepsilon_t$.

DEFINITION 3.2 (THE ARCH(Q) PROCESS).
An ARCH(q) process is an Arch process with

$$\sigma_t = \sqrt{h_t} \qquad (3.2.1)$$

and

$$h_t = \alpha_0 + \alpha_1 \varepsilon_{t-1}^2 + \cdots + \alpha_q \varepsilon_{t-q}^2, \qquad (3.2.2)$$

where z_t is Gaussian distributed. ▲

An immediate consequence of Definition 3.2 is that the innovations ε_t of the ARCH(q) process are conditionally normal,

$$\varepsilon_t | \psi_{t-1} \in N\left(0, h_t\right). \qquad (3.2.3)$$

For the conditional variance of an ARCH(q) process to remain positive, the parameters are restricted to $\alpha_0 > 0$ and $\alpha_i \geq 0, i = 1, \ldots, q$.

The most interesting part of Definition 3.2 is the function h. The definition ensures that the variance of ε_t, conditional on the realized values $\varepsilon_{t-1}, \ldots \varepsilon_{t-q}$, is an increasing function of the magnitude of the lagged ε's, irrespective of their signs. Thus, large returns tend to be followed by large returns of either sign and small returns tend to be followed by small returns of either sign. This behaviour is definitely in line with the previously mentioned observations by (Mandelbrot 1963). The order q of the process determines the volatility persistence and this persistence increases with values of q.

3.2.2 Properties

To illustrate the properties of ARCH(q) processes, an ARCH(4) process has been simulated. The simulated data are plotted in Figure 3.1, and for comparison the geometric returns of Hewlett-Packard are plotted along with Gaussian White Noise in Figure 3.2.

From the figures it is obvious that the simulated ARCH(4) process have some characteristics in common with the Hewlett-Packard returns, characteristics that Gaussian White Noise do not possess. The first thing to notice is that the ARCH(4) data
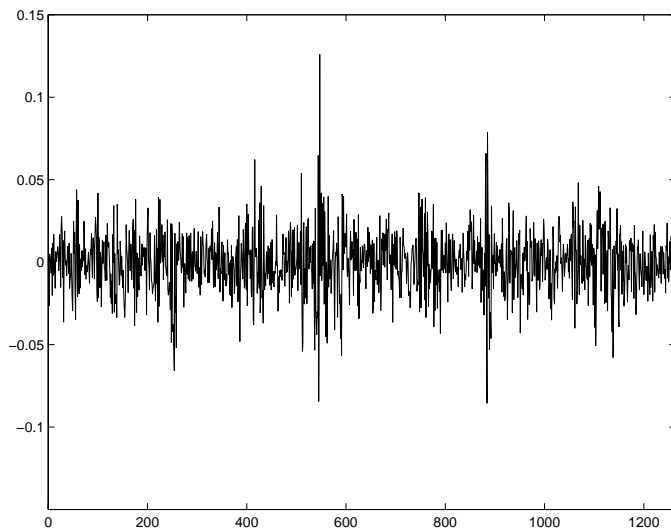
Figure 3.1: Simulated ARCH(4) data. $h_t = 10^{-4} + 0.3\varepsilon_{t-1}^2 + 0.27\varepsilon_{t-2}^2 + 0.18\varepsilon_{t-3}^2 + 0.09\varepsilon_{t-4}^2$.

has a number of outliers. The next thing is that the ARCH data exhibits some of the volatility clustering observable in the Hewlett-Packard returns.

The order and the parameters of the ARCH process in Figure 3.2 have been chosen somewhat arbitrarily as the purpose of the plot is merely to illustrate what ARCH processes are capable of. The in-depth analysis of the ability of ARCH to describe the return series characteristics is deferred to the next chapter where also parameter estimation techniques are described.

3.3 GARCH

3.3.1 Definition

Although ARCH models provide a good description of many return series, it suffers from a practical problem. As one may or may not be able to see from Figure 3.1, the volatility persistence is not quite as high as is the case of the Hewlett-Packard returns. This means that a large lag q is needed for the ARCH(q) process to provide an adequate description of the returns, but, as an ARCH(q) process has $q + 1$ parameters, the immediate consequence of this is that a large number of
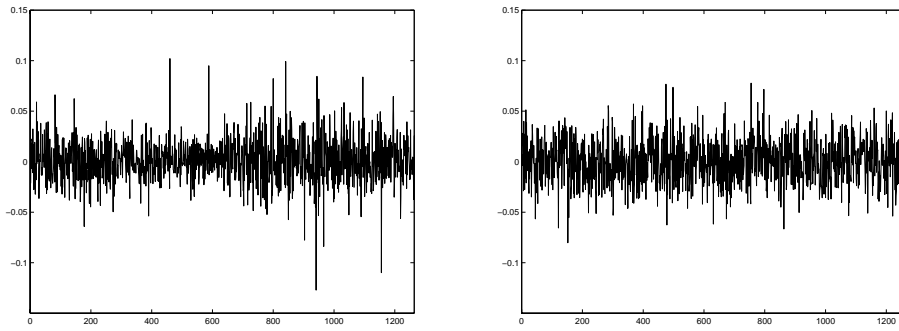
Figure 3.2: Returns of Hewlett-Packard (left) and Gaussian White Noise (right).

parameters have to be estimated, all subject to parameter restrictions. This imposes a serious computational burden.

To circumvent this problem, Bollerslev proposed the GARCH process in (Bollerslev 1986). Formally, the GARCH(p, q) process only differs from the ARCH(q) process in the way that the function h is specified:

DEFINITION 3.3 (THE GARCH(P,Q) PROCESS).
A GARCH(p,q) process is an Arch process with

$$\sigma_t = \sqrt{h_t} \tag{3.3.1}$$

and

$$h_t = \alpha_0 + \sum_{i=1}^{q} \alpha_i \varepsilon_{t-i}^2 + \sum_{j=1}^{p} \beta_i h_{t-i}, \tag{3.3.2}$$

where z_t is Gaussian distributed.   ▲

The difference from ARCH to GARCH is that in the GARCH case the function h_t includes lagged values of itself. In applications, a GARCH process has fewer parameters than an ARCH process as a GARCH(1,1) suffices in most situations. Note that an ARCH(q) process is identical to a GARCH(0, q) process.

To ensure that the conditional variance remains positive, the following parameter restrictions suffices:

$$\alpha_0 \quad > \quad 0 \qquad \qquad (3.3.3a)$$

$$\alpha_i \quad \geq \quad 0, \quad \text{for} \quad i = 1, \ldots, q \qquad \qquad (3.3.3b)$$

$$\beta_j \quad \geq \quad 0, \quad \text{for} \quad j = 1, \ldots, p. \qquad \qquad (3.3.3c)$$

Although the restrictions (3.3.3) are sufficient to ensure non-negativity of the conditional variance, they are not necessary for higher order GARCH processes (p or q greater than one) in which case weaker sufficient conditions may exist, see (Cao & Nelson 1992). These conditions are, however, outside the scope of this thesis as this thesis is primarily concerned with low order GARCH processes.

The motivation of the form of h in (3.3.2) is perhaps more obvious if (3.3.2) is rewritten:

$$h_t = \alpha_0 + \alpha(B)\varepsilon_t^2 + \beta(B)h_t \qquad \qquad (3.3.4)$$

where $\alpha(B) = \sum_{i=1}^{q} \alpha_i B^i$ and $\beta(B) = \sum_{j=1}^{p} \beta_j B^j$ are polynomials in the backshift operator B. If the roots of the polynomial $1 \Leftrightarrow \beta(L)$ lie outside the unit circle, (3.3.2) may be expressed as

$$h_t = \frac{\alpha_0}{1 \Leftrightarrow \beta(1)} + \frac{\alpha(B)}{1 \Leftrightarrow \beta(B)}\varepsilon_t^2 = \alpha_0 \left(1 \Leftrightarrow \sum_{j=1}^{p} \beta_j\right)^{-1} + \sum_{i=1}^{\infty} \delta_i \varepsilon_{t-i}^2 \quad (3.3.5)$$

where δ_i is the coefficient of B^i in the Taylor expansion of $\alpha(B)[1 \Leftrightarrow \beta(B)]^{-1}$. Hence, a GARCH(p, q) process may be looked upon as an infinite order ARCH process.

3.3.2 Properties

At Figure 3.3 the plot of a simulated series of a GARCH(1,1) returns is shown.

From the plot the higher volatility persistence of the GARCH process compared to the ARCH process (Figure 3.1) is obvious. The GARCH process also seems to provide a better description of the Hewlett-Packard returns (Figure 3.2).

To give an idea of the GARCH process ability to incorporate the return series characteristics, it is illustrative to consider the unconditional moments and the autocorrelation structure of the process. Unfortunately, no analytical expression of the unconditional distribution of the GARCH process exist.
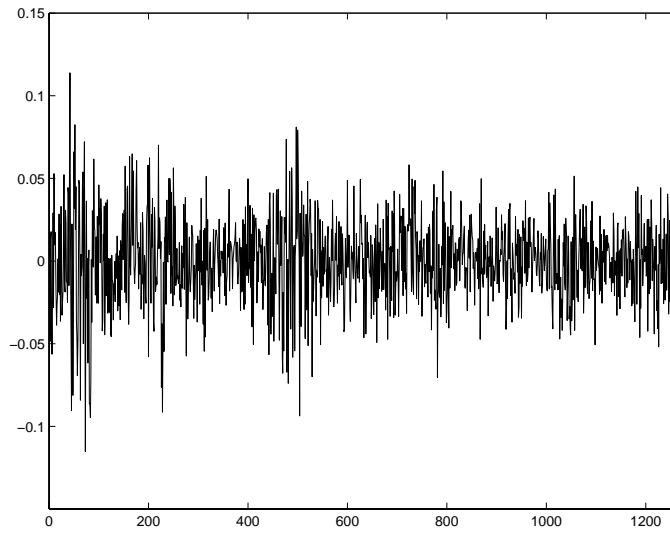
Figure 3.3: Simulated GARCH(1,1) data. $h_t = 2 \times 10^{-5} + 1.0412 \times 10^{-1}\varepsilon_{t-1} + 0.87h_{t-1}$.

3.3.2.1 Unconditional moments

Mean

Using the law of iterated expectations, the unconditional mean may be computed by

$$E(\varepsilon_t) = E[E(\varepsilon_t|\psi_{t-1})] \tag{3.3.6}$$

and, refering to (3.2.3), $E(\varepsilon_t|\psi_{t-1})$ is zero regardless of ψ_{t-1} meaning that $E(\varepsilon_t) = 0$. Thus, GARCH processes have zero mean.

Variance

Since $E\{\varepsilon_t\} = 0$, the unconditional variance is identical to the unconditional sec-

ond order moment which may be computed using

$$
\begin{aligned}
\mathrm{E}(\varepsilon_t^2) &= \mathrm{E}\left[\mathrm{E}(\varepsilon_t^2|\psi_{t-1})\right] \\
&= \mathrm{E}\left[\mathrm{E}(z_t^2\sigma_t^2|\psi_{t-1})\right] \\
&= \mathrm{E}(h_t) \\
&= \alpha_0 + \sum_{i=1}^{q}\alpha_i\mathrm{E}(\varepsilon_{t-i}^2) + \sum_{j=1}^{p}\beta_j\mathrm{E}(h_{t-j}) \\
&= \alpha_0 + \sum_{i=1}^{q}\alpha_i\mathrm{E}(\varepsilon_{t-i}^2) + \sum_{j=1}^{p}\beta_j\mathrm{E}(\varepsilon_{t-j}^2)
\end{aligned}
\tag{3.3.7}
$$

which is a linear difference equation for the sequence of unconditional second order moments. If it is assumed that the process began infinitely far in the past with a finite initial second order moment, the sequence of second order moments converges to

$$
\sigma_\varepsilon^2 = \mathrm{E}(\varepsilon_t^2) = \frac{\alpha_0}{1 \Leftrightarrow \sum_{i=1}^{q}\alpha_i \Leftrightarrow \sum_{j=1}^{p}\beta_j}
\tag{3.3.8}
$$

if

$$
\sum_{i=1}^{q}\alpha_i + \sum_{j=1}^{p}\beta_j < 1.
\tag{3.3.9}
$$

If this condition is satisfied, the variance of the GARCH(p, q) process is equal to (3.3.8).

Skewness

Because z_t in (3.1.1) is normal and independent of σ_t, all the odd higher order moments of ε_t are zero. Hence, the skewness of ε_t is also zero.

Kurtosis

Expressions for even higher order moments exist only for GARCH(1, 1) processes and GARCH(0,1) (that is, ARCH(1)) processes. For an ARCH(1) process the forth order moment exists if $3\alpha_1^2 < 1$ and in this case the kurtosis is

$$
\frac{\mathrm{E}(\varepsilon_t^4)}{\mathrm{E}(\varepsilon_t^2)^2} = 3\left(\frac{1 \Leftrightarrow \alpha_1^2}{1 \Leftrightarrow 3\alpha_1^2}\right)
\tag{3.3.10}
$$

which is greater than three indicating that the ARCH(1) process has excess kurtosis.

For a GARCH(1,1) process the forth order moment exists if $3\alpha_1^2 + 2\alpha_1\beta_1 + \beta_1^2 < 1$ and is then given by

$$E[\varepsilon_t^4] = \frac{3\alpha_0^2(1+\alpha_1+\beta_1)}{(1-\alpha_1-\beta_1)(1-\beta_1^2-2\alpha_1\beta_1-3\alpha_1^2)} \tag{3.3.11}$$

such that the *excess* kurtosis is

$$\frac{E(\varepsilon_t^4)}{E(\varepsilon_t^2)^2} - 3 = \frac{E[\varepsilon_t^4] - 3E[\varepsilon_t^2]^2}{E[\varepsilon_t^2]^2} = \frac{6\alpha_1^2}{1-\beta_1^2-2\alpha_1\beta_1-3\alpha_1^2} \tag{3.3.12}$$

which is greater than zero. Hence, every GARCH(1,1) process has excess kurtosis.

3.3.2.2 Autocorrelation structure

The autocovariance of a GARCH(p, q) process is for $n \geq 1$

$$
\begin{aligned}
E(\varepsilon_t \varepsilon_{t-n}) &= E(z_t \sigma_t z_{t-n} \sigma_{t-n}) \\
&= E(z_t z_{t-n})E(\sigma_t \sigma_{t-n}) = 0
\end{aligned}
\tag{3.3.13}
$$

as the z_t's are mutually independent and independent of σ_t. Thus, a GARCH(p, q) process exhibits no autocorrelation.

The autocovariance of $\{\varepsilon_t^2\}$ is not as easy to derive, but it is possible if the process driving $\{\varepsilon_t^2\}$ is first established.

Define the process $\{v_t\}$ by $v_t = \varepsilon_t^2 - h_t$ (note that v_t has zero mean) and consider again Equation 3.3.2:

$$h_t = \alpha_0 + \sum_{i=1}^{q} \alpha_i \varepsilon_{t-i}^2 + \sum_{j=1}^{p} \beta_i h_{t-i}. \tag{3.3.14}$$

Inserting $h_t = \varepsilon_t^2 - v_t$ yields

$$
\begin{aligned}
\varepsilon_t^2 - v_t &= \alpha_0 + \sum_{i=1}^{q} \alpha_i \varepsilon_{t-i}^2 + \sum_{j=1}^{p} \beta_j (\varepsilon_{t-j}^2 - v_{t-j}) \Leftrightarrow \\
\varepsilon_t^2 &= \alpha_0 + \sum_{i=1}^{q} \alpha_i \varepsilon_{t-1}^2 + \sum_{j=1}^{p} \beta_j \varepsilon_{t-j}^2 - \sum_{j=1}^{p} \beta_j v_{t-j} + v_t. \quad (3.3.15)
\end{aligned}
$$

For ease of notation define $m = \max(p, q)$ and set $\alpha_i \equiv 0$ for $i > q$ and $\beta_j \equiv 0$ for $j > p$ and write

$$
\varepsilon_t^2 = \alpha_0 + \sum_{i=1}^{m} (\alpha_i + \beta_i) \varepsilon_{t-i}^2 - \sum_{j=1}^{p} \beta_j v_{t-j} + v_t, \quad (3.3.16)
$$

as the process driving ε_t^2.

From Equation 3.3.16 the autocorrelation structure of $\{\varepsilon_t^2\}$ may be derived. Recall from (3.3.8) that the mean of ε_t is $\sigma_\varepsilon^2 = \alpha_0 (1 - \sum_{i=1}^{q} \alpha_i - \sum_{j=1}^{p} \beta_j)^{-1}$ meaning that the parameter α_0 may be written $\alpha_0 = \sigma_\varepsilon^2 (1 - \sum_{i=1}^{q} \alpha_i - \sum_{j=1}^{p} \beta_j)$. Inserting this into (3.3.16) yields

$$
\begin{aligned}
\varepsilon_t^2 &= \sigma_\varepsilon^2 \left[1 - \sum_{i=1}^{m} (\alpha_i - \beta_i) \right] + \sum_{i=1}^{m} (\alpha_i + \beta_i) \varepsilon_{t-i}^2 - \sum_{j=1}^{p} \beta_j v_{t-j} + v_t \\
&= \sigma_\varepsilon^2 + \sum_{i=1}^{m} (\alpha_i + \beta_i)(\varepsilon_{t-i}^2 - \sigma_\varepsilon^2) - \sum_{j=1}^{p} \beta_j v_{t-j} + v_t \quad (3.3.17)
\end{aligned}
$$

which is equivalent to

$$
\varepsilon_t^2 - \sigma_\varepsilon^2 = \sum_{i=1}^{m} (\alpha_i + \beta_i)(\varepsilon_{t-i}^2 - \sigma_\varepsilon^2) - \sum_{j=1}^{p} \beta_j v_{t-j} + v_t. \quad (3.3.18)
$$

Multiplying both sides of this equation by $(\varepsilon_{t-n}^2 - \sigma_\varepsilon^2)$ yields

$$
\begin{aligned}
(\varepsilon_t^2 - \sigma_\varepsilon^2)(\varepsilon_{t-n}^2 - \sigma_\varepsilon^2) & \\
= \sum_{i=1}^{m} (\alpha_i + \beta_i)(\varepsilon_{t-i}^2 - \sigma_\varepsilon^2)(\varepsilon_{t-n}^2 - \sigma_\varepsilon^2) & \\
- \sum_{j=1}^{p} \beta_j v_{t-j}(\varepsilon_{t-n}^2 - \sigma_\varepsilon^2) + v_t(\varepsilon_{t-n}^2 - \sigma_\varepsilon^2) &
\end{aligned}
\tag{3.3.19}
$$

and if expectations are taken on both sides:

$$
\begin{aligned}
\mathrm{E}[(\varepsilon_t^2 - \sigma_\varepsilon^2)(\varepsilon_{t-n}^2 - \sigma_\varepsilon^2)] = & \\
\mathrm{E}\left[\sum_{i=1}^{m}(\alpha_i + \beta_i)(\varepsilon_{t-i}^2 - \sigma_\varepsilon^2)(\varepsilon_{t-n}^2 - \sigma_\varepsilon^2)\right] & \\
- \mathrm{E}\left[\sum_{j=1}^{p}\beta_j v_{t-j}(\varepsilon_{t-n}^2 - \sigma_\varepsilon^2)\right] + \mathrm{E}\left[v_t(\varepsilon_{t-n}^2 - \sigma_\varepsilon^2)\right]. &
\end{aligned}
\tag{3.3.20}
$$

The two terms in the last line are both zero as

$$
\begin{aligned}
\mathrm{E}[v_t(\varepsilon_{t-n}^2 - \sigma_\varepsilon^2)] &= \mathrm{E}\left\{\mathrm{E}[v_t(\varepsilon_{t-n}^2 - \sigma_\varepsilon^2)|\psi_{t-1}]\right\} \\
&= \mathrm{E}\left\{(\varepsilon_{t-n}^2 - \sigma_\varepsilon^2)\mathrm{E}[v_t|\psi_{t-1}]\right\} \\
&= \mathrm{E}\{0\} = 0
\end{aligned}
\tag{3.3.21}
$$

and

$$
\begin{aligned}
\mathrm{E}[v_{t-j}(\varepsilon_{t-n}^2 - \sigma_\varepsilon^2)] &= \mathrm{E}\left\{\mathrm{E}[v_{t-j}(\varepsilon_{t-n}^2 - \sigma_\varepsilon^2)|\psi_{t-n}]\right\} \\
&= \mathrm{E}\left\{(\varepsilon_{t-n}^2 - \sigma_\varepsilon^2)\mathrm{E}[v_{t-j}|\psi_{t-n}]\right\} \\
&= \mathrm{E}\{0\} = 0
\end{aligned}
\tag{3.3.22}
$$

for $j < n$ where the fact that v_t has zero mean has been utilized. Thus the autocovariance of a GARCH(p, q) process may be written

$$
\begin{aligned}
\mathrm{Cov}[\varepsilon_t^2, \varepsilon_{t-n}^2] &= \mathrm{E}[(\varepsilon_t^2 - \sigma_\varepsilon^2)(\varepsilon_{t-n}^2 - \sigma_\varepsilon^2)] \\
&= \sum_{i=1}^{m}(\alpha_i + \beta_i)\mathrm{E}\left[(\varepsilon_t^2 - \sigma_\varepsilon^2)(\varepsilon_{t-n}^2 - \sigma_\varepsilon^2)\right] \\
&= \sum_{i=1}^{m}(\alpha_i + \beta_i)\,\mathrm{Cov}(\varepsilon_t^2, \varepsilon_{t-n+i}).
\end{aligned}
\tag{3.3.23}
$$

Dividing both sides with $V[\varepsilon_t^2]$ and denoting the autocorrelation function of ε_t^2 by ρ_n yields

$$\rho_n = \sum_{i=1}^{m} (\alpha_i + \beta_i) \rho_{n-i}, \qquad n > p \tag{3.3.24}$$

which is actually the set of Yule-Walker equations for an ARMA(m, p) process, see (Bollerslev 1988). Hence, the autocorrelation function of $\{\varepsilon_t^2\}$ GARCH(p, q) process is similar to that of an ARMA(m, p) process. This means that estimates of ρ_n may be used to identify the order of an observed GARCH process.

Equation 3.3.24 does not specify the values of the autocorrelation function when $n \leq p$. It is very difficult to find analytical expressions for these autocorrelations, but it has been done by (Bollerslev 1988) and (He & Teräsvirta 1999) for low order Arch processes. For a GARCH(1,1) process $\mathrm{Cor}(\varepsilon_t^2, \varepsilon_{t-1}^2)$ is

$$\rho_1 = \frac{\alpha_1 (1 \Leftrightarrow \alpha_1 \beta_1 \Leftrightarrow \beta_1^2)}{1 \Leftrightarrow 2\alpha_1 \beta_1 \Leftrightarrow \beta_1^2}. \tag{3.3.25}$$

In Figure 3.4 the autocorrelation function for lags 1 to 250 of the GARCH process from Figure 3.3 is plotted. The geometrical decay is obvious.

From the above paragraphs it is readily seen that GARCH processes seem able to model the stylized facts of returns. GARCH innovations have heavy tails, are heteroskedastic and exhibit no autocorrelation in the ordinary innovations, but significant autocorrelation in the squared innovations.

3.4 EGARCH

3.4.1 Definition

One of the most commonly known objections to the GARCH models is that though they are able to model the stylized facts of return series they are not able to take the leverage effect into account. This is due to the fact that GARCH models only deal with the squared returns and thus are unable to discriminate positive and negative returns.

To remedy this problem, (Nelson 1991) introduced the so-called EGARCH process.
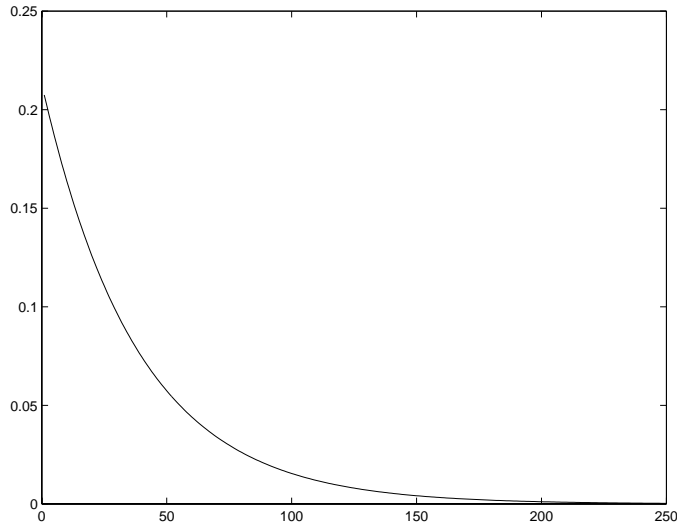
Figure 3.4: Autocorrelation function of the squared $\mathsf{GARCH(1,1)}$ process where $h_t = 2 \times 10^{-5} + 1.0412 \times 10^{-1} \varepsilon_{t-1} + 0.87 h_{t-1}$.

DEFINITION 3.4. The EGARCH(p,q) process
An EGARCH(p,q) process is an Arch process with

$$\sigma_t = \exp(\frac{1}{2} h_t) \tag{3.4.1}$$

and

$$h_t = \alpha_0 + \sum_{i=1}^{q} \alpha_i f(z_{t-i}) + \sum_{j=1}^{p} \beta_j h_{t-j} \tag{3.4.2}$$

and

$$f(z_{t-i}) = \omega z_{t-i} + \lambda(|z_{t-i}| \Leftrightarrow \mathrm{E}|z_{t-i}|) \tag{3.4.3}$$

▲

Again, ψ_{t-1} is the information set available at time $t \Leftrightarrow 1$. If z_t is Gaussian, $\mathrm{E}|z_t| = \sqrt{2/\pi}$. One of the parameters ω and λ is of cource redundant, but they are both

included in the specification because they facilitate an easy interpretation of the model as described below.

A side-effect of the specification is that unlike the case of ARCH and GARCH processes, no parameter restrictions are needed in order to ensure positivity of σ_t^2.

3.4.2 Properties

The properties of the EGARCH process are primarily determined by the function f. The rôle of f is to transform the innovations z_t before "turning them over" to h_t. When z_{t_i} is positive, the derivative of h_t with respect to z_{t-i} is equal to $\alpha_i(\omega + \lambda)$. When z_{t-i} is negative, this derivative is $\alpha_i(\omega \Leftrightarrow \lambda)$. Thus, h_t is affected by both the sign and the size of z_{t-i} and therefore the EGARCH process allows for negative correlation between returns and conditional variances.

The leverage effect (negative correlation between current return and *future* conditional variance) is present when the parameter ω is negative. Suppose that $\lambda = 0$ and $\omega < 0$. Then a negative z_t will cause $f(z_t)$ to be positive thus increasing the future variance.

The volatility clustering is produced by the second term in f. Suppose that $\omega = 0$ and $\lambda > 0$. Whenever a shock occurs ($z_t > \mathrm{E}(z_t)$), $f(z_t)$ is positive hence increasing the future variance.

3.4.2.1 Moments

The unconditional mean of an EGARCH process is zero. This is straightforward to prove using the exact same approach as for the GARCH process.

Expressions for higher order moments are quite difficult to compute, and have thus been left out. The interested reader is referred to (Nelson 1991). It can be said, however, that if a moment of the distribution of z_t is zero, then so is the same order moment of the EGARCH process. Hence, all odd order moments of EGARCH are zero if z_t is Gaussian.

3.5 Alternative Arch models

Since ARCH models was first introduced in 1982, a vast number of models under the Arch framework have been proposed by an equally vast number of researchers.

It is outside the scope of this thesis to dive into all these models, but for the sake of integrity this section provides an overview.

3.5.1 IGARCH

The IGARCH model (see (Bollerslev et al. 1992)) is motivated by the fact that not all returns are best modelled by a stationary model. With the ARMA representation (3.3.16) of GARCH in mind, an ARIMA-style GARCH may be derived by imposing the IGARCH parameter restriction $\sum_{i=1}^{q} \alpha_i + \sum_{j=1}^{p} \beta_i = 1$ instead of the traditional GARCH parameter restriction $\sum_{i=1}^{q} \alpha_i + \sum_{j=1}^{p} \beta_i < 1$. The consequence of moving from GARCH to IGARCH is that the unconditional variance no longer exist (it is infinite) and that the process is no longer covariance stationary.

3.5.2 ARCH-M

The ARCH-M or ARCH-in-Mean model by (Engle, Lilien & Robins 1987) was proposed to model a phenomenon known as the *risk premia*; an assumption that investors are risk averse and thus demand excess profit in return for high risk. In statistical terms, the risk premia surfaces through a positive correlation between the current return and the conditional variance.

The ARCH-M model is derived from the Arch framework by setting

$$y_t = g(x_{t-i}, \sigma_t^2; b) + \varepsilon_t \tag{3.5.1}$$

in (3.1.3). A positive risk premia is obtained by letting the partial derivative of g with respect to σ_t^2 be positive. Despite its name, the ARCH-M model allows for GARCH specification of σ_t^2.

Typically, g is chosen as a linear or logarithmic function of σ_t^2 or σ_t, see (Bollerslev et al. 1992). The most common choice is $\sigma_t = \sqrt{h_t}$ like ARCH and GARCH, and $g = x_t^\top b + \delta\sqrt{h_t}$ where δ is a parameter determining the magnitude of the risk premia.

3.5.3 Non-normal conditional distributions

Although the unconditional distributions of ARCH and GARCH processes have fat tails compared to the normal distribution, some researchers complain that the

tails are not as fat as those of the empirical distributions. These researchers have suggested the use of alternative distributions for the z_t's in (3.1.1). The somewhat obvious choice when looking for fatter tails, the student-t distribution, was suggested as early 1987 in (Bollerslev 1987) who let the degrees of freedom be a parameter to be estimated. Another distribution, the generalized error distribution, was invented by (Nelson 1991) for this purpose. Both the generalized error distribution an the student-t distribution includes the normal distribution as a special case.

The alternative distribution approach are at the conceptual level no more difficult to deal with than the standardized normal distribution. However, the task of estimation is complicated by the fact that the above cited papers leave one or more distribution parameters to be estimated, thereby often increasing the total number of parameters to or beyond the capacity of the available estimation routines.

When introducing non-normal conditional densities it is important to keep (3.1.2) of Definition 3.1 in mind, i.e. to make sure that the innovations $\{\varepsilon_t\}$ have zero mean and σ_t^2 variance. For instance, the general t-distribution $t(n, \mu, \beta)$ with parameters n, μ and β has the distribution

$$f(x) = \frac{\Gamma\left(\frac{n+1}{2}\right)}{\Gamma\left(\frac{n}{2}\right)\beta\sqrt{n\pi}} \left[1 + \frac{(x-\mu)^2}{n\beta^2}\right]^{-\frac{n+1}{2}} \tag{3.5.2}$$

and for $n > 2$ the mean and variance is

$$\mathrm{E}[X] = \mu, \quad V[X] = \frac{n}{n-2}\beta^2. \tag{3.5.3}$$

Thus, to obtain zero mean and variance σ^2 one have to impose the parameter restrictions

$$\mu = 0, \quad \beta = \sigma\sqrt{\frac{n-2}{n}} \tag{3.5.4}$$

thereby casting the distribution as

$$\begin{aligned}
f(x) &= \frac{\Gamma\left(\frac{n+1}{2}\right)}{\Gamma\left(\frac{n}{2}\right)\sigma\sqrt{\frac{n-2}{n}}\sqrt{n\pi}} \left[1 + \frac{x^2}{n\frac{n-2}{n}\sigma^2}\right]^{-\frac{n+1}{2}} \\
&= \frac{\Gamma\left(\frac{n+1}{2}\right)}{\Gamma\left(\frac{n}{2}\right)\sigma\sqrt{n-2}\sqrt{\pi}} \left[1 + \frac{x^2}{(n-2)\sigma^2}\right]^{-\frac{n+1}{2}}.
\end{aligned} \tag{3.5.5}$$

The well-known student-t distribution with ν degrees of freedom is identical to the $t(\nu, 0, 1)$-distribution meaning that an error is almost inevitably introduced if student's t-distribution is used.

3.5.4 Multivariate models

In order to take interactions between financial returns of different sources into account, multivariate ARCH and GARCH models have been proposed. The reader is referred to (Engle, Granger & Kraft 1984), (Bollerslev et al. 1992) and (Bera & Higgins 1993) for more in-depth description of these models.

3.5.5 Other models

The list of Arch family members is long. Other published models are the Multiplicative GARCH model (MGARCH) by (Geweke & Pantula 1986) which specified the conditional variance as

$$\sigma_t = \exp\left(\frac{1}{2}h_t\right) \tag{3.5.6a}$$

$$h_t = \alpha_0 + \sum_{i=1}^{q} \alpha_i \ln\left(\varepsilon_{t-i}^2\right) + \sum_{j=1}^{p} \beta_j h_{t-j} \tag{3.5.6b}$$

proposed to allow unconstrained estimation of its parameters, the GJR-GARCH model by (Glosten, Jagannathan & Runkle 1993)

$$\sigma_t = \sqrt{h_t} \tag{3.5.7a}$$

$$h_t = \alpha_0 + \sum_{i=1}^{q} \alpha_i^{(1)} \varepsilon_{t-i}^2 + \sum_{i=1}^{q} \alpha_i^{(2)} \max(0, \Leftrightarrow \varepsilon_{t-i}) + \sum_{j=1}^{p} \beta_j h_{t-j} \tag{3.5.7b}$$

for $\alpha_0 > 0$, $\alpha_i^{(1)} \geq 0$, $\alpha_i^{(1)} + \alpha_i^{(2)} \geq 0$ and $\beta_j \geq 0$ and the NGARCH

$$\sigma_t = \sqrt{h_t} \tag{3.5.8a}$$

$$h_t = \alpha_0 + \sum_{i=1}^{q} \alpha_i h_{t-i}(z_{t-i} \Leftrightarrow c)^2 + \sum_{j=1}^{p} \beta_j h_{t-j} \tag{3.5.8b}$$

VGARCH

$$\sigma_t = \sqrt{h_t} \tag{3.5.9a}$$

$$h_t = \alpha_0 + \sum_{i=1}^{q} \alpha_i \left(\frac{\varepsilon_{t-i}}{\sqrt{h_{t-i}}} \Leftrightarrow c \right)^2 + \sum_{j=1}^{p} \beta_j h_{t-j} \tag{3.5.9b}$$

models where $\alpha_0 > 0$, $\alpha_i \geq 0$ and $\beta_j \geq 0$ by (Engle & Ng 1993) as alternative ways of incorporating the leverage effect and the Threshold GARCH (TGARCH) by (Zakoian 1994)

$$\sigma_t = \sqrt{h_t} \tag{3.5.10a}$$

$$\sqrt{h_t} = \alpha_0 + \sum_{i=1}^{q} \alpha_i^{(1)} |\varepsilon_{t-i}| + \sum_{i=1}^{q} \alpha_i^{(2)} \max(0, \Leftrightarrow \varepsilon_{t-i})$$

$$+ \sum_{j=1}^{p} \beta_j \sqrt{h_{t-j}} \tag{3.5.10b}$$

where $\alpha_0 > 0$, $\alpha_i^{(1)} \geq 0$, $\alpha_i^{(1)} + \alpha_i^{(2)} \geq 0$ and $\beta_j \geq 0$, to allow for different parameters depending on the sign of ε_{t-1}, just to name a few.

3.6   Augmented GARCH

The vast number of models under the Arch framework are despite their differences very similar. Thus, it would be suitable to encompass all the models into one unified model. Especially, when continuous time Arch models are to be derived from the discrete ones in a subsequent chapter, it would be desirable if the rigourious derivations are to be carried out on one model only.

Unfortunately, the Arch framework itself is not suitable as a unified model as the framework leaves the process $\{\sigma_t\}$ in (3.1.1) unspecified. As a suitable alternative, (Duan 1997) has proposed the *Augmented GARCH* process. The augmented GARCH(p, q) process is defined by

DEFINITION 3.5. The Augmented GARCH(p,q) process
An augmented GARCH(p,q) process is an Arch process with

$$\sigma_t = \sqrt{\tilde{h}_t}, \tag{3.6.1}$$

$$
\begin{aligned}
\phi_t \quad = \quad & \gamma_0 + \sum_{j=1}^{p} \gamma_1^{(j)} \phi_{t-j} \\
& + \sum_{i=1}^{q} \left[\gamma_2^{(i)} |z_{t-i} \Leftrightarrow c|^{\delta} + \gamma_3^{(i)} \max(0, c \Leftrightarrow z_{t-i})^{\delta} \right] \phi_{t-i} \\
& + \sum_{i=1}^{q} \left[\gamma_4^{(i)} f(|z_{t-i} \Leftrightarrow c|, \delta) + \gamma_5^{(i)} f(\max(0, c \Leftrightarrow z_{t-i}), \delta) \right]
\end{aligned}
\tag{3.6.2}
$$

and

$$
\tilde{h}_t = \begin{cases} |\xi \phi_t \Leftrightarrow \xi + 1|^{1/\xi} & \text{if} \quad \xi \neq 0, \\ \exp(\phi_t \Leftrightarrow 1) & \text{if} \quad \xi = 0 \end{cases}
\tag{3.6.3}
$$

where $f(s, \delta) = (s^{\delta} \Leftrightarrow 1)/\delta$ for any $s \geq 0$. ▲

Note that $\lim_{\delta \to 0} f(s, \delta) = \ln s$ so that f is actually a Box-Cox transformation. The purpose the transformation is allow the augmented GARCH process to include models like MGARCH and EGARCH (and models somewhere between EGARCH and GARCH) where the conditional variance is updated using logaritms of lagged residuals and/or variances.

Also note that $\tilde{h}_t$ is actually a continuous function of ξ as

$$\lim_{\xi \to 0} |\xi \phi_t \Leftrightarrow \xi + 1|^{1/\xi} = \exp(\phi_t \Leftrightarrow 1). \tag{3.6.4}$$

The function $\tilde{h}_t$ have been marked with a tilde to distinguish it from the h_t functions of ARCH, GARCH and EGARCH in cases where ambiguities would otherwise be present.

One of the attractive features of the augmented GARCH model is that it separates the specification of the time-series dynamic (ϕ) and the functional form of the transformation ($\tilde{h}_t$). The specification of $\tilde{h}_t$ gives rise to a continuum of models containing the aforementioned Arch models as special cases.

To verify that this is indeed the case, the choices of augmented GARCH parameters leading to the ARCH, GARCH and EGARCH models are given below.

3.6.1 GARCH

THEOREM 3.1.
A GARCH(P,Q) PROCESS IS AN AUGMENTED GARCH(P,Q) PROCESS.

Proof. Setting $\xi = 1$, $c = 0$, $\delta = 2$ and $\gamma_3^{(i)} = \gamma_4^{(i)} = \gamma_5^{(i)} = 0$, $i = 1, \ldots, q$ yields

$$
\begin{aligned}
\phi_t &= \gamma_0 + \sum_{j=1}^{p} \gamma_1^{(j)} \phi_{t-j} + \sum_{i=1}^{q} \gamma_2^{(i)} z_{t-i}^2 \phi_{t-i} \\
&= \gamma_0 + \sum_{j=1}^{p} \gamma_1^{(j)} \phi_{t-j} + \sum_{i=1}^{q} \gamma_2^{(i)} \varepsilon_{t-i}^2
\end{aligned}
\tag{3.6.5}
$$

and

$$
\tilde{h}_t = |\phi_t|.
\tag{3.6.6}
$$

Hence, the augmented GARCH(p,q) process is a GARCH(p,q) process or an I-GARCH(p, q) process when $\gamma_0 = \alpha_0$, $\gamma_1^{(j)} = \beta_j$, $\gamma_2^{(j)} = \alpha_j$ and $\gamma_0 > 0$, $\gamma_1^j \geq 0$ and $\gamma_2^{(i)} \geq 0$. The parameter restrictions ensures a non-negative ϕ_t meaning that h_t of (3.3.2) is identical to ϕ_t.

REMARK 3.1. The augmented GARCH(p,q) process is identical to an ARCH(q) process when $p = 0$ or $\gamma_1^{(j)} = 0$, $j = 1, \ldots, p$.   ▼

3.6.2 EGARCH

THEOREM 3.2.
AN EGARCH(P,Q) PROCESS IS AN AUGMENTED GARCH(P,Q) PROCESS.

Proof. The EGARCH(p,q) process is obtained by setting $\xi = 0$, $c = 0$, $\delta = 1$, $\gamma_2^{(i)} = \gamma_3^{(i)} = 0$, $i = 1,\ldots,q$. Incorporating these parameter restrictions into the augmented GARCH process yields

$$\phi_t = \gamma_0 + \sum_{j=1}^{p} \gamma_1^{(j)}\phi_{t-j} + \sum_{i-1}^{q} \left[\gamma_4^{(i)}|z_{t-i}| + \gamma_5^{(i)}\max(0, -z_{t-i}) \right] \qquad (3.6.7)$$

and

$$\tilde{h}_t = \exp(\phi_t - 1). \qquad (3.6.8)$$

The last term of (3.6.7) may be rewritten using the facts that

$$z_t = \max(0, z_t) - \max(0, -z_t) \qquad (3.6.9a)$$
$$|z_t| = \max(0, z_t) + \max(0, -z_t) \qquad (3.6.9b)$$

and thus by substracting (3.6.9) from (3.6.9)

$$|z_t| - z_t = 2\max(0, -z_t) \qquad (3.6.10)$$

yielding

$$
\begin{aligned}
\phi_t &= \gamma_0 + \sum_{j=1}^{p} \gamma_1^{(j)}\phi_{t-j} + \sum_{i=1}^{q} \left[\gamma_4^{(i)}|z_{t-i}| + \frac{1}{2}\gamma_5^{(i)}(|z_{t-i}| - z_{t-i}) \right] \\
&= \gamma_0 + \sum_{j=1}^{p} \gamma_1^{(j)}\phi_{t-j} \\
&\quad + \sum_{i=1}^{q} \left[\left(\gamma_4^{(i)} + \frac{1}{2}\gamma_5^{(i)} \right) |z_{t-i}| - \frac{1}{2}\gamma_5^{(i)} z_{t-i} \right]. \qquad (3.6.11)
\end{aligned}
$$

Comparing (3.6.8) and (3.6.1) to (3.4.2) it is clear that h_t of (3.4.1) is equal to $\phi_t \Leftrightarrow 1$. Inserting this into the expression above yields

$$
\begin{aligned}
h_t &= \phi_t \Leftrightarrow 1 \\
&= \gamma_0 \Leftrightarrow 1 + \sum_{j=1}^{p} \gamma_1^{(j)}(h_{t-j}+1) \\
&\quad + \sum_{i=1}^{q} \left[\left(\gamma_4^{(i)} + \frac{1}{2}\gamma_5^{(i)} \right) |z_{t-i}| \Leftrightarrow \frac{1}{2}\gamma_5^{(i)} z_{t-i} \right]
\end{aligned}
\tag{3.6.12}
$$

and if the terms are appropriately grouped the following expression is obtained

$$
\begin{aligned}
h_t &= \gamma_0 \Leftrightarrow 1 + \sum_{j=1}^{p} \gamma_1^{(j)} + \sum_{j=1}^{p} \gamma_1^{(j)} h_{t-j} \\
&\quad + \sum_{i=1}^{q} \left(\gamma_4^{(i)} + \frac{1}{2}\gamma_5^{(i)} \right) |z_{t-i}| \Leftrightarrow \sum_{i=1}^{q} \frac{1}{2}\gamma_5^{(i)} z_{t-i}.
\end{aligned}
\tag{3.6.13}
$$

Recalling the EGARCH specification of h_t from (3.4.2) and (3.4.3)

$$
h_t = \alpha_0 + \sum_{i=1}^{q} \alpha_i \left[\omega z_{t-i} + \lambda(|z_{t-i}| \Leftrightarrow \mathrm{E}|z_{t-i}|) \right] + \sum_{j=1}^{p} \beta_j h_{t-j}
\tag{3.6.14}
$$

which is grouped as

$$
\begin{aligned}
h_t &= \alpha_0 \Leftrightarrow \sum_{i=1}^{q} \alpha_i \lambda \mathrm{E}|z_{t-i}| + \sum_{j=1}^{p} \beta_j h_{t-j} + \sum_{i=1}^{q} \alpha_i \lambda |z_{t-i}| \\
&\quad + \sum_{i=1}^{q} \alpha_i \omega z_{t-i}
\end{aligned}
\tag{3.6.15}
$$

the augmented GARCH parameters may be obtained from the EGARCH parameters by relating (3.6.15) to (3.6.13) thus achieving the following expressions for

the augmented GARCH parameters:

$$\gamma_1^{(j)} = \beta_j, \qquad\qquad j = 1, \ldots, p$$

$$\gamma_5^{(i)} = \Leftrightarrow 2\alpha_i\omega, \qquad i = 1, \ldots, q$$

$$\gamma_4^{(i)} = \alpha_i\omega \Leftrightarrow \frac{1}{2}\gamma_5^{(i)}, \quad i = 1, \ldots, q \qquad (3.6.16)$$

$$\gamma_0 = 1 \Leftrightarrow \sum_{j=1}^{p} \gamma_1^{(j)} + \alpha_0 \Leftrightarrow \sum_{i=1}^{q} \alpha_i \lambda \mathrm{E}|z_{t-i}|$$

REMARK 3.2.
By inserting the expression for $\gamma_1^{(j)}$ into the expression for γ_0 and the expression for $\gamma_5^{(i)}$ into the expression for $\gamma_4^{(i)}$, γ_0 and $\gamma_4^{(i)}$ may expressed by

$$\gamma_0 = 1 + \alpha_0 \Leftrightarrow \sum_{i=1}^{q} \alpha_i \lambda \mathrm{E}|z_{t-i}| \Leftrightarrow sum_{j=1}^{p} \gamma_1^{(j)} \qquad (3.6.17a)$$

$$\gamma_4^{(i)} = 2\alpha_i\omega \qquad\qquad\qquad (3.6.17b)$$

▼

3.6.3 Properties

The augmented GARCH process should be looked upon merely as a simple way of working with the various specific Arch models at the same time, not as a model in itself.

The properties of the augmented GARCH are therefore only of interest, if they facilitate an easy way of deriving the properties of the specific Arch models. This is not the case when dealing with moments and autocorrelations, in which case it is much easier to work with the specific models, but when stationarity conditions for the various Arch processes are to be derived, the augmented GARCH is useful.

THEOREM 3.3.
STATIONARITY OF THE AUGMENTED GARCH(1,1) PROCESS

A sufficient condition for strict stationarity of the augmented GARCH(1,1) process is $\gamma_1 \geq 0$, $\gamma_2 \geq 0$ and $\gamma_2 + \gamma_3 \geq 0$ and

$$\gamma_1 + \gamma_2 \mathrm{E}\left[\left|z_t \Leftrightarrow c\right|^\delta\right] + \gamma_3 \mathrm{E}\left[\max(0, c \Leftrightarrow z_t)^\delta\right] \leq 1$$
$$\text{if } \gamma_2 > 0 \text{ or } \gamma_3 \neq 0$$

and

$$\gamma_1 < 1 \text{ if } \gamma_2 = \gamma_3 = 0 \tag{3.6.18}$$

Proof. See appendix of (Duan 1997).

The moments of the second condition of Theorem 3.3 will be computed in a later chapter. For now it suffices to mention that the theorem leads to the well known constraints $\alpha_1 \geq 0$, $\beta_1 \geq 0$ and $\alpha_1 + \beta_1 \leq 1$ for GARCH(1,1) / IGARCH(1,1) and to the constraint $0 \leq \beta_1 < 1$ for EGARCH(1,1).

3.7 Other stochastic volatility models

It should be noted that other discrete time stochastic volatility models than those encompassed by the augmented GARCH process exist. For instance, (Taylor 1986) have proposed the so-called Stochastic Variance model, a model that differs from models under augmented GARCH because they allow the conditional variance to be stochastic. That is, unlike for instance GARCH models, the variance of the return at time t is not known at time $t \Leftrightarrow 1$. The Stochastic Variance models have been less used than GARCH in applications because the parameters of the Stochastic Variance models are more difficult to estimate than GARCH parameters.

Chapter 4

Estimation of Arch model parameters

The previous chapter introduced a lot of models that all could be used to formally explain the stylized facts of stock return series. In this chapter, some of these models are fitted to the data, and an attempt is made to decide which model performs the best in explaining the return series characteristics captured in chapter 2.

Estimating Arch model parameters is a difficult task. The choice of estimation routine and the various parameter restrictions are just a few of the things that complicates the estimation procedure. At the same time, the literature on the subject is very sparse. Researchers in the field seem to be more interested in proposing new models than to take the existing models to work. This chapter takes a close look on Arch parameter estimation. The traditional Maximum Likelihood parameter estimation technique is compared to a recursive estimation routine. The pitfalls of both techniques are carefully described, and the estimation routines are evaluated using real life stock return series as well as simulated data series.

Section 4.1 explain how to decide upon the model orders, a decision that has to be made before the parameter estimation is commenced. Section 4.2 explains Maximum Likelihood estimation and Section 4.3 introduces a statistical test for heteroskedasticity close related to Maximum Likelihood. In Section 4.4 an alternative and recursive parameter estimation routine is introduced. Section 4.5 discusses how to validate the estimated models and Section 4.6 describes the estimation results and compares the two estimation techniques to one another. Section 4.7 concludes.

4.1 Choice of Arch model

When faced with the project of fitting an Arch model to data, the first choices to make are the choices of model class and model order. One has to decide if one wants to make use of an ARCH, a GARCH, an EGARCH or another model, and the model order parameters q and/or p have to be decided on.

4.1.1 Model class

When faced with the choice of model class, the first thing to do is to look at the empirical properties of the models at hand. The characteristics unique to specific Arch models were carefully described in chapter 2. If the data shows strong evidence of the leverage effect, one ought to consider the EGARCH model class, and if risk premia is present, the GARCH-M model may be suitable. If no model class specific characteristic seems to be present in the data, the choice is more difficult. In this situation my suggestion would be to choose a model class with relatively few parameters like GARCH, as least as the preliminary choice.

Seen from a theoretical point of view, one could decide upon the model class by estimating the parameters in the augmented GARCH model, and choosing the Arch model class corresponding to the set of non-zero augmented GARCH parameters. Unfortunately, this approach is unsuitable as the number of augmented GARCH parameters is too vast to allow for a reliable estimation.

The nested model approach *is*, however, suitable when it comes to ARCH and GARCH models. When in doubt which of these model classes to apply, one could start with the GARCH model class and then decide upon the ARCH class if the GARCH-specific parameters appears to be insignificant.

In general, it is not possible to decide the model class in advance, so the most fruitful approach is to estimate one model, valide it, estimate another model, validate it, and then compare the two model validations to decide which model class is the best.

4.1.2 Model order

When it comes to the choice of model order, the situation is not as difficult. The empirical studies in the field of Arch reveals that $p = q = 1$ is by far the most common model orders for GARCH and EGARCH. A few studies have reported

suitable GARCH(2,1) or GARCH(1,2) descriptions of data, but these choices of model order are rare. Higher model orders are *extremely* rare, and should it ever occur that presented data is best described by a high order model, the large number of parameters are so difficult to estimate that the resulting model would probably not provide a suitable description of the data anyway.

Statistical methods to establish the model order exist. As previously mentioned (section 3.3), the empirical autocorrelation function may be used to make preliminary assumptions regarding the model order, see (Bollerslev 1988) for a description of this approach. One should know, however, that these methods probably are not worth the effort. Firstly, as previously stated, because $p = q = 1$ is by far the most common choice, and second, because the statistical methods are not that precise anyway. Recall from Chapter 2 that the asymptotical properties of the autocorrelation function of squared Arch returns are unknown.

The lesson to learn from the above findings is that one should always start by estimating a model with $p = q = 1$ and only use higher order models if the (1,1) model is very poor in describing the data.

4.2   Maximum Likelihood Estimation

The log-likelihoods of various Arch models are easy to obtain from as the conditional distributions are specified by the model framework.

Given T observations, the log-likelihood for an Arch model with parameters θ is computed from the product of all the conditional densities of the ε's.

$$L_T(\theta) = \sum_{t=1}^{T} l_t(\theta) \tag{4.2.1}$$

where

$$l_t(\theta) = \ln f_z(\varepsilon_t | \sigma_t) \tag{4.2.2}$$

and f_z is the distribution of z_t from Equation 3.1.1.

If z_t is Gaussian,

$$f_z(\varepsilon_t | \sigma_t) = \frac{1}{\sqrt{2\pi}} \frac{1}{\sigma} \exp\left[\Leftrightarrow \frac{1}{2} \left(\frac{\varepsilon_t}{\sigma} \right)^2 \right] \tag{4.2.3}$$

and if the distribution of z_t is t with n degrees of freedom, $n > 2$,

$$f_z(\varepsilon_t|\sigma_t) = \frac{\Gamma\left(\frac{n+1}{2}\right)}{\Gamma\left(\frac{n}{2}\right)\sigma\sqrt{n-2}\sqrt{\pi}}\left[1+\frac{\varepsilon_t^2}{\sigma(n-2)}\right]^{-\frac{n+1}{2}}. \tag{4.2.4}$$

4.2.1 ARCH and GARCH

For models like ARCH and GARCH where $\sigma_t = \sqrt{h_t}$, the Gaussian case (4.2.3) is cast as

$$f_z(\varepsilon_t|\sigma_t) = \frac{1}{\sqrt{2\pi}}\frac{1}{\sqrt{h_t}}\exp\left[-\frac{1}{2}\frac{\varepsilon_t^2}{h_t}\right] \tag{4.2.5}$$

and thus, (4.2.2) appears as

$$l_t(\theta) = -\frac{1}{2}\ln(2\pi) - \frac{1}{2}\ln h_t - \frac{1}{2}\frac{\varepsilon_t^2}{h_t}. \tag{4.2.6}$$

If z_t is t-distributed, (4.2.2) is equal to this rather unpleasant expression

$$l_t(\theta) = -\frac{1}{2}\ln\pi + \ln\left[\Gamma\left(\frac{n+1}{2}\right)\right] - \ln\left[\Gamma\left(\frac{n}{2}\right)\right] - \frac{1}{2}\ln(n-2)$$
$$ - \frac{1}{2}\ln(h_t) - \frac{n+1}{2}\ln\left[1+\frac{\varepsilon_t^2}{h_t(n-2)}\right]. \tag{4.2.7}$$

Taking into account that the degrees of freedom is to be estimated, the terms have been grouped to display the constant term first.

4.2.1.1 Initial values

Maximum Likelihood estimation imposes a problem regarding the initial values of h_t. In order to maximize (4.2.1), one must be able to compute all values of h_t, $t = 1, \ldots, T$, but for a GARCH(p, q) process, h_t depends on $\varepsilon_{t-1}, \ldots, \varepsilon_{t-q}$ and $h_{t-1}, \ldots, h_{t-p}$. If the first observation occurs at time $t = 1$, values of $\varepsilon_0, \varepsilon_{-1}, \ldots, \varepsilon_{1-q}$ and $h_0, h_{-1}, \ldots, h_{1-p}$ have to be stipulated before the recursive computations of h_t can be fullfilled.

The numbers of values to be stipulated may be decreased by letting the first available observation arrive at time $t = 1 \Leftrightarrow q$. In this way, it is only the values of $h_0, h_{-1}, \ldots, h_{1-p}$, one has to come up with. The problem is still serious, however, and it is worsened by the fact that the initial values influates the calculation of the standardized residuals (to be described in Section 4.5) that are to be used to validate the estimation routines.

In his original GARCH paper, (Bollerslev 1986) suggests using an estimate of the unconditional variance as initial values, but he does not argue for this choice. Another approach would be to include the initial values in the Maximum Likelihood estimation. These approaches will both be used in the thesis, and the initial values influence on the parameter estimates will be carefully measured.

4.2.2 EGARCH

For a model like EGARCH, σ_t is equal to $\exp\left(\frac{1}{2} h_t\right)$ and when this is inserted into the Gaussian density

$$f_z(\varepsilon_t | \sigma_t) = \Leftrightarrow \frac{1}{\sqrt{2\pi}} \frac{1}{\sqrt{\sigma_t}} \exp\left[\Leftrightarrow \frac{1}{2} \frac{\varepsilon_t^2}{\sigma_t}\right], \tag{4.2.8}$$

Equation 4.2.2 is cast as

$$l_t(\theta) = \Leftrightarrow \frac{1}{2} \ln(2\pi) \Leftrightarrow \frac{1}{2} h_t \Leftrightarrow \frac{1}{2} \varepsilon_t^2 \exp(\Leftrightarrow h_t). \tag{4.2.9}$$

When z_t is t-distributed,

$$\begin{aligned} l_t(\theta) &= \Leftrightarrow \frac{1}{2} \ln \pi \Leftrightarrow \ln\left[\Gamma\left(\frac{n+1}{2}\right)\right] \Leftrightarrow \ln\left[\Gamma\left(\frac{n}{2}\right)\right] \Leftrightarrow \frac{1}{2} \ln(n \Leftrightarrow 2) \\ &\Leftrightarrow \frac{1}{2} h_t \Leftrightarrow \frac{n+1}{2} \ln\left[1 + \frac{\varepsilon_t^2}{n \Leftrightarrow 2} \exp\left(\Leftrightarrow \frac{1}{2} h_t\right)\right]. \end{aligned} \tag{4.2.10}$$

4.2.2.1 Initial values

The initial value problems mentioned in connection with ARCH and GARCH equally applies to EGARCH models.

4.2.3　Parameter uncertainty

A usually great advantage of Maximum Likelihood estimates is that a consistent estimate of the covariance matrix of the parameters may be obtained as the inverse of the information matrix. This is because the variance of a Maximum Likelihood estimator under certain conditions is known to be equal to the lower bound of Cramer-Rao's inequality.

The information matrix $i_T(\theta)$ may be computed using its definition

$$i_T(\theta) = \mathrm{E}\left[\frac{\partial L_T(\theta)}{\partial \theta}\frac{\partial L_T(\theta)}{\partial \theta^\top}\right] \tag{4.2.11}$$

or

$$i_T(\theta) = \Leftrightarrow\mathrm{E}\left[\frac{\partial^2 L_T(\theta)}{\partial\theta\partial\theta^\top}\right] \tag{4.2.12}$$

but the aforementioned problem of the initial values of the conditional variance has an even greater impact when derivatives of $L_T(\theta)$ is to be computed. Using (4.2.1) and (4.2.6) expressions of the derivatives are in the GARCH-Gaussian case

$$\frac{\partial L_T(\theta)}{\partial\theta} = \sum_{t=1}^{T}\frac{\partial l_t(\theta)}{\partial\theta} \tag{4.2.13}$$

and

$$
\begin{aligned}
\frac{\partial l_t(\theta)}{\partial\theta} &= \Leftrightarrow\frac{1}{2}h_t^{-1}\frac{\partial h_t}{\partial\theta} + \frac{1}{2}\varepsilon_t^2 h_t^{-2}\frac{\partial h_t}{\partial\theta}\\
&= \frac{1}{2}h_t^{-1}\frac{\partial h_t}{\partial\theta}\left(\frac{\varepsilon_t^2}{h_t}\Leftrightarrow 1\right)
\end{aligned}
\tag{4.2.14}
$$

and as $h_t = \alpha_0 + \sum_{i=1}^{q}\alpha_i\varepsilon_{t-i}^2 + \sum_{j=1}^{p}\beta_j h_{t-j}$

$$\frac{\partial h_t}{\partial\theta} = \begin{bmatrix} 1 \\ \varepsilon_{t-1}^2 \\ \vdots \\ \varepsilon_{t-q}^2 \\ h_{t-1} \\ \vdots \\ h_{t-p} \end{bmatrix} + \sum_{i=1}^{p}\beta_i\frac{\partial h_{t-i}}{\partial\theta}. \tag{4.2.15}$$

Hence, to compute the information matrix using (4.2.11), values of h_0, h_{-1}, ..., h_{1-p} as well as $\partial h_0/\partial\theta, \partial h_{-1}/\partial\theta, \ldots, \partial h_{1-p}/\partial\theta$ have to be stipulated, a total number of $2p$ values sort of taken out of the blue. As one can imagine, the problems are even worse if (4.2.12) is used.

The values of the partial derivatives, though, need not be specified if the derivatives of the log-likelihood are computed numerically. However, to estimate the expectations of (4.2.11) and (4.2.12) the derivatives ought to be computed for a vast number of realizations of the stochastic parts z_t, $t = 1, 2, \ldots, T$ of L_T. As this is definately intractable, a widely used approach is to estimate $i_T(\theta)$ by

$$i_T(\theta) \simeq H \qquad\qquad (4.2.16)$$

where H is the Hessian

$$H = \Leftrightarrow\frac{\partial^2 L_T}{\partial\theta\partial\theta^\top} \qquad\qquad (4.2.17)$$

computed using the obtained Maximum Likelihood estimate of θ or, equivalently, to approximate the expected value in (4.2.12) by the observed value of the second order derivative. Equation 4.2.12 is used as (4.2.11) will not do since the observed value of $\frac{\partial L_T}{\partial\theta}$ if zero when θ is maximizing L_T. It is emphasized that the approximation is only valid if no parameter constraints are active.

In this thesis the covariance matrix is estimated by the inverse of the observed Hessian computed numerically. Though the Hessian *may* be computed analytically, it will not be done here as the total number of unknown initial values of h, $\frac{\partial h}{\partial\theta}$ and $\frac{\partial^2 h}{\partial\theta^2}$ is so great that an analytical computed value of the Hessian is considered unreliable.

4.2.3.1   Numerical derivatives

In order to compute the above mentioned estimate of the Hessian, a numerical routine to approximate the second order derivatives is called for. General expressions for numerical approximations to second order derivatives are infered below.

Consider a function f of one variable x. The Taylor series expansion of $f(x_0 + u)$ about x_0 where x_0 and u are some real numbers is

$$f(x_0 + u) = f(x_0) + u\frac{\partial f}{\partial x}(x_0) + \frac{1}{2}u^2\frac{\partial^2 f}{\partial x^2}(x_0) + O(u^3). \qquad (4.2.18)$$

Equally, the expansion of $f(x - u)$ is

$$f(x_0 - u) = f(x_0) - u\frac{\partial f}{\partial x}(x_0) + \frac{1}{2}u^2\frac{\partial^2 f}{\partial x^2}(x_0) + O(u^3). \tag{4.2.19}$$

Adding the two expressions to one another yields

$$f(x_0 + u) + f(x_0 - u) = 2f(x_0) + u^2\frac{\partial^2 f}{\partial x^2}(x_0) + O(u^3) \tag{4.2.20}$$

from where the second derivative may be isolated and approximated by

$$\frac{\partial^2 f}{\partial x^2}(x_0) \simeq \frac{f(x_0 + u) + f(x_0 - u) - 2f(x_0)}{u^2} \tag{4.2.21}$$

and this expression is used to compute the diagonal elements of the Hessian. To compute the off-diagonal elements, functions of two variables have to be considered.

Let g be a function of two variables x and y. The Taylor series expansion of $g(x_0 + u, y_0 + u)$ about (x_0, y_0) is

$$g(x_0 + u, y_0 + u) =$$
$$g(x_0, y_0) + u\frac{\partial g}{\partial x}(x_0, y_0) + u\frac{\partial g}{\partial y}(x_0, y_0) + \frac{1}{2}u^2\frac{\partial^2 g}{\partial x^2}(x_0, y_0)$$
$$+ u^2\frac{\partial^2 g}{\partial x \partial y}(x_0, y_0) + \frac{1}{2}u^2\frac{\partial^2 g}{\partial y^2}(x_0, y_0) + O(u^3) \tag{4.2.22}$$

and equally,

$$g(x_0 + u, y_0 - u) =$$
$$g(x_0, y_0) + u\frac{\partial g}{\partial x}(x_0, y_0) - u\frac{\partial g}{\partial y}(x_0, y_0) + \frac{1}{2}u^2\frac{\partial^2 g}{\partial x^2}(x_0, y_0)$$
$$- u^2\frac{\partial^2 g}{\partial x \partial y}(x_0, y_0) + \frac{1}{2}u^2\frac{\partial^2 g}{\partial y^2}(x_0, y_0) + O(u^3) \tag{4.2.23}$$

and

$$g(x_0 - u, y_0 \pm u) =$$
$$g(x_0, y_0) - u\frac{\partial g}{\partial x}(x_0, y_0) \pm u\frac{\partial g}{\partial y}(x_0, y_0) + \frac{1}{2}u^2\frac{\partial^2 g}{\partial x^2}(x_0, y_0)$$
$$\mp u^2\frac{\partial^2 g}{\partial x \partial y}(x_0, y_0) + \frac{1}{2}u^2\frac{\partial^2 g}{\partial y^2}(x_0, y_0) + O(u^3). \qquad (4.2.24)$$

Substracting $g(x_0 + u, y_0 + u)$ from $g(x_0 + u, y_0 - u)$ yields

$$g(x_0 + u, y_0 + u) - g(x_0 + u, y_0 - u) =$$
$$2u\frac{\partial g}{\partial y}(x_0, y_0) + 2u^2\frac{\partial^2 g}{\partial x \partial y}(x_0, y_0) + O(u^3) \qquad (4.2.25)$$

and similarly,

$$g(x_0 - u, y_0 + u) - g(x_0 - u, y_0 - u) =$$
$$2u\frac{\partial g}{\partial y}(x_0, y_0) - 2u^2\frac{\partial^2 g}{\partial x \partial y}(x_0, y_0) + O(u^3). \qquad (4.2.26)$$

Hence, an approximation of $\frac{\partial^2 g}{\partial x \partial y}$ is readily available as

$$\frac{\partial^2 g}{\partial x \partial y}(x_0, y_0) \quad \simeq \quad (4u^2)^{-1}\{g(x_0 + u, y_0 + u) - g(x_0 + u, y_0 - u)$$
$$- [g(x_0 - u, y_0 + u) - g(x_0 - u, y_0 - u)]\}. \quad (4.2.27)$$

What is left is to decide upon the uptimal u. The solution is *not* to choose some small u without consideration. The binary approximation to small numbers used in todays computers may lead to spurious results in this regard, c.f. (Press, Teukolsky, Vetterling & Flannery 1992). Instead it is wise to use a value of

$$\varepsilon_M^{1/3} x_0 \qquad (4.2.28)$$

in the one-dimensional case, where ε_M is the relative floating point accuracy of the computer used to perform the approximation. In the two-dimensional case, a value of

$$\varepsilon_M^{1/3}\sqrt{x_0 y_0} \qquad (4.2.29)$$

is used.

4.2.4 Maximization routine

Applied Maximum Likelihood estimation includes a choice of maximization routine. This choice is of course unimportant as long as the chosen method is able to locate the global maximum of the likelihood function, but, as shall later be revealed, this is not always the case.

Conceptually, there is two different ways of maximizing the likelihood. One way is to use a root-finding algorithm to locate the set of parameters that makes the derived likelihood, the score, vanish. Another way is to use a maximization algorithm on the likelihood function, either a routine that makes use of an explicit expression of the derivative, or a routine that do not. Traditionally, a routine that uses an explicit derivate expression needs less computations than a routine that do not.

In his original GARCH paper, (Bollerslev 1986) suggests using the root-finding approach with the Berndt, Hall, Hall and Hausman algorithm which, according to Bollerslev, is described in (Berndt, Hall, Hall & Hausman 1974). Few other researchers deals with the subject, but the GARCH parameter estimation routine of the SAS system allegedly uses a maximization routine that makes use of an explicit expression of the derivative.

The choice of Bollerslev is disadvantious in the sence that it need an explicit expression of the derivative of the likelihood and, as described in Section 4.2.3, this derivate depends on unknown pre-sample values. It is tempting to assume that Bollerslev's choice is motivated by the fact that maximization routines which do not make use of explicit derivative expressions need a lot more computer power than routines which do. Evidently, the computer power available in 1986 was very small compared to today's resources.

In this thesis the so-called *downhill simplex method in multidimensions* due to Nelder and Mead is the numerical routine used. This routine do not use explicit derivate expressions and so the results produced using this routine are not influenced by poorly chosen initial derivative values. The routine is a minimization routine and as suct it is used on a sign-changed version of the log-likelihood.

Though it is evidently of interest to measure the Arch-related performance of various maximization routines, it is outside the scope of this thesis. Thus, the subject of maximization routines wil not be discussed any further.

4.2.4.1 Parameter restrictions

The parameter constraints that applies to a lot of the various Arch models are not easily combined with the maximization routine due to the way it traverse the parameter space. Routines to perform constrained maximization *do* exist, but they are often slow and thus unsuitable for the computer intensive Arch estimation.

Instead, the parameter constraints are assured using a Lagrange multiplier approach. The parameter constraints are multiplied by a large integer and substracted from the likelihood. For instance, when estimating GARCH(p,q) parameters to which the constraints (3.3.3) and (3.3.9) apply, the minimization routine is put to work on the function

$$
\begin{aligned}
\tilde{L}_T &= \sum_{t=1}^{T} \left(\ln h_t + \frac{\varepsilon_t^2}{h_t} \right) \Leftrightarrow 10^6 \min \left(0, 1 \Leftrightarrow \sum_{i=1}^{q} \alpha_j \Leftrightarrow \sum_{j=1}^{p} \beta_j \right) \\
&\Leftrightarrow 10^{12} \min \left(0, \alpha_1, \ldots, \alpha_q, \beta_1, \ldots, \beta_p \right).
\end{aligned}
\tag{4.2.30}
$$

c.f. (4.2.1) and (4.2.6).

4.3 Lagrange multiplier test

In the context of Maximum Likelihood estimation, a statistical test for heteroskedasticity generally known as *Engle's Likelihood Multiplier Test* deserves a few words. The test appears to be widely used in situations were hetoroskedasticity is expected. It has been deferred to this section because some knowledge of Maximum Likelihood estimation of ARCH and GARCH models is needed in order to understand the test principle.

The test was first described in (Engle 1982) in connection with the first description of the ARCH model, but Engles description is a bit short. Because of this, a more in-depth review is provided below.

Engles test was motivated by the need to test for presence of ARCH effects, i.e. autocorrelations in the squared observations, before actually estimating the order and the parameters of the ARCH model. As we shall see below, the test has other more general applications, but the original focus is important to understand the test principle. Using the Arch framework

$$
\varepsilon_t = z_t \sqrt{h_t}
\tag{4.3.1}
$$

and z_t Gaussian, Engle used a generalization of the ARCH process by specifying h_t as

$$h_t = h(\alpha_0 + \alpha_1 \varepsilon_{t-1} + \cdots + \alpha_q \varepsilon_{t-q}) \qquad (4.3.2)$$

where h is some differentiable function. The reason for generalizing h_t will become clear later, for now it suffices to observe that (4.3.2) contains the ARCH case (Definition 3.2) as a special case.

Engle set out to test the hypothesis

$$H_0 : \alpha_1 = \cdots = \alpha_q = 0 \qquad (4.3.3)$$

or, equivalent $h_t = h(\alpha_0) = h^0$, against the general alternative

$$H_1 : \exists i : 1 \leq i \leq q \wedge \alpha_i \neq 0. \qquad (4.3.4)$$

4.3.1 Test principle

In general, whether applied to a particular field like Arch or not, a Lagrange multiplier test uses the Lagrange multiplier principle known from operations research, c.f. (Kotz 1983). In operations research, the maximum of a function $f(x)$ subject to the constraints $g(x) = 0$ may be found as the unconstrained maximum of the *Lagrangian*

$$\mathcal{L}(x, \lambda) = f(x) \Leftrightarrow \lambda^\top g(x) \qquad (4.3.5)$$

where the vector λ contains the so-called Lagrange-multipliers.

Taking (4.3.5) to statistics it may be used in situations where the maximization of the likelihood function of a statistical model is subject to constaints on the model parameters. Replacing f of (4.3.5) by the log-likelihood, a test for the hypothesis

$$H_0 : g(\theta) = 0 \quad \text{versus} \quad H_1 : g(\theta) \text{ unrestricted} \qquad (4.3.6)$$

where θ is the model parameters, may be constructed as a test that rejects H_0 if λ is too large in some sence. This has been done by (Breusch & Pagan 1979) using

the score and the information of the constrained model. Letting L be the likelihood of the model, defining

$$d = \frac{\partial \ln L}{\partial \theta} \qquad (4.3.7)$$

and

$$I = -\mathrm{E} \left[\frac{\partial^2 \ln L}{\partial \theta \partial \theta^{\top}} \right] \qquad (4.3.8)$$

using the *unconstrained* model and evaluating these quantities at the maximum likelihood estimator $\hat{\theta}$ of the *constrained* model as $\hat{d}$ and $\hat{S}$, the test statistic

$$LM = \hat{d}^{\top} \hat{I}^{-1} \hat{d} \qquad (4.3.9)$$

is asymptotically $\chi^2(s-r)$ distributed under H_0, where s is the dimension of θ, i.e. the number of parameters and r is the number of independent constraints $g_i(\theta) = 0$, $i = 1, \ldots, r$. Note that, as a great advantage in applications, the test statistic is calculated using nothing but estimates of the null model.

4.3.2 Heteroskedasticity

Returning to Engle, the parameters of the unconstrained model are

$$\theta^{\top} = \begin{bmatrix} \alpha_0 & \alpha_1 \cdots \alpha_q \end{bmatrix} \qquad (4.3.10)$$

and the likelihood is

$$L_T(\theta) = \sum_{t=1}^{T} l_t(\theta) \qquad (4.3.11)$$

so that

$$d = \frac{\partial \ln L_T}{\partial \theta} = \sum_{t=1}^{T} \frac{\partial l_t(\theta)}{\partial \theta} \qquad (4.3.12)$$

and by (4.2.14)

$$\frac{\partial l_t(\theta)}{\partial \theta} = \frac{1}{2} h_t^{-1} \frac{\partial h_t}{\partial \theta} \left(\frac{\varepsilon_t^2}{h_t} \Leftrightarrow 1 \right).$$

(4.3.13)

If the vectors u_t, $t = 0, 1, \ldots, T$ are introduced by

$$u_t^\top = \begin{bmatrix} 1 & \varepsilon_{t-1}^2 & \cdots & \varepsilon_{t-p}^2 \end{bmatrix}$$

(4.3.14)

then h_t may be written $h_t = h\left(u_t^\top \theta \right)$ and then

$$\frac{\partial h_t}{\partial \theta} = u_t \left. \frac{\partial h}{\partial \theta} \right|_{u_t^\top \theta}$$

(4.3.15)

Note that as h_t is a scalar and θ is a $(p + 1 \times 1)$ column vector, h_t derived by θ must also be a $(p + 1 \times 1)$ column vector. Because of this, the last part of (4.3.15) must be a scalar and if this scalar is denoted by h' then (4.3.13) may be written as

$$\frac{\partial l_t(\theta)}{\partial \theta} = \frac{h'}{2h_t} u_t \left(\frac{\varepsilon_t^2}{h_t} \Leftrightarrow 1 \right).$$

(4.3.16)

Under the null we have $h_t = h(\alpha_0) = h^0$ and $h' = h^{0\prime}$ which by insertion yields

$$\frac{\partial l_t(\theta)}{\partial \theta} = \frac{h^{0\prime}}{2h^0} u_t \left(\frac{\varepsilon_t^2}{h^0} \Leftrightarrow 1 \right)$$

(4.3.17)

and hence, the score is

$$\begin{aligned} \hat{d} &= \sum_{t=1}^T \frac{h^{0\prime}}{2h^0} u_t \left(\frac{\varepsilon_t^2}{h^0} \Leftrightarrow 1 \right) \\ &= \frac{h^{0\prime}}{2h^0} U^\top f^0 \end{aligned}$$

(4.3.18)

where U is the $(p + 1 \times T)$-matrix defined by $U^\top = \begin{bmatrix} u_1 & u_2 & \cdots & u_T \end{bmatrix}$ and f^0 is the $(T \times 1)$ column vector whose element number t is

$$\frac{\varepsilon_t^2}{h^0} \Leftrightarrow 1.$$

(4.3.19)

The information matrix I of the unconstrained model is

$$I = \Leftrightarrow \mathrm{E}\left[\frac{\partial^2 \ln L_T}{\partial\theta\partial\theta^\top}\right] = \Leftrightarrow \sum_{t=1}^{T} \mathrm{E}\left[\frac{\partial^2 l_t}{\partial\theta\partial\theta^\top}\right] \tag{4.3.20}$$

and using (4.3.13) it is straightforward to show that

$$\frac{\partial^2 l_t}{\partial\theta\partial\theta^\top} = \frac{1}{2h_t}\frac{\partial h_t}{\partial\theta}\left(\Leftrightarrow\varepsilon_t^2\frac{\partial h_t}{h_t^2\,\partial\theta}\right) + \frac{\partial\left[\frac{1}{2h_t}\frac{\partial h_t}{\partial\theta}\right]}{\partial\theta}\left(\frac{\varepsilon_t^2}{h_t}\Leftrightarrow 1\right) \tag{4.3.21}$$

When taking the expected value with respect to the parameters of this expression it should be noted that since the mean of ε_t^2 is h_t, the last term of (4.3.21) vanishes and the last part of the first term is reduced. Hence,

$$I = \sum_{t=1}^{T} \mathrm{E}\left[\frac{1}{2h_t^2}\frac{\partial h_t}{\partial\theta}\frac{\partial h_t}{\partial\theta^\top}\right] = \frac{1}{2}\sum_{t=1}^{T} \mathrm{E}\left[\frac{h_t'^2}{h_t^2}u_t^\top u_t\right]. \tag{4.3.22}$$

Under the null hypothesis this is cast as

$$\hat{I} = \frac{1}{2}\sum_{t=1}^{T} \mathrm{E}\left[\frac{h^{0\prime 2}}{h^{02}}u_t^\top u_t\right] = \frac{1}{2}\left(\frac{h^{0\prime}}{h^0}\right)^2 U^\top U \tag{4.3.23}$$

where $\mathrm{E}\left[U^\top U\right]$ have been replaced by $U^\top U$ as U does not depend on the parameters.

Inserting the expressions of $\hat{d}$ and $\hat{I}$ into the expression of the Lagrange multiplier test statistic (4.3.9) yields

$$\begin{aligned} LM &= \hat{d}^\top \hat{I}^{-1} \hat{d} \\ &= \frac{h^{0\prime}}{2h^0}f^{0\top}U\left[\frac{1}{2}\left(\frac{h^{0\prime}}{h^0}\right)^2 U^\top U\right]^{-1}\frac{h^{0\prime}}{2h^0}U^\top f^0 \\ &= \frac{1}{2}f^{0\top}U\left[U^\top U\right]^{-1}U^\top f^0. \end{aligned} \tag{4.3.24}$$

The unconditional model has $q+1$ parameters and one constraint is active under H_0. Thus, $s = q+1$ and $r = 1$ and LM is asymptotically $\chi^2(q)$-distributed under H_0.

As pointed out by (Engle 1982), all references to h' has disappeared and thus the test is the same for any h which is only a differentiable function of $u_t\theta$. This fact is at the same time the motivation for considering the general setup (4.3.2) instead of just the ARCH(q) model. Had the alternative H_1 only consisted of the ARCH(q) case, the use of the test would be limited to just testing for ARCH. The general setup includes most of the heteroskedastic models in the literature, c.f. (Breusch & Pagan 1979), and thus it can with more right be claimed that the test is actually a test for heteroskedasticity.

Though the test statistic (4.3.24) is easy to compute compared to the task of esti-mating $q+1$ ARCH parameters, an estimate of h^0 *is* needed in order to compute the elements of f^0. This estimation may be avoided by considering the elements of f^0. $\varepsilon_t/\sqrt{h^0}$ is by assumption $N(0, 1)$-distributed under the null and so ε_t^2/h^0 follows a $\chi^2(1)$-distribution. Knowing that the mean of a $\chi^2(1)$-distributed variable is 1, the maximum likelihood estimate of the variance of the same $\chi^2(1)$-distribution is

$$\frac{1}{T}\sum_{t=1}^{T}\left(\frac{\varepsilon_t^2}{h^0}\Leftrightarrow 1\right)^2 = \frac{1}{T}f^{0\top}f^0. \tag{4.3.25}$$

The variance of a χ^2-distribution is known to be equal to 2 and hence, the quantity

$$\frac{2T}{f^{0\top}f^0} \tag{4.3.26}$$

must converge to 1 in probability. Thus, a test statistic asymptotically equivalent to LM would be

$$LM' = T\frac{f^{0\top}U\left[U^{\top}U\right]^{-1}U^{\top}f^0}{f^{0\top}f^0}. \tag{4.3.27}$$

Focusing on the nominator, the term $\left[U^{\top}U\right]^{-1}U^{\top}f^0$ is the OLS estimate of the parameters of the regression of f^0 on U. Because of this, the fraction must be identical to the squared multiple correlation coefficient of the regression of f^0 on U. As adding a constant (one in this case) to the elements of f^0 and multiplying by a scalar (h^0) does not change the size of the multiple correlation coefficient, it is possible to write

$$LM' = TR^2 \tag{4.3.28}$$

where R^2 is the squared multiple correlation coefficient of the regression of ε_t^2 on $\begin{bmatrix} 1 & \varepsilon_{t-1}^2 & \cdots & \varepsilon_{t-q}^2 \end{bmatrix}$.

As the two test statistics (LM and LM') are asymptotically equivalent, LM' is also asymptotically $\chi^2(q)$-distributed under H_0. In this thesis, the LM' statistic will be the one used.

4.3.3 GARCH

The framework defined by (4.3.1) and (4.3.2) is formally not general enough to include GARCH models. This is due to the fact that GARCH models allow h_t to depend on lagged values of itself, a dependency prohibited by (4.3.2). The test is, however, also able to detect the heteroskedasticity of GARCH data, because a GARCH(p,q) model may be interpreted as an infinite order ARCH process, c.f. Section 3.3.

4.3.4 Test results

When the Lagrange multiplier test for heteroskedasticity is applied to the four well-known return series, the test results displayed in Table 4.1 are obtained. Test statistics significant at the 5%-level are written in boldface.

The Lagrange multiplier tests indicate that heteroskedasticity is present in all of the four series. In the case of Hewlett-Packard, H_0 is rejected when the H_1 contains ε's with lags greater than 2. For Sony and Mobil, H_0 is rejected for lags greater than 1. All test results, including the somewhat peculiar results of pepsi, are in line with the Ljung-Box test results (Table 2.5) of Section 2.2.2.3.

4.3.5 Application to augmented GARCH

Because the Likelihood Multiplier test statistic is computed using only parameter estimates obtained from the model under H_0, it may be used in connection with the augmented GARCH process. Specific Arch models may be formulated as parameter constraints imposed on the augmented GARCH model and the Likelihood Multiplier Test may then be used to find out which particular Arch model that best fits the data at hand. This has been done in (Duan 1997), but the approach will not be pursued here because this chapter only deals with GARCH and EGARCH models and though the test statistic only needs estimates of the null model, expres-

Series	HP		Sony		Mobil		Pepsi	
LM	T	Prob	T	Prob	T	Prob	T	Prob
1	2.163	0.141	3.614	0.057	0.515	0.473	3.973	**0.046**
2	2.185	0.335	8.765	**0.012**	11.600	**0.003**	4.485	0.106
3	17.237	**0.001**	9.025	**0.029**	11.607	**0.009**	5.902	0.116
4	18.941	**0.001**	9.666	**0.046**	11.787	**0.019**	6.207	0.184
5	19.811	**0.001**	17.885	**0.003**	16.162	**0.006**	6.206	0.287
6	21.442	**0.002**	20.232	**0.003**	18.322	**0.005**	6.204	0.401
7	22.097	**0.002**	28.439	**0.000**	18.248	**0.011**	7.143	0.414
8	23.170	**0.003**	30.938	**0.000**	22.977	**0.003**	10.087	0.259
9	28.466	**0.001**	30.864	**0.000**	25.738	**0.002**	10.490	0.312
10	28.454	**0.002**	32.392	**0.000**	29.717	**0.001**	11.318	0.333
11	30.323	**0.001**	32.539	**0.001**	34.118	**0.000**	11.331	0.416
12	30.281	**0.003**	32.696	**0.001**	33.256	**0.001**	11.448	0.491
13	30.615	**0.004**	32.747	**0.002**	36.838	**0.000**	11.470	0.571
14	30.632	**0.006**	35.065	**0.001**	45.191	**0.000**	11.479	0.648
15	30.795	**0.009**	36.169	**0.002**	48.166	**0.000**	11.816	0.693

Table 4.1: The test statistics and associated probabilities of the Lagrange multiplier test for heteroskedasticity for orders 1 to 15.

sions of the score and information of the alternative model still have to be derived, a cumbersome task in the case of augmented GARCH.

4.4   Recursive parameter estimation

The difficulties associated with Maximum Likelihood estimation, in particular the problems of the initial values and the dependency on a correct specification of the conditional distribution, motivates a search for estimation procedures less sensitive in these areas.

The search for such procedures has lead to the derivation of recursive versions of the families of estimation methods known from time series analysis as *Prediction Error Methods* (PEM) and *Pseudolinear Regressions* (PLR). The corresponding recursive versions are denoted RPEM and RPLR, respectively.

In this section, the two families RPEM and RPLR of recursive estimation methods are derived. As the usability of the recursive estimation methods is not limited to

Arch models, the Arch notation used in the previous chapters will be abandoned for a while.

Among famous members of the two families are *Recursive Maximum Likelihood* for ARMAX models (member of RPEM), Extended Least Squares for ARMAX models (member of RPLR) and *Recursive Least Squares* for ARX models (member of both families). See eq. (Ljung 1987) or (Ljung & Söderström 1983).

4.4.1 Nonrecursive versions

Consider the pseudolinear model

$$\hat{Y}_{t|t-1} = X_t^\top(\hat{\theta})\hat{\theta}, \tag{4.4.1}$$

where the regressor X is allowed to depend on the parameter vector θ and the prediction error ϵ_t is

$$\epsilon_t = Y_t \Leftrightarrow \hat{Y}_{t|t-1}. \tag{4.4.2}$$

The model (4.4.1) may be static or dynamic, i.e. the regressor X *may* contain lagged values of Y or ϵ.

4.4.1.1 Prediction errors

Intuitively, we are interested in estimation methods that generate prediction errors that are as small as possible in some sense. We would also like our methods to extract as much information from the observations as possible, that is, ideally the prediction errors should be uncorrelated with past observations. As these two criteria do not generally coincide, one has to prefer one from the other. This preference leads to either the PEM methods or the PLR methods.

A PEM seeks to minimize the sum of the squared prediction errors, i.e.

$$V_T = \frac{1}{T}\sum_{t=1}^{T}\frac{1}{2}\epsilon_t^2(\theta). \tag{4.4.3}$$

is minimized with respect to θ. See (Ljung 1987, Chapter 7).

A PLR method seeks to compute its estimates such that the prediction errors are uncorrelated with past data. As this is impossible to assure in practical applications, one may confine oneself to demand the prediction errors to be uncorrelated with the regressors X_t. This makes sense here as $X_t(\theta)$ is a function of relevant past observations and θ.

Formally, a PLR estimate is obtained as the solution to

$$\frac{1}{T} \sum_{t=1}^{T} X_t^{\top}(\theta)\epsilon_t(\theta) = 0. \tag{4.4.4}$$

4.4.2 Prediction Error Methods (PEM)

As it is not possible to minimize the criterion (4.4.3) analytically, a numerical procedure is called for. V_T is minimized by solving

$$\frac{\partial V_T}{\partial \theta} = 0, \tag{4.4.5}$$

a task suitable for the Newton-Raphson procedure. The first derivative with respect to θ is

$$\frac{\partial V_T}{\partial \theta} = \frac{1}{T} \sum_{t=1}^{T} \epsilon_t(\theta) \frac{\partial \epsilon_t(\theta)}{\partial \theta}. \tag{4.4.6}$$

By defining ψ by

$$\psi_t(\theta) = \frac{\partial \hat{Y}_t}{\partial \theta} = \Leftrightarrow \frac{\partial \epsilon_t(\theta)}{\partial \theta}, \tag{4.4.7}$$

Equation 4.4.6 can be rewritten as

$$\frac{\partial V_T}{\partial \theta} = \Leftrightarrow \frac{1}{T} \sum_{t=1}^{T} \epsilon_t(\theta)\psi_t(\theta). \tag{4.4.8}$$

The second derivative (the Hessian) is

$$\frac{\partial^2 V_T}{\partial \theta^2} = \frac{1}{T} \sum_{t=1}^{T} \psi_t(\theta)\psi_t^{\top}(\theta) \Leftrightarrow \frac{1}{T} \sum_{t=1}^{T} \epsilon_t(\theta) \frac{\partial \psi_t(\theta)}{\partial \theta}. \tag{4.4.9}$$

It may be quite difficult to compute the second derivative of ϵ with respect to θ as needed in the last term of the Hessian. This is, however, not necessary as the value of the last term is close to zero when θ is close to the real parameters, and an accurate value of the Hessian is only needed in the proximity of the minimum, see (Ljung 1987, Chapter 10). Thus the last term of (4.4.9) may be discarded and the parameter estimate can be computed iteratively, using the quasi Newton-Raphson algorithm

$$
\begin{aligned}
\hat{\theta}_i &= \hat{\theta}_{i-1} + \left(\frac{\partial^2 V_T}{\partial \theta_{i-1}^2} \right)^{-1} \frac{\partial V_T}{\partial \theta_{i-1}} \tag{4.4.10a} \\
&\simeq \hat{\theta}_{i-1} + \left(\sum_{t=1}^{T} \psi_t(\hat{\theta}_{i-1}) \psi_t^\top(\hat{\theta}_{i-1}) \right)^{-1} \sum_{t=1}^{T} \epsilon_t(\hat{\theta}_{i-1}) \psi_t(\hat{\theta}_{i-1}) \\
&\tag{4.4.10b}
\end{aligned}
$$

where i denotes the iteration number.

4.4.2.1 Recursive Prediction Error Methods (RPEM)

In the following a recursive version of PEM is derived, i.e. an estimation method that can be used to compute a new parameter estimate each time a new observation arrives without the need of a laborious recalculation of the summations in (4.4.10b).

To obtain such a version, consider again Equation 4.4.10b. Computing $\hat{\theta}_T$ using $\hat{\theta}_{T-1}$ and a single iteration yields

$$
\hat{\theta}_T = \hat{\theta}_{T-1} + \left(\sum_{t=1}^{T} \psi_t(\hat{\theta}_{T-1}) \psi_t^\top(\hat{\theta}_{T-1}) \right)^{-1} \sum_{t=1}^{T} \epsilon_t(\hat{\theta}_{T-1}) \psi_t(\hat{\theta}_{T-1}). \tag{4.4.11}
$$

In doing this, it is assumed that only one iteration is needed for $\hat{\theta}_T$ to solve (4.4.5). Nothing precludes the use of more than one iteration per observation, but the improvement in the accuracy of the estimate is insignificant in practical applications.

The last summation in (4.4.11) may be written as

$$\sum_{t=1}^{T} \epsilon_t(\hat{\theta}_{T-1})\psi_t(\hat{\theta}_{T-1})$$

$$= \epsilon_T(\hat{\theta}_{T-1})\psi_T(\hat{\theta}_{T-1}) + \sum_{t=1}^{T-1} \epsilon_t(\hat{\theta}_{T-1})\psi_t(\hat{\theta}_{T-1})$$

$$= \epsilon_T(\hat{\theta}_{T-1})\psi_T(\hat{\theta}_{T-1}) \Leftrightarrow T \left. \frac{\partial V_{T-1}}{\partial \theta} \right|_{\hat{\theta}_{T-1}}, \qquad (4.4.12)$$

where (4.4.8) is used. Assuming that $\hat{\theta}_{T-1}$ minimizes V_{T-1}, the last term is zero.

The first summation in (4.4.11) can be written as

$$\sum_{t=1}^{T} \psi_t(\hat{\theta}_{T-1})\psi_t^{\top}(\hat{\theta}_{T-1})$$

$$= \psi_T(\hat{\theta}_{T-1})\psi_T^{\top}(\hat{\theta}_{T-1}) + \sum_{t=1}^{T-1} \psi_t(\hat{\theta}_{T-1})\psi_t^{\top}(\hat{\theta}_{T-1}), \qquad (4.4.13)$$

but this does not reduce the computations as both terms have to be recalculated each time $\hat{\theta}$ is updated. However, if we decline to update the last summation in (4.4.13) and if we introduce

$$R_T(\hat{\theta}_{T-1}) = \sum_{t=1}^{T} \psi_t(\hat{\theta}_{T-1})\psi_t^{\top}(\hat{\theta}_{T-1}), \qquad (4.4.14)$$

Equation 4.4.11 can be rewritten as the RPEM updating scheme:

DEFINITION 4.1 (THE RECURSIVE PREDICTION ERROR SCHEME).
Let Y_T be defined by (4.4.1), ψ_T by (4.4.7) and R_T by (4.4.14). Then the RPEM updating scheme is defined by

$$\epsilon_T(\hat{\theta}_{T-1}) \quad = \quad Y_T \Leftrightarrow X_T(\hat{\theta}_{T-1})\hat{\theta}_{T-1} \qquad (4.4.15a)$$

$$\hat{\theta}_T \quad = \quad \hat{\theta}_{T-1} + R_T^{-1}(\hat{\theta}_{T-1})\epsilon_T(\hat{\theta}_{T-1})\psi_T(\hat{\theta}_{T-1}) \qquad (4.4.15b)$$

$$R_T(\hat{\theta}_{T-1}) \quad = \quad R_{T-1}(\hat{\theta}_{T-1}) + \psi_T(\hat{\theta}_{T-1})\psi_T^{\top}(\hat{\theta}_{T-1}) \qquad (4.4.15c)$$

▲

For a description of the convergence properties see (Ljung & Söderström 1983).

Note that the expression R_T^{-1} in (4.4.15b) may be computed without the use of explicit matrix inversion. For a description of the Matrix Inversion Lemma see (Ljung 1987, Chapter 11).

4.4.3 Pseudolinear Regressions (PLR)

In order to develop an iterative procedure for the family of pseudolinear regression methods, the left hand side of Equation 4.4.4 is rewritten as

$$
\begin{aligned}
\frac{1}{T}\sum_{t=1}^{T} X_t(\theta)\epsilon_t(\theta) &= \frac{1}{T}\sum_{t=1}^{T} X_t(\theta)\left(Y_t \Leftrightarrow X_t(\theta)\theta\right) \\
&= \frac{1}{T}\sum_{t=1}^{T} X_t(\theta)Y_t \Leftrightarrow \left(\frac{1}{T}\sum_{t=1}^{T} X_t(\theta)X_t^{\top}(\theta)\right)\theta,
\end{aligned}
$$

(4.4.16)

and hence

$$
\hat{\theta} = \left(\sum_{t=1}^{T} X_t(\theta)X_t^{\top}(\theta)\right)^{-1}\sum_{t=1}^{T} X_t(\theta)Y_t.
$$

(4.4.17)

Note that if X is independent of θ, this estimator is identical to the ordinary least squares estimator.

Equation 4.4.17 may be used to find an estimate for θ by iteration

$$
\hat{\theta}_i = \left(\sum_{t=1}^{T} X_t(\hat{\theta}_{i-1})X_t^{\top}(\hat{\theta}_{i-1})\right)^{-1}\sum_{t=1}^{T} X_t(\hat{\theta}_{i-1})Y_t.
$$

(4.4.18)

Using $Y_t = X_t^\top(\hat{\theta}_{i-1})\hat{\theta}_{i-1} + \epsilon_t(\hat{\theta}_{i-1})$ the last factor of (4.4.17) turns into

$$\sum_{t=1}^{T} X_t(\hat{\theta}_{i-1})Y_t$$

$$= \sum_{t=1}^{T} X_t(\hat{\theta}_{i-1})\epsilon_t(\hat{\theta}_{i-1}) + \sum_{t=1}^{T} X_t(\hat{\theta}_{i-1})X_t^\top(\hat{\theta}_{i-1})\hat{\theta}_{i-1}. \qquad (4.4.19)$$

Inserting this into (4.4.18) yields

$$\hat{\theta}_i = \hat{\theta}_{i-1} + \left(\sum_{t=1}^{T} X_t(\hat{\theta}_{i-1})X_t^\top(\hat{\theta}_{i-1})\right)^{-1} \sum_{t=1}^{T} X_t(\hat{\theta}_{i-1})\epsilon_t(\hat{\theta}_{i-1}). \qquad (4.4.20)$$

4.4.3.1   Recursive Pseudolinear Regressions (RPLR)

To obtain a recursive version of PLR a new observation is introduced each time an iteration is performed in (4.4.20), that is

$$\hat{\theta}_T = \hat{\theta}_{T-1} + \left(\sum_{t=1}^{T} X_t(\hat{\theta}_{T-1})X_t^\top(\hat{\theta}_{T-1})\right)^{-1} \sum_{t=1}^{T} X_t(\hat{\theta}_{T-1})\epsilon_t(\hat{\theta}_{T-1}). \qquad (4.4.21)$$

In line with the derivation of the RPEM method we split the last of the two summations into

$$\sum_{t=1}^{T} X_t(\hat{\theta}_{T-1})\epsilon_t(\hat{\theta}_{T-1})$$

$$= X_T(\hat{\theta}_{T-1})\epsilon_T(\hat{\theta}_{T-1}) + \sum_{t=1}^{T-1} X_t(\hat{\theta}_{T-1})\epsilon_t(\hat{\theta}_{T-1}), \qquad (4.4.22)$$

and note that the last term is zero assuming that $\hat{\theta}_{T-1}$ solves (4.4.4) when T is substituted by $T \Leftrightarrow 1$. Similarly for Eq. 4.4.21

$$\sum_{t=1}^{T} X_t(\hat{\theta}_{T-1})X_t^\top(\hat{\theta}_{T-1})$$

$$= X_T(\hat{\theta}_{T-1})X_T^\top(\hat{\theta}_{T-1}) + \sum_{t=1}^{T-1} X_t(\hat{\theta}_{T-1})X_t^\top(\hat{\theta}_{T-1}). \qquad (4.4.23)$$

Introducing

$$R_T(\hat{\theta}_{T-1}) = \sum_{t=1}^{T} X_t(\hat{\theta}_{T-1}) X_t^\top(\hat{\theta}_{T-1}) \qquad (4.4.24)$$

yields the RPLR updating scheme:

DEFINITION 4.2.
(THE RECURSIVE PSEUDOLINEAR REGRESSION SCHEME).
Let Y_T be defined by (4.4.1) and R_T by (4.4.24). Then the RPLR updating scheme is defined by

$$\epsilon_T(\hat{\theta}_{T-1}) \;\;=\;\; Y_T \Leftrightarrow X_T(\hat{\theta}_{T-1})\hat{\theta}_{T-1} \qquad (4.4.25a)$$

$$\hat{\theta}_T \;\;=\;\; \hat{\theta}_{T-1} + R_T^{-1}(\hat{\theta}_{T-1})\epsilon_T(\hat{\theta}_{T-1})X_T(\hat{\theta}_{T-1}) \qquad (4.4.25b)$$

$$R_T(\hat{\theta}_{T-1}) \;\;=\;\; R_{T-1}(\hat{\theta}_{T-1}) + X_T(\hat{\theta}_{T-1})X_T^\top(\hat{\theta}_{T-1}). \qquad (4.4.25c)$$

▲

REMARK 4.1. Notice the close resemblance to the RPEM updating scheme. To change RPLR to RPEM, just swap X with ψ. Also notice that RPLR is identical to RPEM if X is independent of θ, since

$$\psi = \Leftrightarrow\frac{\partial \epsilon}{\partial \theta} = \Leftrightarrow\frac{\partial(Y \Leftrightarrow X^\top \theta)}{\partial \theta} = \frac{\partial(X^\top \theta)}{\partial \theta} = X \qquad (4.4.26)$$

when $\partial X/\partial \theta = 0$.   ▼

4.4.4 Parameter uncertainty

Unfortunately, there is no straightforward way to measure the uncertainty of parameter estimates obtained using RPLR and RPEM. It is not possible to use the inverse Hessian approach from Maximum Likelihood here because this approach

is only valid if the variance of the estimator is identical to the lower bound of Cramer-Rao's inequality, an this is not likely to be the case here. Because of this, no estimates of the parameter uncertainty will be provided for RPEM and RPLR estimates.

4.4.5 Application to Arch

In order to apply RPEM and RPLR methods to a particular application, Y, X and θ must be given an interpretation that matches the particular application as well as the requirements imposed on Y, X and θ by the recursive estimation methods. The most obvious way to search for this interpretation would be to write down the one-step predictor of the particular Arch model in focus, and thereby establish its dependence of the parameters θ.

This approach, however, leads nowhere since the expected value of an Arch-residual is always zero. Instead it is convenient to transform the residuals by setting Y_t equal to some transformation of ε_t depending on the particular Arch model.

Note that the prediction error ϵ is different from the Arch residual ε. ϵ is the prediction error referred to by the RPLR and RPEM methods whereas ε is the residual of the model (3.1.3).

4.4.5.1 GARCH

For GARCH models, the transformation $Y_t = \varepsilon_t^2 = z_t^2 \sigma_t^2$ seems useful. For given parameters, the one-step ahead prediction of Y_t is

$$
\begin{aligned}
\hat{Y}_{t|t-1} &= \mathrm{E}[z_{t|t-1}^2]\mathrm{E}[\sigma_{t|t-1}^2] \\
&= \sigma_t^2 = h_t \\
&= \alpha_0 + \sum_{i=1}^{q} \alpha_i Y_{t-i} + \sum_{j=1}^{p} \beta_j h_{t-j} \\
&= X_t^\top \theta
\end{aligned}
\tag{4.4.27}
$$

where

$$
X_t^\top = \begin{bmatrix} 1 & Y_{t-1} & \cdots & Y_{t-q} & h_{t-1} & \cdots & h_{t-p} \end{bmatrix}
\tag{4.4.28}
$$

and

$$\theta^\top = \begin{bmatrix} \alpha_0 & \alpha_1 \cdots \alpha_q & \beta_1 \cdots \beta_p \end{bmatrix} \tag{4.4.29}$$

This choice of X suggests the use of a RPLR method for estimation since a RPEM method would require an expression for $\partial Y_{t|t-1}/\partial\theta$. This expression would be too cumbersome to compute due to its recursive nature caused by the fact that $Y_{t-1}, \ldots, Y_{t-q}$ of (4.4.28) depends on θ.

4.4.5.2 ARCH

The choices of Y, X and θ in the ARCH case follows directly from above, as an ARCH(q) process is identical to a GARCH(0, q) process. Again $Y_t = \varepsilon_t^2$ but this time

$$\hat{Y}_{t|t-1} = \alpha_0 + \sum_{i=1}^{q} \alpha_i Y_{t-i} \tag{4.4.30}$$

and

$$X_t^\top = \begin{bmatrix} 1 & Y_{t-1} \cdots Y_{t-q} \end{bmatrix} \tag{4.4.31}$$

and

$$\theta^\top = \begin{bmatrix} \alpha_0 & \alpha_1 \cdots \alpha_q \end{bmatrix} \tag{4.4.32}$$

meaning that the RPEM method is just as suitable as the RPLR method now that X_t no longer depends on the parameters, c.f. Remark 4.1.

4.4.5.3 EGARCH

Finding interpretations of Y, X and θ that allows an EGARCH model to be estimated using RPEM or RPLR is a bit more difficult than what was the case with ARCH and GARCH. The reason why is the more "exotic" relationship

$$\sigma_t = \exp\left(\frac{1}{2}h_t\right) \tag{4.4.33}$$

between σ_t and h_t that applies to EGARCH. However, there is a way out. If the constant k is defined by $k = \mathrm{E}[\ln z_t^2]$ and Y is chosen so that $Y_t = \ln \varepsilon_t^2 - k$ then it is possible to write

$$
\begin{aligned}
\hat{Y}_{t|t-1} &= \mathrm{E}[\ln \varepsilon_{t|t-1}^2] - k \\
&= \mathrm{E}[\ln z_{t|t-1}^2 + \ln \sigma_{t|t-1}^2] - k \\
&= \mathrm{E}[\ln z_t^2] + \ln \sigma_t^2 - k \\
&= \ln \sigma_t^2 = h_t \\
&= \alpha_0 + \sum_{i=1}^{q} \alpha_i \left[\omega z_{t-i} + \lambda \left(|z_{t-i}| - \mathrm{E}|z_{t-i}| \right) \right] + \sum_{j=1}^{p} \beta_j h_{t-j} \\
&= \alpha_0 + \sum_{i=1}^{q} \alpha_i \omega z_{t-i} + \sum_{i=1}^{q} \alpha_i \lambda \left(|z_{t-i}| - \mathrm{E}|z_{t-i}| \right) \\
&\quad + \sum_{j=1}^{p} \beta_j h_{t-j}. \tag{4.4.34}
\end{aligned}
$$

The RPLR method applies if X_t and θ is chosen to be the two $(1 + 2q + p)$-dimensional vectors

$$
X_t = \begin{bmatrix} 1 \\ z_{t-1} \\ \cdots \\ z_{t-q} \\ |z_{t-1}| - \sqrt{\frac{2}{\pi}} \\ \cdots \\ |z_{t-q}| - \sqrt{\frac{2}{\pi}} \\ \ln \sigma_{t-1}^2 \\ \cdots \\ \ln \sigma_{t-p}^2 \end{bmatrix}, \qquad
\theta = \begin{bmatrix} \alpha_0 \\ \alpha_1 \omega \\ \cdots \\ \alpha_q \omega \\ \alpha_1 \lambda \\ \cdots \\ \alpha_q \lambda \\ \beta_1 \\ \cdots \\ \beta_p \end{bmatrix}. \tag{4.4.35}
$$

Note that, as described in Chapter 3, not all parameters are identifiable. Also note that since the value $\sqrt{2/\pi}$ has been added to the absolute innovations in the X-

vector, the setup has been limited to cases where z is Gaussian distributed[1]. It is of course straightforward to choose another distribution, compute $E|z|$ and insert the obtained value instead of $2/\pi$ in X.

The setup (4.4.35) may not be suitable for model order parameters greater than (1,1). The reason why is that for higher order EGARCH models, the setup does not guarantee that $\begin{bmatrix} \alpha_1\omega \cdots \alpha_q\omega \end{bmatrix}$ is proportional to $\begin{bmatrix} \alpha_1\lambda \cdots \alpha_q\lambda \end{bmatrix}$. Hence, when estimating higher order EGARCH models one may want to consider an alternative approach where one of the parameters ω or λ is fixed prior to estimation.

What remains is to compute k. By its definition,

$$
\begin{aligned}
k &= E[\ln z_t^2] \\
&= \int_{-\infty}^{\infty} \frac{1}{\sqrt{2\pi}} \ln u^2 e^{-\frac{u^2}{2}} du \\
&= 2\int_{0}^{\infty} \frac{1}{\sqrt{2\pi}} \ln u^2 e^{-\frac{u^2}{2}} du \\
&= 2\int_{0}^{\infty} \sqrt{\frac{2}{\pi}} \ln u e^{-\frac{u^2}{2}} du \\
&= \frac{4}{\sqrt{\pi}} \int_{0}^{\infty} \frac{1}{\sqrt{2}} \left[\ln\left(\frac{u}{\sqrt{2}}\right) + \ln\sqrt{2} \right] e^{-\left(\frac{u}{\sqrt{2}}\right)^2} du \\
&= \frac{4}{\sqrt{\pi}} \left[\int_{0}^{\infty} \ln t\, e^{-t^2} dt + \int_{0}^{\infty} \frac{\ln\sqrt{2}}{\sqrt{2}} e^{-\frac{u^2}{2}} du \right] \\
&= \frac{4}{\sqrt{\pi}} \left[\frac{\Leftrightarrow\sqrt{\pi}}{4}(\gamma + 2\ln 2) + \frac{\ln\sqrt{2}}{\sqrt{2}} + \frac{1}{2}\sqrt{2\pi} \right] \\
&= \Leftrightarrow(\gamma + 2\ln 2) + \ln 2 = \Leftrightarrow(\gamma + \ln 2) \\
&\simeq \Leftrightarrow 1.270362845
\end{aligned}
\tag{4.4.36}
$$

where γ is Euler's constant.

4.5 Model validation

Now that the estimation procedures to be used are in place, it is time to estimate the parameters of GARCH and EGARCH. Before the estimation procedures are

---

[1]Strictly speaking, the restriction is not to Gaussianity but to distributions of z where $E|z| = \sqrt{2/\pi}$.

applied, it may be illustrative to discuss how to validate parameter estimates of Arch models.

As the motivation of Arch models is the return series characteristics captured in chapter 2, the model validation should measure the abilities of the models to describe these characteristics.

This validation is obtained by observing Equation 3.1.1 of the Arch framework. As $\varepsilon_t = z_t \sigma_t$, the standardized residuals r_t,

$$r_t = \frac{\varepsilon_t}{\sigma_t} \tag{4.5.1}$$

ought to have properties identical to those of z_t. That is, $\{r_t\}$ should be white noise with zero mean and unit variance. There also should be no autocorrelation in $\{r_t\}$ or in $\{r_t^2\}$, and the distribution of r_t should be identical to the distribution of z_t. Thus, estimated models are validated by applying the investigations of Chapter 2 to the standardized residuals to see if all the relevant return series characteristics have been accounted for.

In those cases where the parameters are estimated using simulated data, the true parameter values are included in the validation process.

4.6 Estimation results

This section presents a number of parameter estimation results obtained by applying the Recursive Pseudolinear Regression (RPLR) method and the Maximum Likelihood method to various GARCH and EGARCH models. As previously stated, the main goals are to compare the two estimation methods to one another and to find out which Arch model provides the best description of the data. The subgoals are to measure the impact on the estimates of stipulated initial values.

The section is divided as folows: First, GARCH(1,1) with Gaussian innovations is fitted to the four return series using Maximum Likelihood, and the obtained estimates are carefully analysed. Then, these estimations are repeated using RPLR, thereby enabling a comparison of the performance of RPLR to Maximum Likelihood. Afterwards, a GARCH(1,1) model with Gaussian innovations is estimated from simulated data using Maximum Likelihood as well as RPLR to allow for a comparison of the obtained estimates to the true parameters. Subsequently, a GARCH(1,1) model with t-distributed innovations is fitted to real life return series as well as simulated data in order to find out if the t-distribution is to be preferred

to the normal distribution. The impact of different initial values on Maximum Likelihood and RPLR is measured in the succeeding subsection, and a study of GARCH(1,1) applied to small data series is provided thereafter. The section is concluded by an estimation of EGARCH(1,1) applied to return series.

4.6.1 GARCH(1,1)

4.6.1.1 Return series – Maximum Likelihood

The first estimation results to be presented are those of a GARCH(1,1) model with Gaussian innovations fitted to the four return series using Maximum Likelihood. The parameter estimates are presented in Table 4.2 along with their standard errors and the value of the log-likelihood. As previously stated, a GARCH(1,1) likelihood needs a stipulated value of h_0 to be computed and for the estimates in Table 4.2 h_0 has been set to the unconditional variance, c.f. Equation 3.3.8.

Series	HP	Sony	Mobil	Pepsi
α_0	7.3100×10^{-5}	3.2958×10^{-6}	4.8515×10^{-6}	2.9344×10^{-5}
	(3.0331×10^{-5})	(1.3138×10^{-6})	(1.8363×10^{-6})	(1.0580×10^{-5})
α_1	1.0411×10^{-1}	6.0622×10^{-2}	4.7775×10^{-2}	7.7883×10^{-2}
	(3.7029×10^{-2})	(1.1253×10^{-2})	(1.2551×10^{-2})	(2.4485×10^{-2})
β_1	7.3705×10^{-1}	9.2884×10^{-1}	9.1611×10^{-1}	8.0033×10^{-1}
	(9.2735×10^{-2})	(1.3554×10^{-2})	(2.1343×10^{-2})	(5.8673×10^{-2})
L_T	3.0953×10^3	3.4982×10^3	3.8688×10^3	3.5005×10^3

Table 4.2: Maximum Likelihood parameter estimates of GARCH(1,1) applied to return series.

To validate these estimation results, each of the four series are examined in turn along the lines suggested in Section 4.5. The estimates of Hewlett-Packard will be thoroughly examined to fully illustrate the advantages and disadvantages of the estimated GARCH model, the estimates of Sony, Mobil and Pepsi with less rigour to avoid tiresome repetitions.

Hewlett-Packard
The estimate of β_1 of Hewlett-Packard is 0.737 with a standard error of 0.0927. The usual t-test statistic for a null hypothesis of $\beta_1 = 0$ against the alternative[2]

---

[2]Recall from (3.3.3) that negative GARCH parameters are prohibited and from Section 4.2.4.1 that these constraints have been incorporated into the estimation procedure.

$\beta_1 > 0$ is 7.948 and the 99%-fractile of the $t(1261)$-distribution is 2.3293 meaning that the estimate of β_1 is significantly different from zero. The estimate of α_1 is 0.104 with a standard error of 0.0370, corresponding to a t-test statistic of 2.8116 which means that α_1 too is significantly different from zero. The same conclusion applies to α_0 estimated to a value of 7.310^{-5} with a standard error of 3.033×10^{-5} and a test statistic of 2.4101. The significance of α_0 is particular important as a value of $\alpha_0 = 0$ means that the unconditional variance of the process is zero and thus $\alpha_0 = 0$ is disallowed by the GARCH parameter constraints. In order to verify that the process driving the Hewlett-Packard returns is not an IGARCH process, it is noted that the value of $\alpha_1 + \beta_1$ is 8.412 with a standard error of 0.0999 causing the test statistic of the hypothesis $\alpha_1 + \beta_1 = 1$ against the alternative $\alpha_1 + \beta_1 < 1$ to be $\Leftrightarrow 1.591$ which is significant at all levels at or above 5.6%. The estimated model is therefore on the brink of being an IGARCH model.

In Figure 4.1 the 1264 geometric returns of Hewlett-Packard are plotted above 1264 realizations from a GARCH(1,1) process[3] with parameters equal to those of the estimates in the first column of Table 4.2.

The first thing to note about the simulation is that the simulated data is quite homogeneous. By a close look, however, the figure reveals that there is a small amount of clustering in the data and there is also a small amount of volatility persistence present. There are at most one significant outlier in the simulated data. When comparing the plot of the simulated data to the plot of the actual Hewlett-Packard returns it is seen that the clustering and volatility persistence features are quite similar, but that the single outlier of the simulated data is clearly outnumbered by the outliers of the actual returns. In this respect the similarities in the two data sets are quite small, indicating that the conditional normality assumption may not be correct and that a more heavy-tailed conditional distribution may be more appropriate.

The standardized residuals of Hewlett-Packard are plotted in Figure 4.2. With the goal of the standardized residuals being Gaussian White Noise in mind, it is difficult to be impressed by looking at the plot. Discregarding the scale difference, the plot of the standardized residuals is hard to distinguish from the plot of the actual Hewlett-Packard returns at the upper part of Figure 4.1.

A formal investigation of the standardized residuals along the lines of Section 2.2.2 reveals that the kurtosis estimate is 5.6503 and the Jarque-Bera test statistic is 374.8, a highly significant value. These numbers confirms that the estimated GARCH(1,1) model fails to explain the heavy tails of the Hewlett-Packard returns. However, the kurtosis of the Hewlett-Packard return was in Chapter 2 estimated to 6.089

---

[3]The simulation have been performed using the same h_0 as the log-likelihood.

Figure 4.1: The geometric returns of Hewlett-Packard (top) and realizations from a GARCH(1,1) process using the parameter values of the second column of Table 4.2 (bottom).
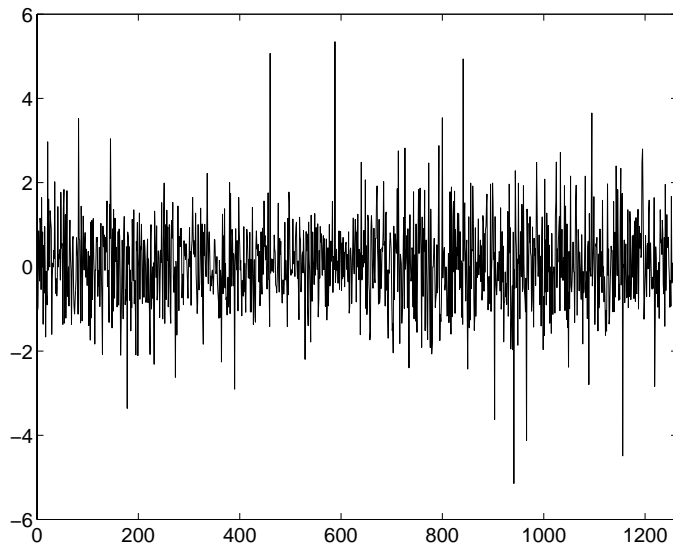
Figure 4.2: The standardized residuals of Hewlett-Packard.

and the Jarque-Bera test statistic was 502.9 indicating that the GARCH(1,1) model has explained *some* part of the heavy tails. The autocorrelation function estimates of the standardized residuals and the squared standardized residuals are as plotted in Figure 4.3.

The approximative 95% critical region for the autocorrelation function estimates is $\pm 2/\sqrt{1264} = \pm 0.0563$ and as no value of the autocorrelation function of the squared Hewlett-Packard returns exceeds this limit (the estimate of the autocorrelation function of the squared standardized residuals at lag 25 is 4.85), the estimated GARCH(1,1) model has actually been able to explain this characteristic of the Hewlett-Packard returns. The autocorrelation structure of the ordinary standardized residuals is identical to the autocorrelation structure of the actual Hewlett-Packard returns, c.f. Figure 2.4, as expected. These findings are confirmed by the Ljung-Box test statistic which is 35.644 for lag 30, insignificant at all levels at or below 22% and the Portmanteau Q-test with a test statistic of 14.227 at lag 30, insignificant at all levels at or below 99%! The Lagrange multiplier test for heteroskedasticity has a test statistic of 13.551 at lag 30 which is also insignificant at all levels below 99%.

To assess the ability of the GARCH(1,1) model to predict the conditional variance, a plot of the squared Hewlett-Packard returns, the estimates of the conditional vari-
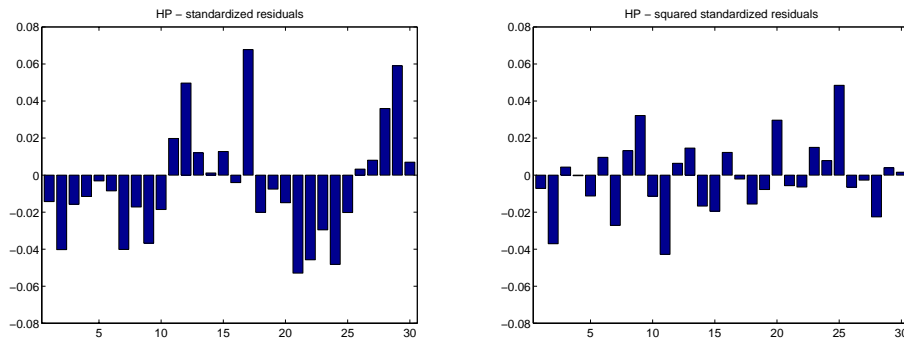
Figure 4.3: Estimated autocorrelation functions of the standardized residuals of Hewlett-Packard (left) and the squared standardized residuals (right).

ance h_t using the estimated GARCH(1,1) parameters and the estimated unconditional variance is presented in Figure 4.4. To allow for a closer inspection, only observations and associated estimates number 201 through 400 of the total number of 1264 returns have been plotted.

From the plot it is readily seen that the estimated conditional variance does not vary much from the unconditional variance and in periods of low volatility, e.g. from observation number 285 to 305 and 345 to 355, the estimated conditional variance is much larger than the data. Thus, the GARCH(1,1) volatility estimates are upward biased during these periods. Moreover, as it may be visible from the plot, the GARCH estimates exhibits some delay relative to the observations, i.e. when the volatility of the data changes from high to low, it takes some time for the GARCH model to adjust. When the volatility reverts from low to high, the delay seems shorter. The delay is a of course an inherent property of a model based on historical information, but in applications where a quick response is desired , one might wish to look for other models. The delay may suggest that the volatility persistence of the GARCH(1,1) model is actually too high compared to the return series data. The estimated GARCH(1,1) model predicts the unconditional variance to be 4.6021×10^{-4}, the second order moment of the Hewlett-Packard returns is 4.4730×10^{-4} and the mean of the estimated conditional variances is 4.5513×10^{-4}. Thus, the estimated GARCH(1,1) model is upward biased with respect to the uncondtional and conditional variances.

Sony
The estimate of β_1 of Sony is 0.929 with a standard error of 0.0135. As the param-
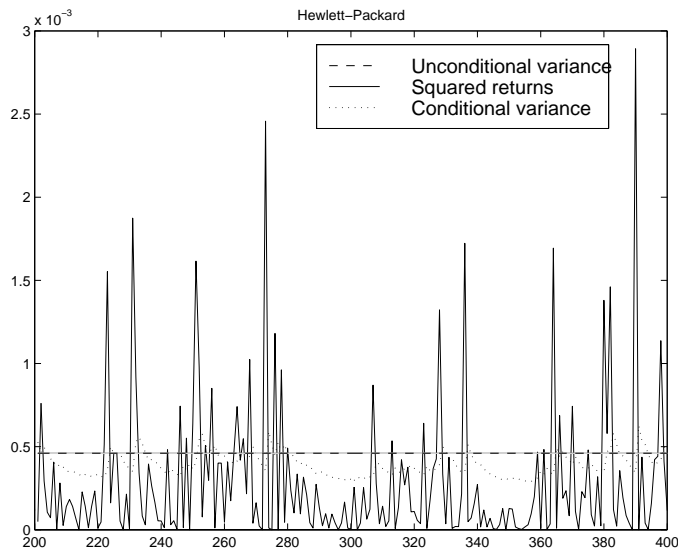
Figure 4.4: Squared Hewlett-Packard returns (solid line), estimates of the conditional volatility of the GARCH(1,1) model (dotted line) and the estimated unconditional variance (dashed line).

eter estimate is larger than β_1 of Hewlett-Packard and the standard error is smaller, β_1 of Sony is significant. The estimate of α_1 is 0.0606 with a standard error of 0.0125 yielding a t-test statistic of 5.3872 which is highly significant. The same conclusion applies to α_0 which is estimated to a value of 3.296×10^{-6}, a standard error of 1.3138×10^{-6} and a t-test statistic of 2.5084 which is significant at all levels below 99%. The hypothesis of $\alpha_1 + \beta_1 = 1$ has a test statistic of $\Leftrightarrow 57.63$ so an IGARCH(1,1) process is out of the question.

The geometric returns of Sony and realizations from a GARCH(1,1) process with the estimated parameter values are plotted in Figure 4.5.

Again, the GARCH(1,1) model does not seem able to replicate the heavy tails of the returns but the model do appear to capture the volatility structure. The kurtosis of the standardized residuals is 6.8585, a bit lower than the kurtosis of the actual Sony returns which was estimated to a value of 7.725. The Jarque-Bera test statistic is 822.0, highly significant but lower than the value 1269 of the actual returns. An expection of the autocorrelation function estimates plotted in Figure 4.6 reveals that the GARCH model has failed to remove the significant autocorrelation at lag 18, but that the autocorrelations at lags 1, 2, 5, 6, 7, 8, 10, 14, 15 and 26, c.f.
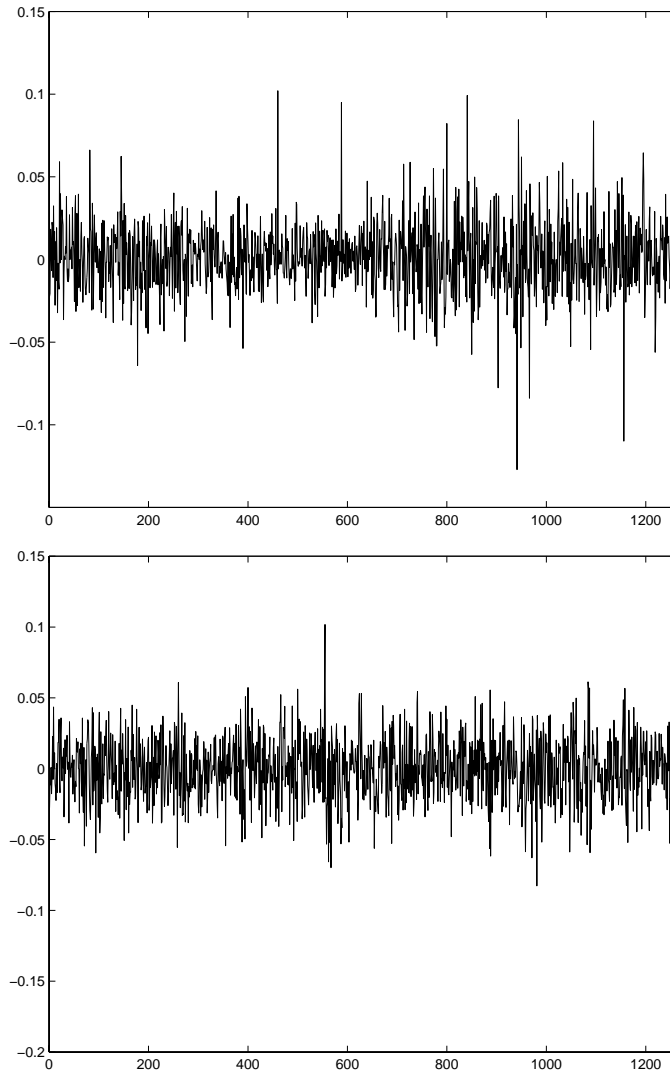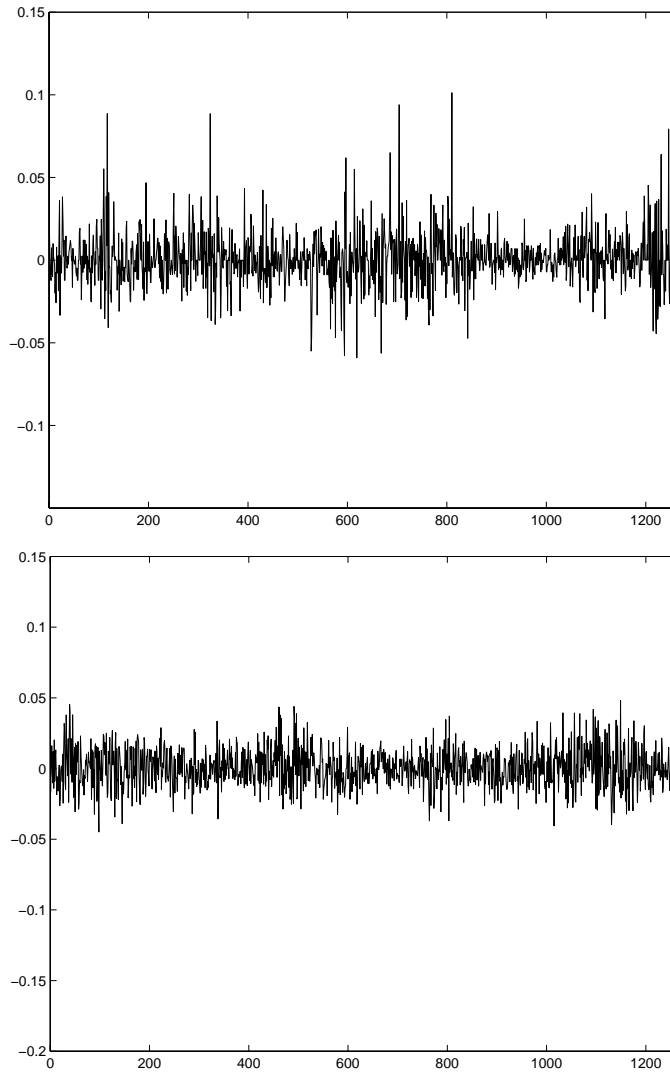
Figure 4.5: The geometric returns of Sony (top) and realizations from a GARCH(1,1) process using the parameter values of the third column of Table 4.2 (bottom).

Figure 2.4 have been captured. A preliminary assumption would be that the Sony returns are actually a bit too heteroskedastic to GARCH(1,1), an assumption partly backed by the fact that the Portmanteau test statistic of the residuals is 24.808 at lag 30, insignificant at all levels below 26% and thus not as "elegant" as in the case of Hewlett-Packard.
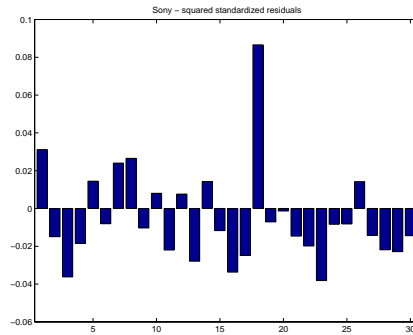
Figure 4.6: Estimated autocorrelation function of the squared standardized residuals of Sony.

The remarks regarding the ability of the GARCH(1,1) model fitted to the Sony returns to predict the conditional variance are no different than those applying to Hewlett-Packard.

Mobil

The t-test statistics of significance of the Mobil parameters are as presented in Table 4.3 along with the probabilities of observing more extreme values of the test statistics under the H_0's.

Hypothesis	$\beta_1 = 0$	$\alpha_1 = 0$	$\alpha_0 = 0$	$\alpha_1 + \beta_1 = 1$
Test statistic	42.923	3.8065	2.6420	$\Leftrightarrow 1.4584$
Probability	0	$< 1\%$	$< 1\%$	7.25%

Table 4.3: Test statistics and associated probabilities of various hypothesis regarding the estimated Mobil parameters.

As seen from the table, all parameter values are highly significant and, as was the case of Hewlett-Packard, it is generally not possible to reject the hypothesis that the underlying process driving the Mobil returns is actually an IGARCH(1,1).

With regard to the failure of GARCH(1,1) to explain the heavy tails of the Hewlett-

Packard and Sony returns, one might hope that the heavy tails of Mobil can actually be captured by GARCH as the tails of the Mobil returns are actually not that fat, recall from Table 2.1 that the kurtosis of the Mobil returns is only 4.040. However, the kurtosis of the standardized Mobil returns is 3.889 so, again, the GARCH(1,1) model is only capable of reducing the heaviness of the tails. The Jarque-Bera test applied to the standardized residuals yields a highly significant test statistic of 42.37, a bit lower than the statistic of the actual returns which in Chapter 2 was estimated to a value of 57.14.

The conclusions regarding the ability of the GARCH(1,1) model to capture the heteroskedasticity and autocorrelation structure of the Mobil returns is no different from the conclusions obtained in the previous paragraphs. Thus, there is no need to repeat this analysis.

Pepsi

The resuls of testing the significance of the estimated Pepsi parameters are as displayed in Table 4.4. All three parameters are highly significant and an IGARCH(1,1) model is definately inadequate.

Hypothesis	$\beta_1 = 0$	$\alpha_1 = 0$	$\alpha_0 = 0$	$\alpha_1 + \beta_1 = 1$
Test statistic	13.6407	3.1808	2.7736	$\Leftrightarrow 17.645$
Probability	0	$< 1\%$	$< 1\%$	0

Table 4.4: Test statistics and associated probabilities of various hypothesis regarding the estimated Pepsi parameters.

Recalling the Ljung-Box and Portmanteau test results from Chapter 2 and the Likelihood multiplier test results from Section 4.3 of this chapter, the Pepsi returns are different from the returns of Hewlett-Packard, Sony and Mobil in the sence that the squared returns do not exhibit significant autocorrelation for lags greater than one.

The standardized residuals of Pepsi are, however, no different from the previously analysized standardized residuals, as the squared standardized residuals of Pepsi do not exhibit autocorrelation at any lag. The fitted GARCH(1,1) model has reduced the autocorrelation in the squared returns, significant or not. The Portmanteau test statistics of the standardized residuals at lag one and lag 30 are insignificant at all levels below 52% and 88%, respectively.

The inability of GARCH(1,1) to capture the heavy tails of the returns equally applies to Pepsi.

Summary

Fitting a GARCH(1,1) model to return series using Maximum Likelihood yields
models that are perfectly able to capture the heroskedasticity and autocorrela-
tion structure of the returns, thus satisfying the objective of Gaussiannity of the
standardized residuals residuals. GARCH(1,1) is, however, unable to replicate
the heavy tails of the return distributions. Furthermore, estimates of the condi-
tional variance of the returns obtained using the fitted GARCH models reveals that
GARCH(1,1) is a bit slow in adapting changes to the magnitude of the conditional
volatility, especially when the volatility changes from high to low.

4.6.1.2   Return series – Recursive Pseudolinear Regression

The parameter estimates obtained when a GARCH(1,1) model is fitted to the four
return series using the recursive pseudilinear regression method (RPLR) developed
in Section 4.4 are presented in Table 4.5.

Series	HP	Sony	Mobil	Pepsi
α_0	2.2755×10^{-5}	6.9024×10^{-6}	6.5947×10^{-7}	7.9248×10^{-5}
α_1	3.2087×10^{-2}	3.3573×10^{-2}	9.8273×10^{-3}	5.2765×10^{-2}
β_1	9.1701×10^{-1}	9.3969×10^{-1}	9.8515×10^{-1}	6.0847×10^{-1}

Table 4.5: Recursive parameter estimates of GARCH(1,1) applied to return
series.

The RPLR parameter estimates are quite different from the Maximum Likelihood
estimates. Though the parameters of Table 4.5 are clearly of the same magnitude
as the Maximum Likelihood estimates, there are substantial differences. The val-
ues of α_1 obtained using RPLR are for all series substantially smaller than the
Maximum Likelihood estimates. The ratio ranges from approximately 1.5 (Pepsi)
to about 4.9 (Sony). The pattern is reversed when the parameter estimates of β_1
is considered, at least for the Hewlett-Packard, Sony and Mobil returns where the
parameter estimates obtained using RPLR are larger than the parameter estimates
obtained using Maximum Likelihood. Surpricingly, RPLR estimates β_1 of Pepsi to
a value considerably smaller than the Maximum Likelihood procedure does. There
does not seem to be uniform relation between the RPLR and Maximum Likelihood
estimates of α_0. No standard error of the parameter estimates is presented as the
RPLR method does not allow for an immediate computation of the parameter co-
variance matrix. When the standard errors are absent, no t-test for significance can
be performed.

The problem when faced with two sets of parameter values is to decide which one to rely on. This can be done only by performing an analysis similar to the one performed in the context of Maximum Likelihood, but though this analysis may answer the question of which estimation routine performes the best with regard to the autocorrelation structure, it will not answer the question when it comes to the heavy tails of the returns. The reason why is that the RPLR estimation procedure do not assume a particular (conditional) distribution of the returns, nor does it specify any distribution parameters.

Hewlett-Packard

The sum of the RPLR parameter estimates of α_1 and β_1 is 0.949 which is considerably larger than the corresponding value of the Maximum Likelihood estimates which is 0.841. This means that the RPLR method estimates the process driving the Hewlett-Packard returns to have a larger unconditional variance than the Maximum Likelihood estimation procedure does. The parameter based estimate of the unconditional variance (Formula 3.3.8) is 4.4702×10^{-4} in the RPLR case and 4.6021×10^{-4} in the Maximum Likelihood case. For comparison, the second order moment of the Hewlett-Packard returns is 4.4730×10^{-4} so, as far as the unconditional variance is considered, RPLR outperforms the Maximum Likelihood method. Moreover, the mean of the estimated conditional variances is 4.4730×10^{-4} (!) in the RPLR case and 4.5513×10^{-4} in the Maximum Likelihood case.

The standardized residuals of Hewlett-Packard obtained using the conditional standard deviation estimates of the RPLR fitted model have a kurtosis of 6.0135 which is larger than the kurtosis of the standardized residuals obtained using Maximum Likelihood (5.6503) and only slightly lower than the kurtosis of the actual Hewlett-Packard returns (6.089). Thus, RPLR performs even worse than Maximum Likelihood when it comes to the heavy tails. As there are no visible differences between the standardized residuals obtained using Maximum Likelihood and the standardized residuals obtained using RPLR, the latter are not shown. However, in Figure 4.7, the estimated autocorrelation function of the squared standardized residuals obtained using RPLR is presented.

By comparing this plot to the rightmost plot in Figure 4.3, it is observed that the RPLR estimated GARCH(1,1) model has removed more of the largest "spikes". The spikes of lag 2, 9, 20 and 25 are smaller at the RPLR plot than at the Maximum Likelihood plot, but the differences are too small to support a judgement as to which method performs the best in explaining the autocorrelation in the squared returns. Moreover, the Portmanteau Q-test at lag 30 has a test statistic of 14.321,
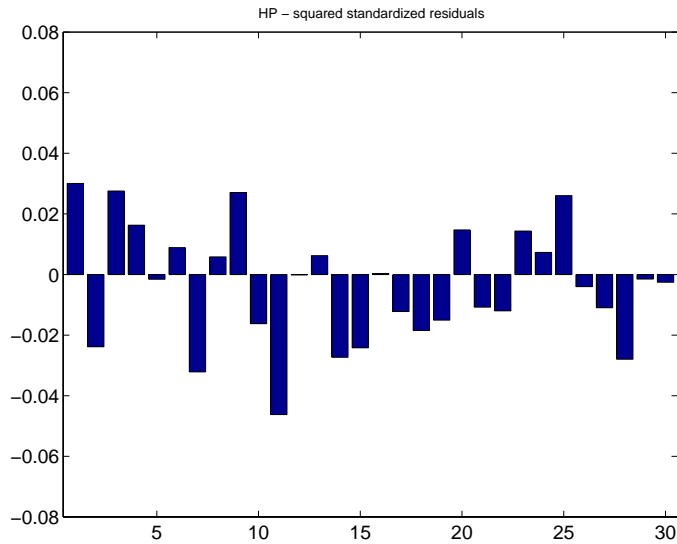
Figure 4.7: Estimated autocorrelation functions of the squared standardized residuals of Hewlett-Packard.

0.94 higher than the corresponding figure of the Maximum Likelihood estimated standardized residuals and the Likelihood multiplier test statistic is 14.081 of the RPLR residuals, 0.530 higher than the test statistic of the Maximum Likelihood returns. Thus, if a conclusion is to be drawn regarding heteroskedasticity it will be that the two estimation routines performs equally well.

In order to measure the ability of the RPLR-estimated GARCH(1,1) model in predicting the conditional variance, a plot similar to the one in Figure 4.4 have been constructed. It is presented in Figure 4.8.

When comparing the plot in Figure 4.8 to Figure 4.4, the higher volatility persistence of the RPLR-estimated model is obvious. The RPLR-estimated model is even more reluctant to follow the data than the Maximum Likelihood estimated model. This is because the estimate of the parameter β_1 is much higher in the RPLR-estimated model than in the Maximum Likelihood estimated model, so one should not use this observation to jump to the conclusion that the Maximum Likelihood estimated model is superior in this regard. After all, one could just choose a set of model parameters with a relatively low value of β_1, and this model would then have a lower volatility persistence than the two estimated models, thus adjusting faster to changes in volatility. The speed of volatility adjustment should only be used as a measure of the ability of GARCH(1,1) to explain return series
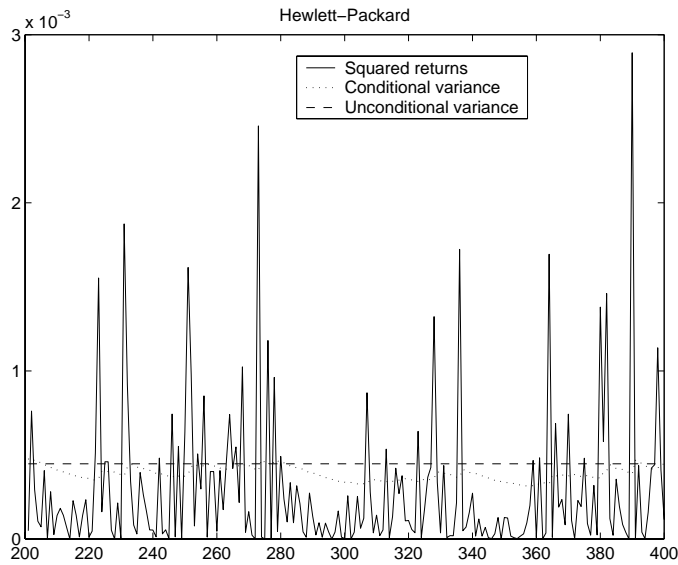
Figure 4.8: Squared Hewlett-Packard returns (solid line), estimates of the conditional volatility of the GARCH(1,1) model (dotted line) and the estimated unconditional variance (dashed line).

characteristics, not to choose between Maximum Likelihood and RPLR.

The conditional variance estimates of the RPLR-estimated model are not biased as opposed to the estimates of the Maximum Likelihood estimated model. The sum of the squared differences between the squared returns and the conditional variance estimates is 1.2725×10^{-3} in the RPLR case and 1.2826×10^{-3} in the Maximum Likelihood case, indicating that the RPLR-estimed model *do* a better job in estimating the conditional variance than does the Maximum Likelihood estimated model. But the price is as mentioned a slower responce to changes in the conditional variance.

Sony

In the case of Sony, the sum of the parameters α_1 and β_1 is 0.973 in the RPLR case and 0.989 in the Maximum Likelihood case. Thus, the situation is just the opposite of Hewlett-Packard as RPLR estimates Sony to have a lower unconditional variance than does Maximum Likelihood. On the other hand, the second order moment of the Sony returns is 2.5845×10^{-4}, the RPLR parameter based estimate of the unconditional variance is 2.5817×10^{-4} and the Maximum Likelihood parame-

ter based estimate is 3.1281×10^{-4}. This means that (again) RPLR outperforms Maximum Likelihood when it comes to estimating the unconditional variance.

The standardized residuals computed using the RPLR estimated parameters have a kurtosis of 6.9699 which, as was the case with Hewlett-Packard, is lower than the kurtosis of the actual returns (7.7250) but higher than the kurtosis of the Maximum Likelihood based standardized residuals (6.8585).

Unsimilar to the Hewlett-Packard situation, the ability of the RPLR estimated model to capture the autocorrelation structure of the actual returns is not as good as the Maximum Likelihood estimated model. This is observed by comparing the estimated autocorrelation function of the squared standardized residuals of RPLR in Figure 4.9 to the estimated autocorrelations of the squared standardized residuals of Maximum Likelihood in Figure 4.6. Several spikes are highest on the RPLR plot, in particular the significant spike of lag 18.
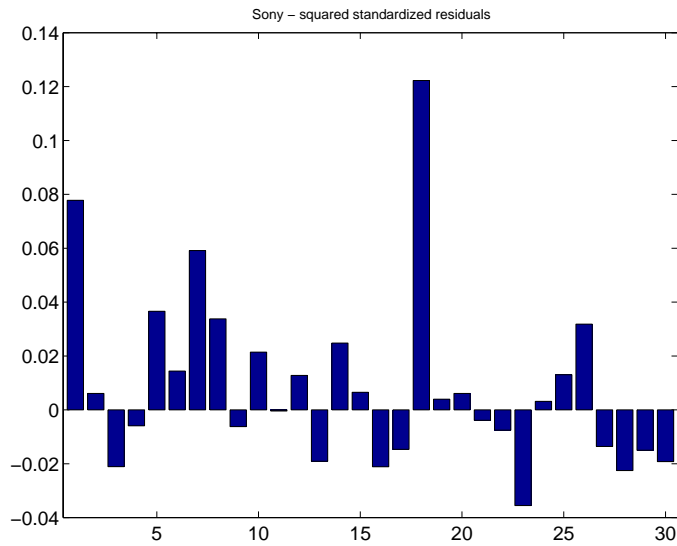
Figure 4.9: Estimated autocorrelation functions of the squared standardized residuals of Sony.

The Portmanteau test statistic of the standardized residuals at lag 30 (43.411), insignificant at all levels below 5.39% and the Likelihood multiplier test for heteroskedasticity at lag 30 (42.247), insignificant at all levels below 6.82% are in line with this assumption. Considering the 5% level, the RPLR-estimated model barely manages to capture the heteroskedasticity of Sony.

The sum of the squared differences between the squared returns of Sony and the conditional variance estimates is 5.6555×10^{-4} in the RPLR case and 5.6790×10^{-4} in the Maximum Likelihood case.

Mobil

The sum of the α_1 and β_1 estimates is 0.995 higher than the corresponding figure of the Maximum Likelihood estimated model which is 9.6389. Thus, the RPLR estimated model is very close to IGARCH(1,1). The second order moment of the mobil returns is 1.3182×10^{-4}, the unconditional variance of the RPLR estimated model is 1.3133×10^{-4} and the corresponding figure of the Maximum Likelihood estimated model is 1.3435×10^{-4}. Once again, the RPLR method is the best when the second order moment is considered.

The usual pattern is, however, *not* recognized when the kurtosis estimates are considered. The kurtosis of the RPLR generated standardized residuals is 3.393, the kurtosis of the Maximum Likelihood generated standardized residuals is 3.89 and the kurtosis of the actual Mobil returns is 4.040. Hence, when it comes to Mobil, the RPLR method is the one that is best at explaining the heavy tails.

The ability of the RPLR estimated model to model the autocorrelation structure of the mobil returns is indistinguishable from the ability of the Maximum Likelihood estimated model. Thus, the autocorrelation function estimates as well as the Portmanteau and Likelihood ratio test statistics have been left out. The sum of the squared differences between the squared mobil returns and the conditional variance predictions is 6.6036×10^{-5} in the RPLR case and 6.5473×10^{-5} in the Maximum Likelihood case. Hence, in this regard the Maximum Likelihood estimated model provides the best prediction of the conditional variances, but its predictions are biased.

Pepsi

The RPLR parameter estimates of the Pepsi returns are, as previously stated, interesting because Pepsi is the only one of the four series for which it is the case that the estimate of β_1 is lower than the corresponding Maximum Likelihood estimate. The sum of α_1 and β_1 of the RPLR estimates is only 0.6612, the corresponding value of the Maximum Likelihood estimates is 0.8782. The parameter based estimate of the unconditional variance is 2.3393×10^{-4} in the RPLR case, 2.4095×10^{-4} in the Maximum Likelihood case and the second order moment of the Pepsi returns is 2.3394×10^{-4}. Hence, in the Pepsi case as in the cases of Hewlett-Packard, Sony and Mobil, the GARCH(1,1) model estimated using RPLR outperforms the model

fitted by Maximum Likelihood when it comes to estimation of the unconditional variance.

Similar to the situation of Mobil, the RPLR estimated model is best in explaining the heavy tails of the returns. The kurtosis of the Pepsi returns is 7.6009, the kurtosis of the standardized residuals obtained using RPLR is 6.8301 and the kurtosis of the residuals obtained using Maximum Likelihood is 6.9054.

The ability of the RPLR estimated model to capture the autocorrelation structure is (as usual) no different from the ability of the Maximum Likelihood estimated model. The Portmanteau statistic of the standardized residuals at lag 30 is insignificant at all levels below 96% better than the corresponding statistic of the Maximum Likelihood estimated model which is insignificant at all levels below 88%.

The sum of the squared differences between the squared Pepsi returns and the predictions of the conditional variances is 4.5744×10^{-4} for RPLR and 4.5797×10^{-4} for Maximum Likelihood.

Summary

The parameter estimates obtained when fitting a GARCH(1,1) model to the four return series are different from the estimates obtained using Maximum Likelihood though the estimates are of the same scale. The models fitted using RPLR are much better at estimating the unconditional variances of the return series and better at predicting the conditional variances than the Maximum Likelihood fitted models are.

The above analysis provide no straight answer to the question of which of the RPLR and the Maximum Likelihood fitted models that are best at capturing the heavy tails of the empirical return series distributions. The light tales of the estimated GARCH(1,1) models is a property of the GARCH model not of the estimation routines, and the two routines performs equally bad in this regard. Also, it has not been possible to distinguish the two estimation routines by the way the fitted models capture the autocorrelation structure of the return series.

If, upon the above findings, one estimation routine is to be preferred to another, it must be the RPLR routine. This routine does not perform worse than Maximum Likelihood when it comes to the heavy tails, it performs equally well when it comes to capture the autocorrelation structures, but it is better at estimating the variances, in particular the unconditional ones. However, the performance of Maximum Likelihood may be badly affected by the normality assumption which, for all that the findings of this subsection matters, is incorrect. In addition, the analysis is ob-

scured by the fact that the "true" parameters are unknown. But, if the task at hand is stricty limited to that of estimating a GARCH(1,1) with Gaussian innovations to geometric returns of stocks, my recommendation is to use RPLR.

4.6.1.3 Simulated series - Gaussian innovations

To measure the performance of RPLR versus Maximum Likelihood, the two estimation routines may be compared using simulated data. In this way, the assumed conditional distribution is assured to be correct and the true parameter values are available for comparison. However, before commencing the simulations, a number of choices have to made, choices that may effect the performance of the two estimation routines.

First of all, one have to decide which parameter values to use for simulation. This decision is difficult as one can not know in advance if one of the estimation routines performes best for some parameter values and worst for others. Theoretically, one could remedy this problem by exhausting the parameter space, but in practice this is not possible within a reasonable amount of time. Instead, one could use a Monte-Carlo approach by choosing a reasonable number of parameter vectors randomly and then compare the two estimation routines using these vectors. To increase the applicability of the conclusion obtained in this way, it is observed from Table 4.2 and Table 4.5 that GARCH(1,1) parameters of return series are not uniformly distributed in the space of allowed GARCH(1,1) parameters. The α_0 parameter seems to have values somewhere around 10^{-5}, the values of α_1 seems to be near 0.1 and the value β_1 is to be found around 0.8 and 0.9. In line with these findings, a simulation study have been carried out using 20 randomly chosen parameter vectors from the parameter space

$$\Theta = \{(\alpha_0, \alpha_1, \beta_1) \quad | \quad 6.2 \times 10^{-5} \leq \alpha_0 \leq 7.2 \times 10^{-5} \wedge$$
$$0.04 \leq \alpha_1 \leq 0.15 \wedge 0.70 \leq \beta_1 \leq 0.95 \wedge$$
$$\alpha_1 + \beta_1 < 1\} \tag{4.6.1}$$

Each simulation contains 1264 "observations" just like the return series.

The parameter estimates obtained from the simulation study are presented in Table 4.6.

Looking at the table it seems like Maximum Likelihood does the best in estimating the parameters, though both estimation routines have serious difficulties obtaining the correct values. Measured by the geometrical distance between the individual

Sim	$\alpha_0 \times 10^5$			$\alpha_1 \times 100$			$\beta_1 \times 10$		
	True	ML	RPLR	True	ML	RPLR	True	ML	RPLR
1	7.6	12.4	18.0	9.0	9.3	9.6	8.6	8.2	7.8
2	6.7	8.4	4.3	7.7	9.0	5.5	8.4	8.0	8.9
3	6.9	8.6	10.2	10.8	15.6	12.9	7.0	6.0	5.7
4	7.7	29.8	21.0	5.9	8.9	9.0	8.2	4.4	5.8
5	6.6	3.9	1.1	6.4	5.7	2.2	7.4	8.3	9.5
6	7.0	12.1	19.6	13.6	17.5	16.0	7.4	6.2	5.0
7	7.3	26.6	12.6	4.6	9.0	7.6	7.0	0.5	4.7
8	6.9	9.9	15.3	12.1	10.3	12.2	7.5	7.0	5.8
9	7.1	8.6	19.0	15.0	16.1	19.0	7.7	7.4	5.9
10	7.1	7.1	5.1	5.7	5.7	3.4	9.2	9.3	9.5
11	7.3	9.7	11.6	8.8	13.4	13.4	7.7	6.8	6.4
12	7.6	7.6	104.5	7.7	8.1	0.1	8.8	8.8	3.7
13	6.6	4.8	1.5	11.1	9.7	4.4	7.8	8.3	9.3
14	6.5	5.2	24.9	11.0	11.0	18.2	8.1	8.1	4.4
15	6.6	8.0	6.0	9.4	8.9	5.4	7.1	6.7	7.6
16	6.7	10.3	7.3	8.1	11.1	8.8	7.6	6.5	7.4
17	7.5	9.7	12.2	6.1	8.3	9.4	8.5	7.9	7.5
18	6.7	6.1	6.4	12.2	16.5	17.7	7.1	7.0	6.8
19	6.5	6.6	4.6	11.3	13.7	9.0	7.8	7.7	8.4
20	6.2	8.2	8.7	12.7	16.5	13.7	8.2	7.7	7.9

Table 4.6: True and estimated parameter values from 20 simulations of a GARCH(1,1) process with Gaussian innovations.

true parameter values and the estimated values, Maximum Likelihood outperforms RPLR in simulations 1, 2, 5, 9, 11 through 14 and 17, a total number of 9 simulations (45%). RPLR outperforms Maximum Likelihood in simulations 7 and 16 (10%), and the remaining 9 simulations (45%) are undecided as no estimation routine is closest to the true parameter values for all three parameters.

In order to compare the Maximum Likelihood and the RPLR estimation results of Table 4.6 across simulations, one may suggest to compute the total distance from the true parameter values to the estimated ones for each estimation method and use these two measures to compare the methods. This approach, however, have been abandoned because it is not possible to relate estimation errors of α_0 to estimation errors of α_1 or β_1. Is, for instance, the Maximum Likelihood estimate of simulation number 15 to be preferred over the RPLR estimate of the same simulation? The RPLR method gets the value of α_0 about right but is completely wrong with respect to α_1, whereas Maximum Likelihood fails to catch α_0 but is not far from correct with its estimate of α_1. Instead, the unconditional variances as estimated using Formula 3.3.8 are compared because the the unconditional variance is a functional of all three parameters and the results of this comparison is available in Table 4.7.

From the table it seems like RPLR does the best job in estimating the parameters of the simulated series. However, the general picture is heavily distorted my the near IGARCH model of isimulations number 10. If simulation number 10 were to be discregarded, the mean of the Maximum Likelihood errors would have a value around 2×10^{-4} lower than the value displayed in the table, whereas the mean of the RPLR errors would be around 2×10^{-5} lower, thus making Maximum Likelihood the best performing method. Anyway, the average of the RPLR estimates *is* closest to the average of the true values and RPLR *do* have the lowest sum af squared differences.

The results of simulation number 10 perfectly illustrates the problem of comparing the two estimation methods. Maximum Likelihood actually does a very good job in estimating the parameters, its estimates of α_0 and α_1 are identical to the true values, β_1 is just 0.2 above the true value whereas RPLR grossly underestimates the values of α_0 and α_1 and slightly overestimates β_1. But because the true model is close to IGARCH and Maximum Likelihood overestimates $\alpha_0 + \alpha_1$ just a little bit, its estimate of the unconditional variance is far above the true value.

The preliminary conclusion must be that Maximum Likelihood is the preferred estimation routine when simulated GARCH(1,1) series are to be estimated, at least hwen the true model is not very close to being unstationary.

Sim	True ($\times 10^4$)	ML ($\times 10^4$)	|True-ML| ($\times 10^6$)	RPLR ($\times 10^4$)	|True-RPLR| ($\times 10^6$)
1	14.9	14.3	60.9	14.3	60.2
2	7.8	8.0	12.9	7.9	7.4
3	3.7	3.5	20.7	3.4	24.3
4	6.2	6.3	4.6	6.3	5.0
5	3.4	3.6	20.1	3.6	17.4
6	5.8	5.9	13.0	5.8	5.5
7	2.9	3.1	22.1	2.8	12.7
8	5.1	5.1	6.5	5.1	5.4
9	9.1	9.1	6.6	8.7	48.1
10	31.3	70.2	3891.2	36.9	560.3
11	5.3	5.2	10.1	5.2	12.6
12	20.1	22.0	196.6	16.7	340.0
13	6.2	6.4	23.4	6.4	19.9
14	7.8	6.6	119.6	6.5	129.1
15	3.4	3.3	11.9	3.3	13.2
16	4.3	4.3	0.6	4.3	0.0
17	8.1	7.9	26.7	7.9	25.9
18	3.9	4.4	52.9	4.4	55.8
19	6.3	7.2	86.7	7.1	74.6
20	12.5	12.3	19.1	12.0	50.4
Mean ($\times 10^5$)			23.031		7.3390
Stdev ($\times 10^4$)			9.7719		1.5890

Table 4.7: Estimated unconditional variances using Maximum Likelihood and RPLR. Column 4 and 6 display the difference between the true parameter values and the Maximum Likelihood and RPLR estimates, respectively.

4.6.1.4 Simulated series - t-distributed innovations

When the Maximum Likelihood and the RPLR estimation routines was applied to the return series, an assumption was made that the relative poor performance of the Maximum Likelihood method is due to the fact that the empirical distributions has fatter tails than the assumed Gaussian distribution. To test this assumption, a simulation study have been carried out, estimating the parameters of 20 simulated GARCH(1,1) processes with innovations drawn from a standardized t-distribution[4] with four degrees of freedom. The Maximum Likelihood routine still "assumes" that the innovations are Gaussian. The simulation results are presented in Table 4.8.

The results in the table seems to give some support to the assumption. Here, Maximum Likelihood outperforms RPLR in simulations number 1, 2, 7, 11, 14 and 19, a total number of 6 simulations (30%) whereas RPLR outperforms Maximum Likelihood in simulations number 3, 5 and 18 (15%) with the remaining 11 simulations (55%) left undecided. Hence, it seems like RPLR has gained some performance relative to Maximum Likelihood, but, using this measure, Maximum Likelihood is still the best. However, with 55% of the simulations undecided more analysis is required before a conclusion is obtained.

The estimated unconditional variances are as displayed in Table 4.9.

This table supports the preliminary conclusion based on Table 4.8, the Maximum Likelihood routine seems to be performing a better job. Counting the number of simulations where Maximum Likelihood outperforms RPLR and vice versa yields $12 \Leftrightarrow 8$ in favor of Maximum Likelihood, meaning that it is not just the tendency of RPLR to be really wrong when its wrong as in simulation 7 that makes RPLR perform worse than Maximum Likelihood. The results of Table 4.9 are especially interesting as they seem to reject the preliminary assumption that the RPLR estimates are to be preferred when the conditional distribution of the returns has heavy tails. This rejection makes it difficult to explain why RPLR is better at predicting the conditional and unconditional variances of the real return series. Maybe it is just because the above simulation studies gives a wrong picture of the performance of RPLR, maybe it is because the t-distribution of the above simulations is not a suitable approximation of the heavy tails of the real returns or maybe there is something else about the return series than just the heavy tails that impede the performance of Maximum Likelihood. In a later subsection, the sensitivities of the estimation routines to initial values of parameters, conditional variance e.t.c. are analysed.

---

[4]Standardized (here): Unit variance and zero mean. Recall from Section 3.5 that is it *not* the student-t distribution.

Param	$\alpha_0 \times 10^5$			$\alpha_1 \times 100$			$\beta_1 \times 10$		
Sim	True	ML	RPLR	True	ML	RPLR	True	ML	RPLR
1	7.0	6.4	47.3	11.8	14.9	33.9	8.3	7.9	2.0
2	6.6	13.3	28.6	6.5	7.2	12.6	8.8	8.2	6.4
3	6.4	1.0	9.4	4.2	0.9	1.0	7.3	9.5	6.5
4	7.4	4.7	4.2	10.9	7.8	8.8	7.1	8.0	8.0
5	6.6	16.0	9.8	6.6	13.6	3.9	7.8	5.1	7.4
6	7.2	10.3	4.9	13.8	13.2	6.4	8.0	7.3	8.7
7	6.7	6.7	27.3	8.2	8.2	7.5	9.0	9.0	8.9
8	7.2	14.8	11.3	6.5	6.6	3.9	7.1	5.2	6.4
9	7.5	10.5	35.4	8.5	5.5	5.7	8.4	8.1	5.0
10	6.9	8.2	8.1	7.4	9.4	0.1	7.1	6.3	7.3
11	6.5	11.6	23.2	4.4	3.4	0.2	8.4	7.8	6.3
12	7.3	10.1	20.3	14.5	19.1	18.4	7.5	6.5	4.8
13	7.5	20.1	62.6	7.7	6.1	7.6	8.5	7.2	2.3
14	7.3	4.5	1.1	11.9	13.5	2.1	7.3	7.8	9.6
15	7.0	3.9	4.2	13.6	13.0	6.8	7.8	8.1	8.6
16	6.5	6.4	7.3	11.6	10.0	7.9	7.1	7.2	7.1
17	6.3	2.3	3.9	12.4	6.3	6.2	7.3	8.8	8.3
18	7.4	12.9	2.9	13.3	24.9	5.1	7.7	5.9	9.1
19	7.4	11.7	2.3	12.2	12.5	3.5	8.0	7.6	9.4
20	7.4	2.3	32.8	6.8	3.7	9.1	8.1	9.3	3.9

Table 4.8: True and estimated parameter values from 20 simulations of a GARCH(1,1) process with t-distributed innovations.

Sim	True $(\times 10^4)$	ML $(\times 10^4)$	|True-ML| $(\times 10^6)$	RPLR $(\times 10^4)$	|True-RPLR| $(\times 10^6)$
1	12.6	10.5	212.6	10.1	249.1
2	12.2	12.2	0.4	12.4	19.6
3	2.8	2.8	2.4	2.8	2.5
4	4.0	3.8	18.4	3.9	11.5
5	4.3	4.5	20.4	4.4	5.2
6	10.8	7.4	338.0	7.1	364.0
7	40.6	40.6	0.0	81.5	4088.9
8	3.3	3.5	24.4	3.5	23.3
9	9.6	8.0	155.1	7.9	162.9
10	3.2	3.0	25.3	3.0	24.4
11	5.5	6.4	85.4	6.4	82.5
12	6.7	6.2	49.5	6.0	64.1
13	10.0	9.0	100.1	9.0	98.3
14	4.9	5.1	24.7	4.7	19.6
15	8.5	6.9	158.7	6.2	230.7
16	3.8	3.6	19.7	3.5	24.9
17	4.3	3.8	47.5	3.7	57.0
18	7.7	8.0	25.7	7.1	62.4
19	9.9	9.9	8.3	10.2	33.8
20	6.0	6.6	56.0	6.3	30.3
Mean $(\times 10^5)$			6.8637		28.275
Stdev $(\times 10^5)$			8.6868		90.109

Table 4.9: Estimated unconditional variances using Maximum Likelihood and RPLR. Column 4 and 6 display the differences between the true parameter values and the Maximum Likelihood and RPLR estimates, respectively.

4.6.1.5   Return series – t-distributed innovations

Now, in the context of t-distributed innovations, it is illustrative to measure the performance of Maximum Likelihood relative to RPLR when Maximum Likelihood is implemented using the log-likelihood function corresponding to the t-distribution, i.e. the log-likelihood from Equation 4.2.7. The Maximum Likelihood estimates of GARCH(1,1) with t-distributed innovations are as displayed in Table 4.10. The row labeled df is the estimates degrees of freedom.

Series	HP	Sony	Mobil	Pepsi
α_0	1.1572×10^{-6}	5.9043×10^{-6}	4.4666×10^{-6}	1.5819×10^{-5}
	8.8560×10^{-7}	3.0074×10^{-6}	2.1789×10^{-6}	8.1411×10^{-6}
α_1	1.3189×10^{-2}	8.6821×10^{-2}	4.8135×10^{-2}	3.7531×10^{-2}
	4.0361×10^{-3}	2.2953×10^{-2}	1.4552×10^{-2}	1.6496×10^{-2}
β_1	9.8434×10^{-1}	8.9713×10^{-1}	9.1875×10^{-1}	8.9250×10^{-1}
	4.4285×10^{-3}	2.7851×10^{-2}	2.5870×10^{-2}	4.6918×10^{-2}
df	6.1946	4.1533	9.9503	5.8196
	9.3981×10^{-1}	4.6334×10^{-1}	2.4899	8.5450×10^{-1}
L_T	3.1458×10^3	3.5792×10^3	3.8842×10^3	3.5592×10^3

Table 4.10: Maximum Likelihood parameter estimates of GARCH(1,1) with t-distributed innovations applied to return series.

By comparing the results of Table 4.10 to the ones of Table 4.2, it is observed that the parameter estimates are actually quite different. Recall that the lower degree of freedom, the heavier the tails of the t-distribution and compare the df-estimates of Table 4.10 to the kurtosis values of the return series in Table 2.1. The higher the kurtosis, the lower the degrees of freedom. Also note that the lower degree of freedom, the closer are the estimates of Table 4.10 to the estimates of Table 4.2. For instance, there is almost no difference between the Mobil estimates of t-distributed innovations and the Mobil estimates of Gaussian distributed innovations.

In order to establish wheather the t-distributed innovations have enabled the GARCH(1,1) model to capture the heavy tails of the return series distribution, 1264 realizations from a GARCH(1,1) process with t-distributed innovations and with the exact same parameter values as those of Hewlett-Packard in Table 4.10 is plotted in Figure 4.10.

If Figure 4.10 is compared to the two plots in Figure 4.1, it might be possible to observe that the GARCH(1,1) model with t-distributed innovations is better at capturing the heavy tails of the empirical distributions than the GARCH(1,1) model
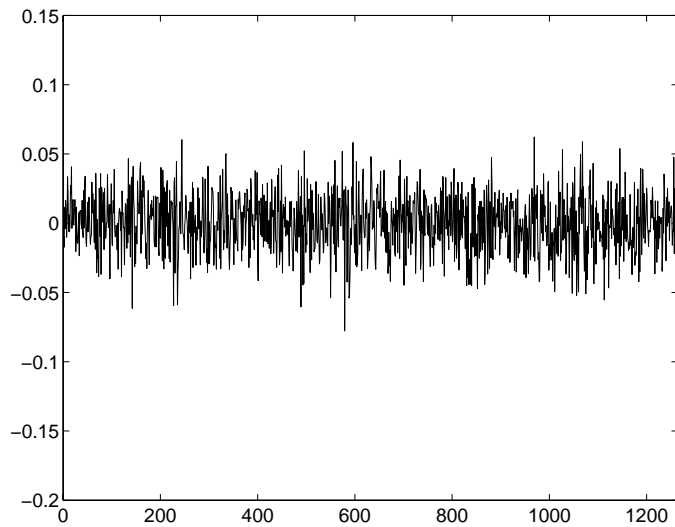
Figure 4.10: Realizations from a GARCH(1,1) process with t-distributed innovations using the parameter values of the second column of Table 4.10.

with Gaussian innovations. The kurtosis of the realizations from the t-GARCH model is 4.94, whereas the kurtosis of the Hewlett-Packard returns is 6.09 and the kurtosis of the realizations of the Gaussian-GARCH model is 3.11, numbers which indicate that the GARCH(1,1) model with t-distributed innovations is far better in capturing the heavy tails of the returns than the GARCH(1,1) with Gaussian innovations is. The kurtosis estimates also indicate, however, that exchanging the Gaussian distribution with a t-distribution may not be sufficient if the objective is to construct a model that produces distribution tails as heavy as the tails of the real life returns.

In Figure 4.11, the conditional variance estimates of Hewlett-Packard produced by the GARCH(1,1) model with t-distributed innovations are plotted above the conditional variance estimates produced by the GARCH(1,1) model with Gaussian innovations.

From the two plots in Figure 4.11 it is observed that the GARCH(1,1) model with t-distributed innovations is less affected by outliers, a clearly desirable property. The subject of influence of outliers on the conditional variance estimates will be further discussed in a later subsection when the EGARCH(1,1) process has been estimated. For now it suffices to say that the reason for the conditional variance estimates of the t-GARCH to appear more "calm" is that the heavier tails of the

Figure 4.11:   Estimated conditional variances of Hewlett-Packard by GARCH(1,1) with t-distributed innovations (top) and by GARCH(1,1) with Gaussian innovations (bottom).

t-distribution compared to the normal distributions allow the GARCH(1,1) model to explain outliers, outliers that could otherwise only be accounted for by drastic adjustments of the conditional variance.

4.6.2 RPLR – Initial values

To find out why RPLR seems to be behaving better than Maximum Likelihood on return series, but poor in the simulation studies, a closer look at the RPLR method is needed.
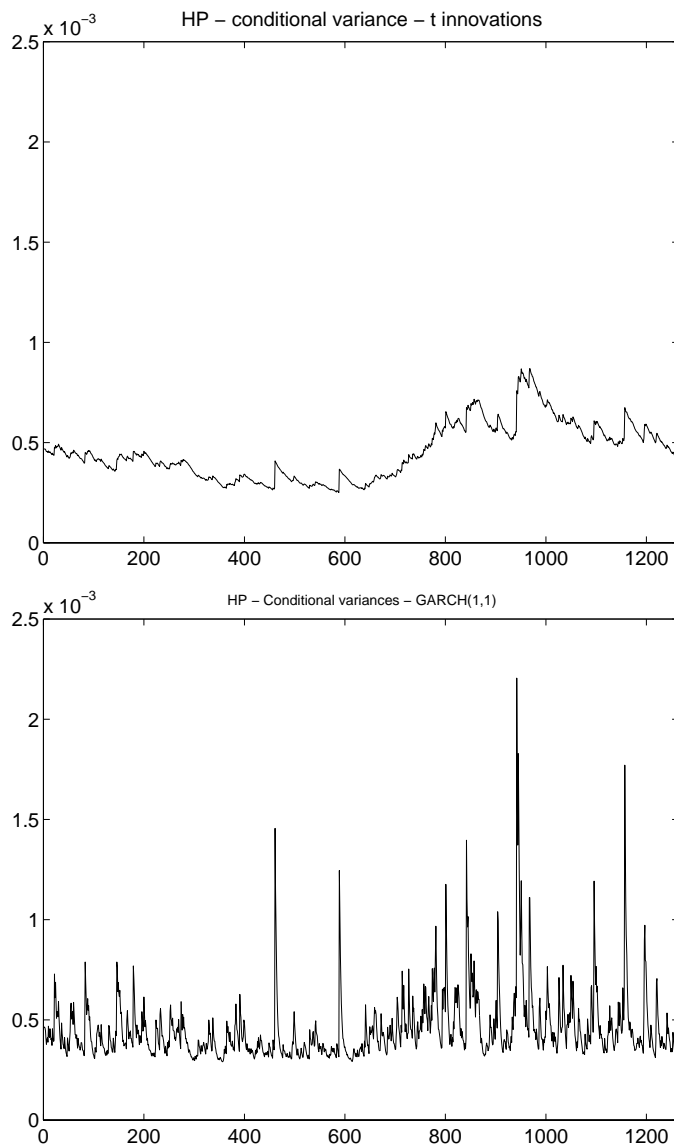
Recall that the RPLR method finds its parameter estimates by iterating over the RPLR updating scheme of Definition 4.2:

$$\epsilon_T(\hat{\theta}_{T-1}) = Y_T \Leftrightarrow X_T(\hat{\theta}_{T-1})\hat{\theta}_{T-1} \tag{4.6.2a}$$

$$\hat{\theta}_T = \hat{\theta}_{T-1} + R_T^{-1}(\hat{\theta}_{T-1})\epsilon_T(\hat{\theta}_{T-1})X_T(\hat{\theta}_{T-1}) \tag{4.6.2b}$$

$$R_T(\hat{\theta}_{T-1}) = R_{T-1}(\hat{\theta}_{T-1}) + X_T(\hat{\theta}_{T-1})X_T^\top(\hat{\theta}_{T-1}). \tag{4.6.2c}$$

and in the GARCH(1,1) case, $Y_T = \varepsilon_T^2$, $X_T^\top = \begin{bmatrix} 1 & \varepsilon_{T-1}^2 & h_{T-1} \end{bmatrix}$ and $\theta^\top = \begin{bmatrix} \alpha_0 & \alpha_1 & \beta_1 \end{bmatrix}$. With this choice of X, all the elements of X will be nonnegative at all times. This means that the elements of R will increase each time an iteration is performed and because of this, the difference between θ_T and θ_{T-1} will decrease when the number of performed iterations gets high.

4.6.2.1 Initial value of R

To start up the iteration, initial values of R, h and θ are needed. Theoretically this means that the RPLR method needs a total number of 12 initial values, far more than Maximum Likelihood. In practice, R may be chosen to be proportional to the identity matrix, $R_0 = kI$ meaning that the total number of initial values to be stipulated is 5. Still, however, the choice of R_0, that is k, is important. Because of the steady increase of R and the decrease of $|\theta_T \Leftrightarrow \theta_{T-1}|$, k will have to be assigned a small value if the parameter guess θ_0 is far from the true parameters. At the same time, k will have be larger if the initial parameter vector is close to the

true parameters, because otherwise the term $R_T^{-1}(\hat{\theta}_{T-1})\epsilon_T(\hat{\theta}_{T-1})X_T(\hat{\theta}_{T-1})$ of the θ_T updating equation (4.6.2c) is likely to "divert" θ_T by choosing a value of θ_T far from θ_{T-1}.

To illustrate how RPLR behave for different choices of the initial values, let us consider the parameter trajectory of the RPLR estimation of Hewlett-Packard, i.e. the estimation of the second column of Table 4.5. The initial values of this estimation are $k = 10^{-4}$, $h_0 = 4.473 \times 10^{-4}$ (the second order moment of the Hewlett-Packard returns), $\alpha_0 = 10^{-5}$, $\alpha_1 = 0.1$ and $\beta_1 = 0.9$.

Because the trajectories of the three parameters are very similar to each other (apart from the parameter value), it suffices to consider the trajectory of one parameter. The trajectory of β_1 of the first 50000 iterations is presented as the upper plot in Figure 4.12. Because the 1264 observations of the return series is too small a number for the RPLR estimation routine to converge, the iteration has been restarted a number of times, i.e. the values of R, θ and h have been preserved, but the routine has started over with return number 1. This explains the small spikes of the trajectory, spikes that occur each time the iteration restarts and moves from return number 1264 to return number 1. The large number of iterations may lead one to believe that the RPLR routine takes a long time to compute its estimates, but this is not the case. In fact it takes less time for RPLR to compute 100000 iterations than it takes the Maximum Likelihood routine to find the its estimate, even with the true parameters as starting values.

As seen from the plot, the first part of the trajectory is very turbulent whereas the trajectory is more predictable for higher iteration numbers. This corresponds to the above findings, where the difference $|\theta_T \Leftrightarrow \theta_{T-1}|$ was predicted to decrease as the number of performed iterations increases.

The lower plot in Figure 4.12 shows the same trajectory for the same number of iterations, but where k has been set to the value 0.1. As it is readily seen, the RPLR method not just barely changes the value of the parameter from one iteration to the next, it is also moving away from the true value.

Note that the situation of $k = 0.1$ not only illustrates the behaviour of the RPLR method when the initial value of k is too high. It also illustrates a situation where the monotonously increasing elements of the R matrix gets too high before the parameter value θ has converged. It is actually quite possible that this is what has happened to some of the estimations of the previously described estimation studies, but because of lack of storage space, it was not possible to save the parameter trajectories of the estimations, only the final values, and because of this it is not possible to find out exactly what went wrong.
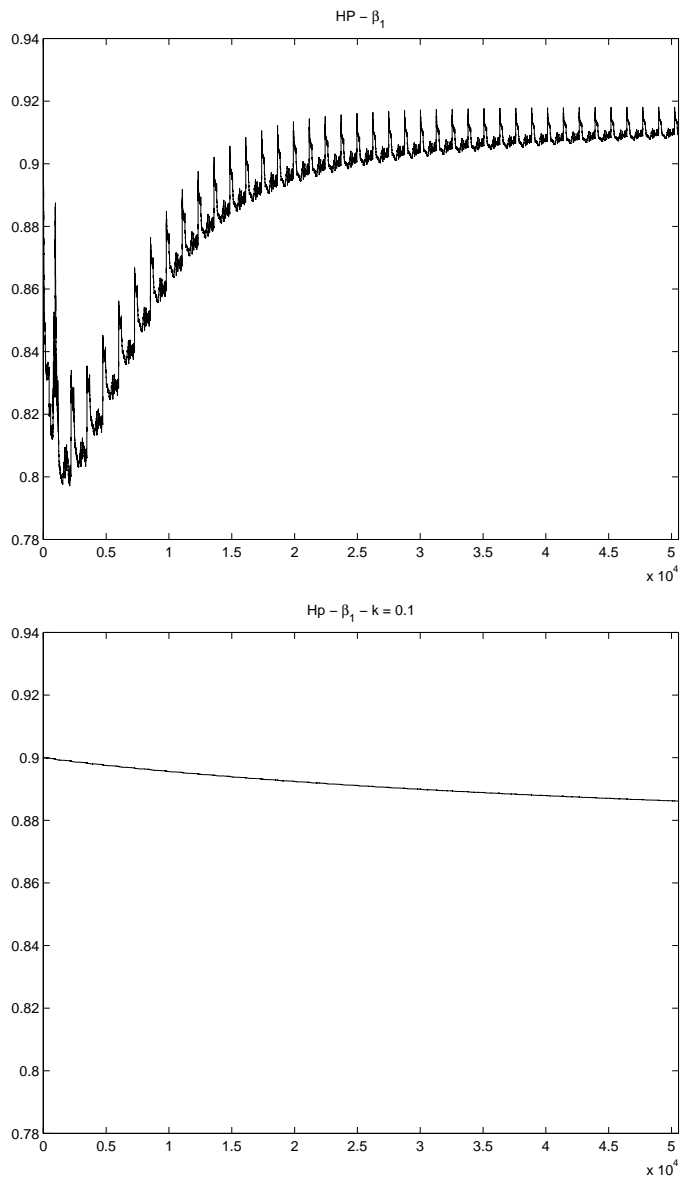
Figure 4.12: The trajectory of β_1 estimates for Hewlett-Packard. The upper plot shows the trajectory when $k = 10^{-4}$, the lower plot shows the trajectory when $k = 0.1$.

To show that a too high value of the matrix R is not the only thing that may cause the RPLR method to produce spurious parameter estimates, an estimation similar to the two estimations above but with $k = 10^{-12}$ have been performed. The resulting parameter estimate is $\alpha_0 = 7.3 \times 10^{-4}$, $\alpha_1 = 4.37 \times 10^{-2}$ and $\beta_1 = \Leftrightarrow 6.98 \times 10^{-1}$. What has happened is that the initial part of the trajectories have been so turbulent that it has pushed the final value of β_1 so far away from the initial value that it has become negative. To show that the estimation have actually converged, the trajectory of β_1 is displayed in Figure 4.13.
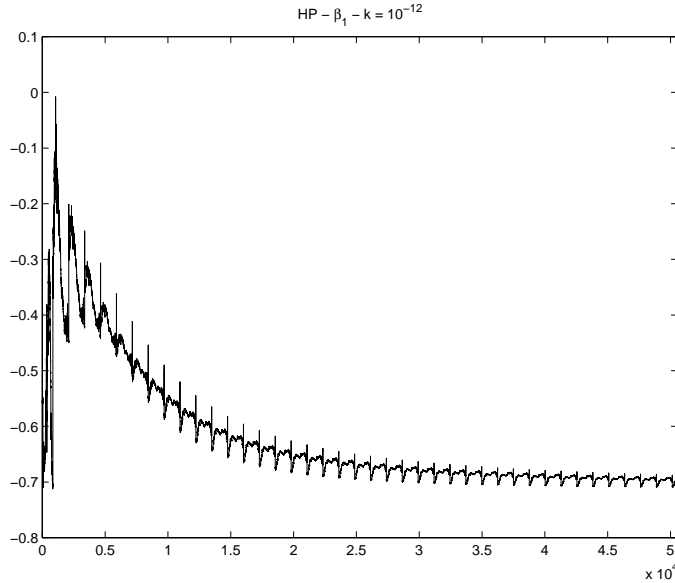
Figure 4.13: The trajectory of β_1 estimates for Hewlett-Packard, $k = 10^{-12}$.

The above analysis explains why it is so difficult to measure the performance of RPLR. There is no general optimal value of R_0, the optimal choice depends on the (unknown) distance between the true parameter values and θ_0. Furthermore, as the plot in Figure 4.13 illustrates, RPLR may in some cases converge to a wrong set of parameters. If this results in parameter values that violates the GARCH constraints, one may have an idea that something is wrong, but if no constraints are violated it may be very difficult to assess the validity of the obtained estimates. This is unlike Maximum Likelihood where, as shall later be described in detail, the validity of the obtained estimates may be asserted by e.g. an estimate of the Hessian because valid Maximum Likelihood estimates are known to maximize the log-likelihood.

The problems associated with the choice of the initial value of R naturally also

applies to simulation studies. Shall one pretend to have no prior knowlegde about the parameters and start the iteration with a low value k, or shall one use ones knowledge about the true parameter values and start the estimation close to the true parameter values with a higher value of k? In the previously described estimation studies, the first approach have been taken and k has been set to $\times 10^{-4}$ as it is clearly unnatural in a real life situation to know the parameter values prior to estimation, but this choice of k may be responsible for some of the poor RPLR estimation results of Table 4.6 and Table 4.8.

4.6.2.2   Initial values of GARCH(1,1) parameters

An analysis of the influence of the choice of initial parameter values θ_0 on the performance of the RPLR method is, as described above, difficult to perform because of the interaction between R_0, θ_0 and the true parameters. However, to give some kind of illustration of what the RPLR method is capable of, the parameters of Hewlett-Packard have been estimated using initial parameter values of $\alpha_0 = 0.01$, $\alpha_1 = 0.4$ and $\beta_1 = 0.5$, values which are actually quite far from the previously estimated values. The values of k and h_0 are 10^{-4} and 4.473×10^{-4} as above. The obtained parameter estimates of this study are, after about one million iterations, identical to those of Table 4.5 and the parameter trajectory of β_1 is plotted in Figure 4.14.

As it is readily seen from the plot, the RPLR routine have no problems finding the correct estimates, even with initial parameter values far from the true ones. If the Maximum Likelihood procedure is started with these values of the initial parameters, it fails to find the maximum of the log-likelihood.

4.6.2.3   Initial value of h

To measure the sentivity of the RPLR method to the choice of h_0, a number of estimations (not shown) have been performed on the Hewlett-Packard return series. The conclusion of this estimation study is that extremely high values of h_0 like $h_0 = 1$ or $h_0 = 0.1$ will push the RPLR procedure "off the track", thus yielding obscure estimation results, but "sensible" values of h_0 not far from the unconditional variance (even $h_0 = 0$) will always yield a proper estimation provided that R_0 and θ_0 are not poorly chosen. Hence, the RPLR method can be said to be insensitive to the choice of h_0.
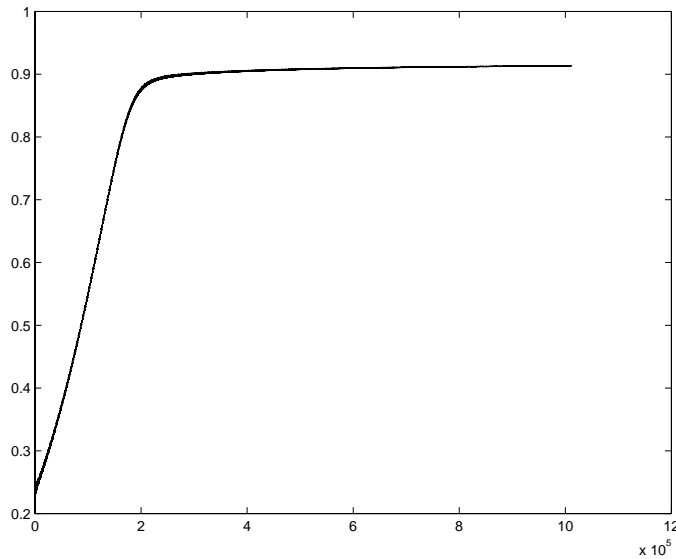
Figure 4.14: The trajectory of β_1 estimates for Hewlett-Packard, initial parameter values far from the true ones.

4.6.3 Maximum Likelihood – Initial values

The Maximum Likelihood routine needs values of the parameter vector θ and a value of h_0 for startup.

4.6.3.1 Initial value of h

To measure the influence of the initial value of h on the ability of Maximum Likelihood to obtain correct parameter estimates, an estimation study has been performed where the value of h_0 have been included in the Maximum Likelihood setup as a parameter. In this way, the Maximum Likelihood routine is allowed to change the value of h_0 if the stipulated value is suboptimal. The estimation results obtained from this estimation study are presented in Table 4.11.

Comparing the results of Table 4.11 to the results of Table 4.2 where h_0 was identical to the unconditional variance, $h_0 = \alpha_0/(1 \Leftrightarrow \alpha_1 \Leftrightarrow \beta_1)$, it is observed that the two sets of parameter estimates are slightly different. The largest change in parameter values seems to have occured for Sony where quite a lot of "parameter mass" have moved from α_0 to α_1. It is also observed that with α_1 of Sony as the only

Series	HP	Sony	Mobil	Pepsi
α_0	7.2395×10^{-5}	3.0587×10^{-6}	4.8343×10^{-6}	2.7275×10^{-5}
	2.8286×10^{-5}	1.2833×10^{-6}	1.7576×10^{-6}	1.0241×10^{-5}
α_1	1.0362×10^{-1}	6.4384×10^{-2}	4.7505×10^{-2}	7.3205×10^{-2}
	3.5513×10^{-2}	1.1953×10^{-2}	1.2506×10^{-2}	2.4132×10^{-2}
β_1	7.3949×10^{-1}	9.2883×10^{-1}	9.1589×10^{-1}	8.1422×10^{-1}
	8.6578×10^{-2}	1.3266×10^{-2}	2.0729×10^{-2}	5.7720×10^{-2}
h_0	2.2030×10^{-4}	1.2166×10^{-4}	2.0981×10^{-4}	3.3467×10^{-5}
	3.5059×10^{-4}	9.4419×10^{-5}	1.1716×10^{-4}	7.8642×10^{-5}
L_T	3.0954×10^{3}	3.4994×10^{3}	3.8691×10^{3}	3.5015×10^{3}

Table 4.11: Maximum Likelihood parameter estimates of GARCH(1,1) applied to return series with h_0 part of the estimation.

exception, all standard errors have been reduced. This is, however, not surpricing as some of the uncertainty is now constributed to h_0.

One of the main problems of the estimates of Table 4.2 was that the estimates were biased in the sence of the unconditional variance. To see if this is also the case of the estimates in Table 4.11, estimates of the unconditional variance for the two tables are presented in Table 4.12 along with the RPLR estimates and the second order moments of the four return series.

	HP	Sony	Mobil	Pepsi
Maximum Likelihood, estimated h_0	4.6144	4.5085	1.3206	2.4229
Maximum Likelihood, fixed h_0	4.6021	3.1281	1.3435	2.4095
RPLR	4.4702	2.5817	1.3133	2.3393
Empirical second order moment	4.4730	2.5845	1.3182	2.3394

Table 4.12: Estimates of the unconditional variance based on Maximum Likelihood estimation with h_0 included in the estimation (first row), Maximum Likelihood estimation with h_0 fixed (second row), RPLR estimation (third row) and the empirical second order moment of the return series (forth row). All estimates have been multiplied by 10^4 before they have been inserted into the table.

From the table it does not seems like including h_0 in the estimation improves the ability of Maximum Likelihood to estimate the unconditional variance. The unconditional variance estimate of Mobil is slightly better for an estimated h_0 than for a fixed h_0, but the estimates of Hewlett-Packard and Pepsi are slightly worsened when h_0 is included in the estimation and clearly, the unconditional variance

estimate of Sony is seriously negatively affected. Hence, the remarks made previously regarding the distraction of the Maximum Likelihood estimation routine by heavy tailed empirical distributions seem to apply also to Maximum Likelihood estimation including h_0.

An estimation study (not shown) where the very same 2×20 simulated series that were used in Table 4.6 and Table 4.8 have been estimated using Maximum Likelihood estimation with an estimated h_0 reveals that including h_0 in the estimation improves the performance measured in terms of the distance between estimated and true parameter values, but that the performance measured in terms of the estimates of the unconditional variance are slighty impeded when the parameters of GARCH(1,1) simulations with t-distributed innovations are estimated. These findings indicate that maybe h_0 should not be included in the estimation procedure if special weight is put upon a correct estimate of the unconditional variance.

An alternative to the two above approaches of including h_0 in the estimation or fixing h_0 at the unconditional variance is to fix h_0 at a value different from the unconditional variance. To measure the influence of the choice of h_0 in this case, a study have been performed where the GARCH(1,1) parameters of the Hewlett-Packard returns have been estimated using 7 different values of h_0. The parameter estimates obtained in this way are presented in Table 4.13.

h_0	α_0	α_1	β_1
0	8.2563×10^{-5}	1.1474×10^{-1}	7.0747×10^{-1}
1×10^{-4}	7.4798×10^{-5}	1.0627×10^{-1}	7.3194×10^{-1}
2×10^{-4}	7.2555×10^{-5}	1.0379×10^{-1}	7.3900×10^{-1}
4×10^{-4}	7.2729×10^{-5}	1.0403×10^{-1}	7.3825×10^{-1}
1×10^{-3}	7.8459×10^{-5}	1.1024×10^{-1}	7.1978×10^{-1}
1×10^{-2}	1.1721×10^{-4}	1.4522×10^{-1}	6.0286×10^{-1}
1×10^{-1}	1.8080×10^{-4}	1.7973×10^{-1}	4.3324×10^{-1}

Table 4.13: Maximum Likelihood parameter estimates of GARCH(1,1) applied to Hewlett-Packard for different choices of h_0.

Based on these estimates it is without much doubt concluded that the choice of h_0 has a serious impact on the parameter estimates. Because of this it is recommended that h_0 is included in the estimation procedure unless the value of the fixed h_0 has been especially accounted for. The table illustrates that fixing h_0 at the unconditional variance (compare with Table 4.12) is not optimal, though not highly inappropriate.

4.6.3.2 Initial value of GARCH(1,1) parameters

The Nelder-Mead simplex routine that is used to maximize the log-likelihood function needs a guess at the parameters for startup. When the algorithm terminates is has either located a global maximum, located a local maximum or it has been unable to locate a maximum. All but the first one of these situations are of course unwanted.

Because termination of the Nelder-Mead simplex routine does not itself guarantee that a global maximum has been located, it is important to be able to establish which one of the above three situations that are present.

Whether it is the last one, i.e. no maximum located, or one of the first ones, i.e. a global or a local maximum located, may established by computing e.g. an estimate of the Hessian. If the computed parameter vector is a maximum, all elements of the Hessian must be positive. Often it is sufficient to compute two values of the log-likelihood, one using the obtained parameter vector and one using the obtained parameter vector mutiplied by a scalar close, but not equal, to one. If the obtained parameter vector does not maximize the log-likelihood, the latter of the two values of the log-likelihood will ofteb be larger than the first.

If the parameter estimate at hand appears to be maximizing the log-likelihood, it is important to assess whether the maximum is local or global, but this is difficult in a large space of possible parameter values. It is therefore interesting to know whether the log-likelihood surface contains local maxima or not.

To answer this question, one may start by examining the literature in the field of GARCH Maximum Likelihood estimation, but, at previously mentioned, the literature in this field is very sparse, though some researchers claim that the log-likelihood surface has multiple local maxima, c.f. (Bisgaard 1998). Another approach is to consider the derived log-likelihood $\partial L_T / \partial \theta$ and analyse whether there is more than one solution to the equation $\partial L_T / \partial \theta = 0$. This aproach leads nowhere because the equation contains several unknown quantities such as $\partial h_0 / \partial \theta$ and the like. Remaining is the conceptually simple but laborious method of scanning the entire parameter space for log-likelihood maxima.

Using the last of these methods, an extremely careful investigation of the log-likelihood surface of Hewlett-Packard with h_0 fixed at the unconditional variance have been performed. Values of the log-likelihood have been computed for each combination of α_0, α_1 and β_1 where α_0 have been equal to each of the 42 equidistant values between 6.2×10^{-5} and 7.84×10^{-5}, $alpha_1$ have been equal to each of the 226 equidistant values between 0.06 and 0.15 and where β_1 have been equal

to each of the 626 equidistrant values between 0.7 and 0.95, a total number of 5941992 different parameter vectors.

This huge estimation study discovered no local minima. This means that either the log-likelihood has no local maxima are present, that the maxima are so local that they where not captured or that local maxima do exist, but outside the investigated part of the parameter space.

To illustrate the appearance of the log-likelihood surface, the pseudo likelihood surface around the Maximum Likelihood estimate in the $(\tilde{L}, \alpha_1, \beta_1)$ space with α_0 fixed at the value of the minimum is plotted in Figure 4.15. Recall that Nelder-Mead is actually a minimization routine and that the parameter estimates are thus found at the minimum of the pseudo log-likelihood surface.
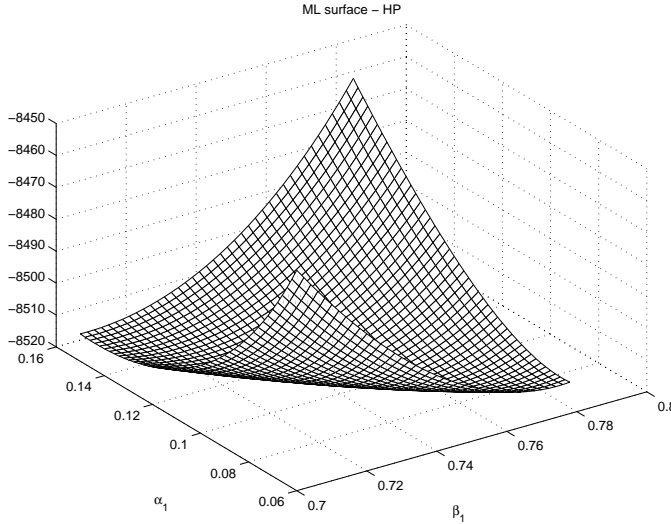
Figure 4.15: The surface of the pseudo log-likelihood around the global minimum in the $(\tilde{L}, \alpha_1, \beta_1)$ space.

From the plot it is observed that the pseudo log-likelihood surface is actually quite steep in the $\alpha_1 + \beta_1$ direction, whereas it is quite flat in the $\alpha_1 \Leftrightarrow \beta \Leftrightarrow 1$ direction. This corresponds to a correlation in the parameter space as estimated by the inverse of the Hessian in the case of Hewlett-Pavkard which yields a correlation matrix of

$$R = \begin{bmatrix} 1.0000 & 0.5269 & \Leftrightarrow 0.9463 \\ 0.5269 & 1.0000 & \Leftrightarrow 0.7478 \\ \Leftrightarrow 0.9463 & \Leftrightarrow 0.7478 & 1.0000 \end{bmatrix} \tag{4.6.3}$$

where the parameter order is α_0, α_1, β_1.

It is likely to believe that it is this substantial correlation that causes estimation problems rather than local maxima of the log-likelihood function.

The estimation study have been repeated with a likehood function corresponding to t-distributed innovations, again applied to the Hewlett-Packard returns. This estimation confirms that there are no local maxima (except the global one) in the investigated part of the parameter space.

That there does not seem to be more than one local maximum of the log-likelihood does not mean that Maximum Likelihood estimation always occurs without problems. Many times, the minimization procedure takes a long time to find the minimum or terminates at a point that is not a minimum. For instance, when the GARCH(1,1) parameters of the four return series were estimated with h_0 as part of the estimation (Table 4.11), the Nelder-Mead simplex routine only succeeded in finding the correct minima of Hewlett-Packard and Mobil given starting values of $\alpha_0 = 0.001$, $\alpha_1 = 0.3$, $\beta_1 = 0.3$ and h_0 equal to the standard deviation of the return series in focus. The parameter estimates of Sony and Pepsi were only obtained after the numerical routine had been given the Mobil estimates as starting values, i.e. starting values much closer to correct parameters. Estimating the parameters of a GARCH(1,1) model with t-distributed innovations fitted to return series data is very difficult. The numerical routine barely manages to find the maximum, ven when the starting values of the parameters are very close to the final ones.

4.6.3.3   Number of available observations

The 1264 observations that has been available for the studies described in this thesis corresponds to five yearls of daily observations. This is quite a lot, and maybe it is incorrect to assume that the GARCH parameters are unchanged for such a long period. To investigate the validity of this assumption, each of the four return series have been divided, first in two parts and then in four parts each, and a GARCH(1,1) model have been fitted to each of these subseries. The parameter estimates obtained in this way are presented in Table 4.14. For brevity, the standard errors of the estimates have been left out.

As it is immediately observed, all four return series exhibits substantial differences from one subseries to another, but it is difficult to know whether these differences are caused by "true" changes in the parameters or if they occur because of suboptimalities. For instance, by looking at Figure 2.1 it is observed that the returns series contains a number of extreme observations. The tendency of a GARCH model to

Subseries	Parameter	HP	Sony	Mobil	Pepsi
1-632	α_0	–	2.14×10^{-5}	8.28×10^{-6}	3.83×10^{-5}
	α_1	–	8.70×10^{-2}	4.78×10^{-2}	8.57×10^{-2}
	β_1	–	8.31×10^{-1}	8.78×10^{-1}	7.18×10^{-1}
633-1264	α_0	1.57×10^{-4}	1.04×10^{-6}	3.55×10^{-6}	3.07×10^{-5}
	α_1	1.12×10^{-1}	5.42×10^{-2}	4.00×10^{-2}	5.81×10^{-2}
	β_1	6.22×10^{-1}	9.43×10^{-1}	9.38×10^{-1}	8.34×10^{-1}
1-316	α_0	8.88×10^{-5}	3.01×10^{-5}	1.73×10^{-5}	2.97×10^{-5}
	α_1	2.69×10^{-2}	1.64×10^{-1}	4.82×10^{-2}	6.02×10^{-2}
	β_1	7.32×10^{-1}	7.10×10^{-1}	8.15×10^{-1}	7.75×10^{-1}
317-632	α_0	–	5.72×10^{-6}	5.27×10^{-6}	4.11×10^{-5}
	α_1	–	6.52×10^{-2}	3.34×10^{-2}	1.12×10^{-1}
	β_1	–	9.26×10^{-1}	9.10×10^{-1}	6.95×10^{-1}
633-948	α_0	1.37×10^{-4}	–	3.84×10^{-6}	6.34×10^{-5}
	α_1	8.00×10^{-2}	–	3.37×10^{-2}	2.08×10^{-1}
	β_1	7.08×10^{-1}	–	9.42×10^{-1}	4.82×10^{-1}
949-1264	α_0	–	1.63×10^{-6}	3.44×10^{-6}	–
	α_1	–	7.47×10^{-2}	4.45×10^{-2}	–
	β_1	–	9.16×10^{-1}	9.33×10^{-1}	–

Table 4.14: Maximum Likelihood parameter estimates of GARCH(1,1) applied to subseries of the returns. Table cells containing – corresponds to unconverged estimations.

produce extreme observations depends on its parameters and vice versa, so if the estimation routine observe no or only a few outliers in the subseries it considers, it may "believe" that the subseries it actually generated by another GARCH model than the true one, even if the real reason for absence of outliers is that the probability of the true model generating an outlier within the observations of the subseries is less than 1.

The results in Table 4.14 shows that one should be careful when using financial time series and GARCH. It is likely that the model parameters changes from one year to the next or that a large number of observations are needed to obtain a reliable parameter estimate.

Parameters that changes over time calls for a recursive estimation routine. Unfortunately, RPLR adapts too slow to parameter changes because of its low speed of convergence. To increase this speed, it may be suggested to apply forgetting and gain to the RPLR method, as described in (Ljung 1987, chapter 11). Unfortunately,

forgetting and gain leads nowhere when the RPLR method is applied to GARCH. It appears to be impossible to choose a forgetting factor that is not either so high that it prohibits fast adaptation, or so low that the RPLR method do not converge.

4.6.4 EGARCH(1,1)

The last estimation results to be presented are those of EGARCH(1,1) applied to the four return series. In the sections above it has already been demonstrated how to analyse obtained estimates and because of this the analysis of the EGARCH parameter estimates will be brief.

The estimation results obtained when fitting an EGARCH(1,1) model to the four return series are presented in Table 4.15 along with the associated standard errors. Taught by experience, the estimation procedure have been designed so that the initial value h_0 of the conditional variance has been included in the estimation.

Recall from Chapter 3 that not all parameters of EGARCH are identifiable and because of this, row number two and three of the table contain estimates of parameter products instead of sole parameters.

	HP	Sony	Mobil	Pepsi
α_0	$\Leftrightarrow 8.6871 \times 10^{-2}$	$\Leftrightarrow 9.6209 \times 10^{-2}$	$\Leftrightarrow 4.0000 \times 10^{-1}$	$\Leftrightarrow 8.8182 \times 10^{-1}$
	6.9589×10^{-4}	6.6728×10^{-4}	7.0449×10^{-4}	2.7105×10^{-3}
$\alpha_1\omega$	$\Leftrightarrow 3.5310 \times 10^{-2}$	$\Leftrightarrow 1.1248 \times 10^{-2}$	$\Leftrightarrow 4.0552 \times 10^{-2}$	$\Leftrightarrow 1.2725 \times 10^{-2}$
	1.1097×10^{-2}	9.5178×10^{-3}	1.8028×10^{-2}	1.2215×10^{-2}
$\alpha_1\lambda$	5.0479×10^{-2}	1.2546×10^{-1}	1.1714×10^{-1}	1.8584×10^{-1}
	1.2660×10^{-2}	2.3452×10^{-2}	2.4935×10^{-2}	3.5396×10^{-2}
β_1	9.8828×10^{-1}	9.8697×10^{-1}	9.5485×10^{-1}	8.9354×10^{-1}
	6.5798×10^{-5}	2.6703×10^{-4}	7.6948×10^{-5}	2.4223×10^{-4}
h_0	$\Leftrightarrow 7.7540$	$\Leftrightarrow 9.0733$	$\Leftrightarrow 8.4785$	$\Leftrightarrow 9.6049$
	3.7075×10^{-1}	7.7434×10^{-1}	6.6615×10^{-1}	8.3648×10^{-1}
L_T	3.1086×10^3	3.4995×10^3	3.8720×10^3	3.5092×10^3

Table 4.15: Maximum Likelihood parameter estimates of EGARCH(1,1) applied to return series.

By comparing the last row of Table 4.15 to the last row of Table 4.11 it is noted how close the values of the log-likelihood of EGARCH(1,1) are to the values of the log-likelihood of GARCH(1,1).

The RPLR estimates of EGARCH(1,1) applied to the four return series are actually

quite different from the GARCH(1,1) estimates. The RPLR estimates are presented in Table 4.16.

	HP	Sony	Mobil	Pepsi
α_0	$\Leftrightarrow 7.3347 \times 10^{-1}$	$\Leftrightarrow 1.0194$	$\Leftrightarrow 4.2214 \times 10^{-1}$	$\Leftrightarrow 8.6764 \times 10^{-1}$
$\alpha_1 \omega$	$\Leftrightarrow 1.1195 \times 10^{-1}$	2.2286×10^{-2}	$\Leftrightarrow 1.2547 \times 10^{-2}$	2.8303×10^{-2}
$\alpha_1 \lambda$	1.7426×10^{-1}	3.5814×10^{-1}	1.3179×10^{-1}	2.0117×10^{-1}
β_1	9.0573×10^{-1}	8.7900×10^{-1}	9.5255×10^{-1}	8.9246×10^{-1}

Table 4.16: RPLR parameter estimates of EGARCH(1,1) applied to return series.

When the two sets of parameter estimates are so far from each other, it is important to find out which one is correct, if any. For the GARCH(1,1) parameter estimates earlier described, a part of this validation was a comparison of the unconditional variance estimates. Unfortunately, no explicit expression for the unconditional variance of an EGARCH process has been derived, but it is possible to use the mean of the conditional variances instead. These measures are presented in Table 4.17

	HP	Sony	Mobil	Pepsi
Maximum Likelihood	4.464	2.621	1.320	2.348
RPLR	40.200	31.052	51.295	32.486
Empirical second order moment	4.473	2.585	1.318	2.339

Table 4.17: Estimates of the unconditional variance based on mean of the conditional variance estimates of Maximum Likelihood (first row), mean of the conditional variance estimates of RPLR (second row) and the empirical second order moment of the return series (third row). All estimates have been multiplied by 10^4 before they have been inserted into the table.

From the table it seems evident that RPLR has got the parameter estimates completely wrong. Investigations of the RPLR parameter trajectories (not shown) confirms this impression, the RPLR is not able to converge, even for a substantial number of iterations. Hence, the RPLR estimates will be disregarded.

The Maximum Likelihood based estimates of the unconditional variance are more precise than the corresponding estimates of GARCH(1,1), c.f. Table 4.12. This may indicate that the EGARCH(1,1) model is easier to estimate and/or that the EGARCH(1,1) model provides a better description of the return series data than GARCH(1,1) does. To investigate the behaviour of the estimated EGARCH(1,1) process, the conditional variance estimates of Hewlett-Packard are presented as the

upper plot in Figure 4.16 and for comparison, the corresponding estimates from GARCH(1,1) are plotted below.

From the two plots it seems like the trajectory of the EGARCH estimates of the conditional variances is a smoothed version of the trajectory of the GARCH esti- mates. Whether to rely on the GARCH or EGARCH estimates may be a matter of preferences, but one should take into account that the apparant higher adaptiv- ity of the GARCH(1,1) has costs. The large spike in the trajectory of GARCH at number 942 occurs because the Hewlett-Packard returns have a large spike at number 941, not because GARCH(1,1) possess a divined capability of predicting outliers. Because of this, the higher ability of GARCH to follow changes in the returns compared to EGARCH makes GARCH easy to disturb by single extreme observations.

The upper plot in Figure 4.16 does not look very different from the plot of condi- tional variance estimates of a GARCH(1,1) model with t-distributed innovations in Figure 4.11. This resemblance leads one to believe that EGARCH is less sensitive to the choice of the conditional distribution than GARCH is.

Recalling Chapter 2, the EGARCH process was motivated by the so-called leverage effect and so, it may be interesting to see if EGARCH has been able to capture the leverage effect of the returns. Since it was only possible to see signs of the leverage effect in the returns of Hewlett-Packard, these returns will be the only one considered here.

As the leverage effect includes a tendency of the conditional variance to fall in responce to an increase and fall in response to a decrease of the stock price, the ability of a volatility model to capture the leverage effect may be measured by comparing conditional variance estimates folowing positive returns to conditional variance estimates following negative returns. If leverage effect is present in the returns, the latter of the two conditional variance estimates should be larger. The mean of the conditional variance estimates of the Maximum Likelihood estimated EGARCH(1,1) model is 4.3636×10^{-4} for conditional variance estimates fol- lowing positive returns and 4.5619×10^{-4} for estimates following negative re- turns. The corresponding means for the RPLR estimated GARCH(1,1) model[5] is 4.4777×10^{-4} and 4.4546×10^{-4}, respectively. Based on these numbers it is tempting to conclude that EGARCH *has* been able to capture the leverage effect of Hewlett-Packard. It should be noted, however, that the material supporting this conclusion is of low quality. First of all, the presence of the leverage effect in the

---

[5]The RPLR estimated GARCH(1,1) model has been considered because the Maximum Likeli- hood estimated model is biased with respect to the conditional variance estimates.
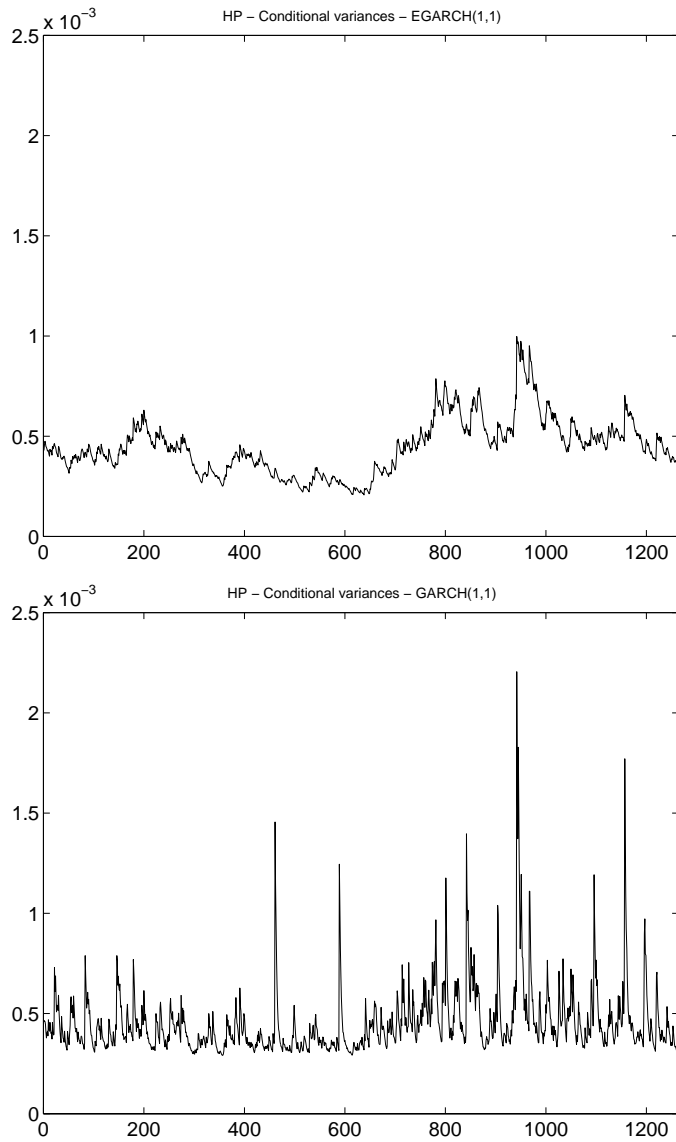
Figure 4.16: Estimated conditional variances of Hewlett-Packard by EGARCH(1,1) (top) and GARCH(1,1) (bottom).

Hewlett-Packard returns was difficult to assess in the first place, and the two above stated means of the EGARCH conditional variance estimates may not be significantly different from each other. In addition, only one return series have been investigated, a very sparse data material.

4.7 Summary

The above section presented a number of estimation results, results obtained using a lot of different models, estimation methods and initial values. Because the results where produced under very different conditions, a summary may be needed.

4.7.1 Estimation procedure

The performance of the Maximum Likelihood estimation procedure versus RPLR has been measured using the four return series as well as simulated data.

4.7.1.1 Maximum Likelihood

The Maximum Likelihood estimation procedure appears to be a reliable estimation procedure, valid for GARCH as well as EGARCH. It is easy to verify that parameter estimates obtained using the Maximum Likelihood procedure are indeed maximizing the log-likelihood function, and standard errors of the obtained estimates are readily available. Parameter constraints are easily incorporated into the Maximum Likelihood estimation procedure.

The Maximum Likelihood procedure seems to get into problems, if the conditional distribution of returns is different from the one assumed by the estimation procedure. If this is the case, Maximum Likelihood is likely to produce parameter estimates that results in a biased estimate of the conditional variance.

The Maximum Likelihood precedure is very sensitive to the stipulated value of the initial conditional variance. Because of this, it is recommended that the initial conditional variance is included in the estimation procedure, though this may increase the sensitivity of the estimation procedure to the choice of the initial parameter values.

4.7.1.2 RPLR

The RPLR estimation procedure is difficult to manage. Unless the RPLR routine is started with a carefully chosen initial value of the matrix R, the estimation procedure may very well converge to wrong parameter estimates. This is a problem as it is difficult to assess the validity of parameter estimates obtained using RPLR. The RPLR procedure does not provide standard errors of the obtained estimates, and it is not possible to incorporate parameter constraints into the estimation procedure. The RPLR procedure does not seem to be able to estimate the parameters of an EGARCH(1,1) model.

In order for the RPLR estimation to converge, quite a few iterations are needed. This is not a problem with todays amount of available computer resources, but is a problem with regard to the recursive nature of the RPLR procedure. Often, a recursive estimation procedure has the advantage of its ability to adjust to changes in the model parameters, changes that occur over time. But, because the RPLR procedure need so many observations to converge, it will take an awful lot of time before the procedure has adjusted to a parameter change.

The RPLR may have two advantages compared to Maximum Likelihood. First, RPLR does not assume a particular return series distribution and because of this it does not suffer from the problems related to a wrongly specified distribution. Second, it seems like RPLR, carefully nursed, is better than Maximum Likelihood in estimating the parameter of GARCH if the initial guess on the parameter values is far from the true parameter values.

The RPLR method is insensitive to the stipulated value of the initial conditional variance.

4.7.2 Arch models

In the previous section, three different Arch models were fitted to return series and simulated data. The estimated models were GARCH(1,1) with Gaussian innovations, GARCH(1,1) with t-distributed innovations and EGARCH(1,1) with Gaussian innovations.

4.7.2.1 GARCH(1,1)

The GARCH(1,1) model is, regardless of the type of the conditional distribution, a good choice when the autocorrelation structure of return series is to be described. When it comes to the heavy tails of the return series distribution, the hoice is less obvious. A GARCH(1,1) model with Gaussian innovations is clearly unable to cope with the heavy tails, the heavy tails causes the estimate of the unconditional and conditional variances to be biased and the conditional variance estimates to be heavily disturbed by extreme returns. The problems associated with the heavy tails do not seem to apply to the GARCH(1,1) model with t-distributed innovations, not in the same degree at least. Unfortunately, the latter model is very difficult to estimate and because of this, it is not possible to recommend one model at the expence of another. If, however, it is possible to fit a GARCH(1,1) model with t-distributed innovations to the data at hand, this model is to be preferred.

4.7.2.2 EGARCH(1,1)

The EGARCH(1,1) model has not been as closely examined as the GARCH(1,1) model. However, the EGARCH(1,1) model appears to be less sensitive to the choice of conditional distribution. Some evidence of an ability to capture the leverage effect have been discovered, but the data material supporting this assumption is sparse.

4.7.3 Future work

Because estimation studies takes a very long time to perform, it has only been possible to investigate a subset of the many adjustable elements of Arch models and the RPLR and Maximum Likelihood estimation procedures. A few examples of what I would have investigated if more time had been available is provided below.

1. Even the t-distribution was not able to capture the heavy tails of the return series distribution and thus it would be interesting to look for other conditional distribution candidates, e.g. the Generalized Error Distribution of (Nelson 1991).

2. The RPLR method had the benefit of being able to estimate GARCH(1,1) parameters without assuming a conditional distribution, but this means that

the method was not able to estimate any distribution parameters. It would be interesting to try to combine the RPLR method with a method to estimate distribution parameters to see if these two methods in cooperation provides better estimates than Maximum Likelihood.

3. Stochastic volatility models are often used for option pricing. In a financial market where even a small deviation from the fair price allows for arbitrage trading, it would be interesting to investigate the impact of parameter standard errors on the option price estimate.

4. Two recursive methods were derived, RPLR and RPEM. RPEM does not fit to GARCH and EGARCH, but it would be interesting to analyse its behaviour when applied to ARCH.

5. Only GARCH(1,1) and EGARCH(1,1) were investigated. Though higher order models are difficult to estimate, it might be interesting to see if models like GARCH(1,2) or GARCH(2,1) possess properties that more closely resembles those of return series.

Chapter 5

From discrete to continuous time

As mentioned in the introduction to this thesis, continuous time volatility models are preferred to discrete time models in situations where the models have to be dealt with and/or justified theoretically. For instance, relatively complicated tasks like finding the (theoretical) fair price of an option are best performed using a continuous time volatility model. This again means that when the problem at hand is to find the fair option price, fitting a discrete time volatility model to the stock price observations may not provide the solution.

Continuous time models, however, are often difficult to estimate, partly because observations occur in discrete time, not continuous time.

Quite a few researchers have tried to remove this trade-off between the two model kinds by bridging the gap between discrete and continuous time. These attempts include the *indirect inference* approach by (Gourieroux, Monfort & Renault 1993), the temporal aggregation approach by (Drost & Nijman 1993) and (Drost & Werker 1996) and the approximating augmented GARCH process by (Duan 1997). These different approaches all use a fitted discrete time model like Arch to infer or deduce an underlying continuous time stochastic volatility model.

The indirect inference approach seeks to find a continuous time stochastic volatility model that, observed at discrete time instants, yields the same discrete time model parameter estimates as the real data at hand. The indirect inference approach requires a lot of potential continuous time stochastic volatility models to be

135

simulated, and is thus very time consuming.

The temporal aggregation approach seeks to establish an algebraic relationship between discrete time GARCH parameters and continuous time stochastic volatility model parameters, thus allowing the continuous time stochastic volatility model parameter estimates to be computed as soon as discrete time model parameter estimates are available.

The approximating augmented GARCH approach by Duan seeks, like the temporal aggregation approach, to establish an algebraic relationship between the parameters of the two model kinds. The *modus operandi* of Duan is, however, very different from the approach taken in temporal aggregation.

The first two of the above mentioned approaches have been carefully studied by (Bisgaard 1998). Thus, this thesis only deals with the approximating augmented GARCH approach by Duan. Before diving into his approach, though, a brief introduction to continuous time volatility models is provided.

5.1   Continuous time volatility models

This section requires some basic knowledge of stochastic calculus. For an introduction to the subject, see (Madsen, Nielsen & Baadsgaard 1998, chapter 7).

5.1.1   Definition

In their famous paper (Black & Scholes 1973), Black and Scholes suggested to model the price process of a stock by the diffusion process

$$dS_t = \alpha S_t dt + \sigma S_t dW_t \tag{5.1.1}$$

where α and σ are parameters and W_t is a standard Wiener process. The parameter σ defines the volatility of the stock.

As previously stated, the model $(5.1.1)$ soon turned out to be a too simple description of the stock price process, in particular because of its assumption of constant volatility. Because of this, the Black-Scholes model is often enlarged by allowing the volatility to vary according to its own diffusion process, hence introducing the *continuous time stochastic volatility model*.

$$dS_t = \alpha(S_t, \theta) dt + \sigma_t dW_{1,t} \tag{5.1.2a}$$
$$d\varphi(\sigma_t) = a(\sigma_t, S_t, \theta) dt + b(\sigma_t, S_t, \theta) dW_{2,t} \tag{5.1.2b}$$

where φ is some mapping of σ_t and $W_{1,t}$ and $W_{2,t}$ are correlated Wiener processes with correlation coefficient ρ and θ is a parameter vector. The function α describe the drift of the stock price, the functions a and b describe the instantaneous conditional mean and standard deviation of the volatility process σ. The model (5.1.2) constitutes the basis of the definition of continuous time stochastic volatility, which is defined as σ_t of (5.1.2). In applications, a less general setup than (5.1.2) is often considered, namely

$$dS_t = \alpha S_t dt + \sigma_t S_t dW_{1,t} \tag{5.1.3a}$$
$$d\varphi(\sigma_t) = a(\sigma_t, S_t, \theta) dt + b(\sigma_t, S_t, \theta) dW_{2,t} \tag{5.1.3b}$$

where α is a constant parameter.

5.1.2 Common models

The literature describes a number of different models fitting into (5.1.3). Some of the more prominent of these are

5.1.2.1 The Hull - White model

In a paper that discussed the pricing of options under stochastic volatility, (Hull & White 1987) proposed the model

$$dS_t = \alpha S_t dt + \sigma_t S_t dW_{1,t} \tag{5.1.4a}$$
$$d\sigma_t^2 = a\sigma_t^2 dt + b\sigma_t^2 dW_{2,t} \tag{5.1.4b}$$

where α, a and b are constants.

5.1.2.2 The model of Scott, Stein & Stein and Heston

In (Scott 1987), (Stein & Stein 1991) and (Heston 1993) first Scott, then Stein and Stein and finally Heston considered the model

$$dS_t = \alpha S_t dt + \sigma_t S_t dW_{1,t} \tag{5.1.5a}$$
$$d\sigma_t = \kappa(\beta - \sigma_t)dt + \theta dW_{2,t}. \tag{5.1.5b}$$

The model incorporates a mean-reverting tendency into the behaviour of σ_t. β is the stationary value of σ_t and κ is denoted the *speed of adjustment* parameter.

5.1.2.3 The Black, Scholes and Courtadon model

$$dS_t = \alpha S_t dt + \sigma_t S_t dW_{1,t} \tag{5.1.6a}$$
$$d\sigma_t = \kappa(\beta - \sigma_t)dt + \theta \sigma_t dW_{2,t} \tag{5.1.6b}$$

a slight extension of the Scott, Stein & Stein and Heston model.

5.1.2.4 Correlation

The two Wiener processes of the above listed models are correlated with correlation ρ. It should be noted that the models may be defined using only uncorrelated Wiener processes. Let $\tilde{W}_{1,t}$ and $\tilde{W}_{2,t}$ be two standard uncorrelated Wiener processes. If the Wiener increments above are defined by $dW_{1,t} = d\tilde{W}_{1,t}$ and $dW_{2,t} = \rho d\tilde{W}_{1,t} + \sqrt{1-\rho^2}d\tilde{W}_{2,t}$ then the above models may be written using only $d\tilde{W}_{1,t}$ and $d\tilde{W}_{2,t}$. For instance, the Black, Scholes and Courtadon model has the following structure when only uncorrelated Wiener processes are used:

$$dS_t = \alpha S_t dt + \sigma_t S_t d\tilde{W}_{1,t} \tag{5.1.7a}$$
$$d\sigma_t = \kappa(\beta - \sigma_t)dt + \rho\theta\sigma_t d\tilde{W}_{1,t} + \sqrt{1-\rho^2}\theta\sigma_t d\tilde{W}_{2,t} \tag{5.1.7b}$$

5.1.3 Discretization

When working with stochastic differential equations and continuous time stochastic volatility models it is often convenient to use a discrete time approximation. For instance, as there is no way to simulate a process in continuous time, a discrete time approximation is needed for this purpose.

A number of different ways to discretize the solution to a stochastic differential equation exist. Here, only the so-called Euler discretization scheme shall be considered, as it is the most simple and easy-to-use discretization scheme. Other, more precise discretization schemes exist, but the Euler scheme suffices for the purposes of this thesis. For an in-depth description of discretization schemes, see (Kloeden & Platen 1995) or (Madsen et al. 1998).

5.1.3.1   The Euler scheme

When presenting the Euler scheme, we consider the solution X_t to the continuous time stochastic volatility model

$$X_t = X_0 + \sum_{j=1}^{d} \int_0^t a^j(s, X_s)ds + \sum_{j=1}^{m} b^j(s, X_s)dW_s^j \tag{5.1.8}$$

where m is dimension of the Wiener process (that is, the number of independent standard Wiener processes) and d is the dimension of the drift. The discretizised approximation to X is denoted Y.

In the one-dimensional case the Euler scheme takes the form

$$Y_{n+1} = Y_n + a(n, Y_n)\Delta\tau_n + b(n, Y_n)\Delta W_n \tag{5.1.9}$$

where $\Delta\tau_n = \tau_{n+1} \Leftrightarrow \tau_n$ is the length of the time discretization, and $\Delta W_n = W_{\tau_{n+1}} \Leftrightarrow W_{\tau_n}$ is the $N(0, \Delta\tau_n)$ increments of the Wiener process W on the time interval $[\tau_{n+1}, \tau_n]$.

In the multi-dimensional case the k'th component of the discretization vector is

$$Y_{n+1}^k = Y_n^k + a^k(n, Y_n)\Delta\tau_n + \sum_{j=1}^{m} b^{k,j}\Delta W_{\tau_n}^j. \tag{5.1.10}$$

As an example, consider again the Black, Scholes and Courtadon model (5.1.7). The discrete time approximation using the Euler scheme at equidistant discretization times $t = 1, 2, 3, \ldots$ is

$$
\begin{aligned}
S_{t+1} &= S_t + \alpha S_t + \sigma_t S_t \Delta W_{1,t} & \text{(5.1.11a)} \\
\sigma_{t+1} &= \sigma_t + \kappa(\beta \Leftrightarrow \sigma_t) + \rho\theta\sigma_t\Delta W_{1,t} + \sqrt{1 \Leftrightarrow \rho^2}\sigma_t\Delta W_{2,t} & \text{(5.1.11b)}
\end{aligned}
$$

where

$$\begin{bmatrix} W_{1,t} \\ W_{2,t} \end{bmatrix} \in N\left(\begin{bmatrix} 0 \\ 0 \end{bmatrix}, \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix}\right). \qquad (5.1.12)$$

5.1.3.2 Convergence

In the context of discretization it might be useful to introduce the concept of convergence of an approximation to a stochastic differential equation. Traditionally, one distinguishes between strong and weak convergence, c.f. (Kloeden & Platen 1995, chapter 3).

DEFINITION 5.1 (STRONG CONVERGENCE).
A general discrete time approximation $Y^{\Delta t}(t)$ with maximum step size Δt converges strongly to X at time T if

$$\lim_{\Delta t \to 0} \mathrm{E}\left[\left|X(T) \Leftrightarrow Y^{\Delta t}(T)\right|\right] = 0, \qquad (5.1.13)$$

and if there exist positive constants C and α independent of Δt and a finite $\Delta_0 > 0$ such that

$$\mathrm{E}\left[\left|X(T) \Leftrightarrow Y^{\Delta t}(T)\right|\right] \leq C(\Delta t)^{\alpha} \qquad (5.1.14)$$

for each $0 \leq \Delta t \leq \Delta_0$ then $Y^{\Delta t}$ is said to converge strongly with order α.   ▲

DEFINITION 5.2 (WEAK CONVERGENCE).
 A general discrete time approximation $Y^{\Delta t}(t)$ with maximum step size Δt converges weakly to X at time T with respect to a class $\mathcal{C}$ of test functions g if

$$\lim_{\Delta t \to 0}\left|\mathrm{E}\left\{g(X(T))\right\} \Leftrightarrow \mathrm{E}\left\{g(Y^{\Delta t}(T))\right\}\right| = 0 \qquad (5.1.15)$$

for all $g \in \mathcal{C}$ and if there for each *polynomial* G exist positive constants C and β independent of Δt and a finite $\Delta_0 > 0$ such that

$$\left|\mathrm{E}\left\{G(X(T))\right\} \Leftrightarrow \mathrm{E}\left\{G(Y(T))\right\}\right| \leq C(\Delta t)^{\beta} \qquad (5.1.16)$$

for each $0 \leq \Delta t \leq \Delta_0$.   ▲

In connection with the two convergence definitions it should be noted that strong convergence is also denoted pathwise convergence as its definition regards the distance between the path of X and the path of Y. It should also be noted that if the class $\mathcal{C}$ of Definition 5.2 contains all polynomials, weak convergence implies convergence of all moments.

EXAMPLE 5.1. The Euler approximation (5.1.10) (usually) converges weakly with order $\beta = 1$ and strongly with order $\alpha = 1/2$. ♦

5.2   From discrete to continuous time

As previously suggested, the purpose of relating discrete time models to continuous time models is to obtain the parameters of the continuous models using the parameter estimates from the discrete time models.

One apparent plausible way to bridge the gap between discrete and continuous time stochastic volatility models is to look at the continuous time models as discrete time models sampled using very small sampling intervals. In this light, one could seek to bridge the gap by allowing the sampling interval of the discrete time models to get arbitrarily small and then choose the continuous time model that arises in the limit. Strictly speaking, one considers a sequence of models M_s, where s is the size of the sampling interval and investigates the model M in the limit, $M = \lim_{s \to 0} M_s$. The process of changing the size of the sampling interval is called temporal aggregation.

Unfortunately, as showed by (Drost & Nijman 1993), Arch models are not closed under temporal aggregation. This means that if observations sampled, say, daily follow an Arch process, then observations sampled, say, every other day will in general not follow an Arch process. In fact, the innovations z_t from (3.1.1) of the every-other day model will not be independently distributed, hence violating the Arch definition[1].

That Arch models fail to be closed under temporal aggregation raises a serious question when one wants to change the size of the sample interval. Shall one insist that the model at hand must be closed under temporal aggregation, or shall one accept that the model is only valid for a particular choice of the sample interval size? The first approach is taken in (Drost & Nijman 1993). The authors set out to find a "core", a collection of properties that GARCH processes possess and which

---

[1]Strictly speaking, Drost and Nijman investigate GARCH models, not Arch models. But, since they show that the z_t's fail to be i.i.d., their conclusion must apply to the whole class of Arch models.

is preserved under temporal aggregation. They succeed in doing so and they denote the preserved core *weak GARCH*. Unfortunately, weak GARCH does not specify the conditional distribution of z_t, thus leaving an incompletely specified sequence $\{M_s\}$ of models.

The alternative approach is pursued by (Duan 1997). Duan accepts that Arch models are not closed under temporal aggregation, but instead of removing those part of the Arch specification that prohibits the aggregation, he accepts the fact that at most one element of the sequence of models $\{M_s\}$ is correct, and decides that the other elements are merely approximations. Thus, if one take the limiting model M to be the correct one (as one does if one uses this model to price options), then all the other models are merely approximations.

To see how Duan derived the diffusion limit of the augmented GARCH process, consider the augmented GARCH(1,1) process, now written in a way slightly different from (3.6.1) and (3.6.2). The definition is taylored to the application to price processes, so the geometric returns of Equation 2.2.2 are considered.

DEFINITION 5.3 (THE AUGMENTED GARCH(1,1) PROCESS).
 Let $\{z_t\}$ be a sequence of indepent identical distributed stochastic variables with zero mean and unit variance. The returns R_t of a stock price process is said to follow an augmented GARCH(1,1) process if

$$R_t = \omega(\tilde{h}_t) + z_t\sqrt{\tilde{h}_t} \tag{5.2.1a}$$

$$\phi_t = \gamma_0 + \left(\gamma_1 + \gamma_2 Z_{t-1}^{(2)} + \gamma_3 Z_{t-1}^{(3)}\right)\phi_{t-1} + Z_{t-1}^{(4)} \tag{5.2.1b}$$

$$\tilde{h}_t = \begin{cases} |\xi\phi_t \Leftrightarrow \xi + 1|^{1/\xi} & \text{if} \quad \xi > 0, \\ \exp(\phi_t \Leftrightarrow 1) & \text{if} \quad \xi = 0 \end{cases} \tag{5.2.1c}$$

where

$$Z_k^{(2)} = |z_k \Leftrightarrow c|^\delta \tag{5.2.2a}$$

$$Z_k^{(3)} = \max(0, c \Leftrightarrow z_k)^\delta \tag{5.2.2b}$$

$$Z_k^{(4)} = \gamma_4 f(|z_k \Leftrightarrow c|, \delta) + \gamma_5 f(\max(0, c \Leftrightarrow z_k), \delta), \tag{5.2.2c}$$

$f(x, \delta) = (x^\delta \Leftrightarrow 1)/\delta$ where ω some well-behaved function of $\tilde{h}_t$.   ▲

Note that the function ω of Definition 5.3 should not be confused with the parameters ω_i of the augmented GARCH process.

Compared to the definition of the general augmented GARCH process (Definition 3.5), the functional form of R_t have been restricted. The reason for this and the reason for introducing the Z's will become clear later. Negative values of ξ have been disallowed in order to rule out the possibility of explosion.

Note that if it is recalled that $R_t = \ln S_t \Leftrightarrow \ln S_{t-1}$, Equation 5.2.1a in Definition 5.3 may also be written

$$\ln S_t = \omega(\tilde{h}_t) + \ln S_{t-1} + z_t \sqrt{\tilde{h}_t} \tag{5.2.3}$$

5.2.1   Approximating augmented GARCH

From its definition it is clear that the augmented GARCH(1,1) process does not facilitate an investigation of its behaviour when the size of the sample interval is changed. To allow for different sample interval sizes, (Duan 1997) introduced *the approximating augmented GARCH process*. He considered the finite time interval $[0, T]$ and divided it into nT subintervals each of length $s = 1/n$. By introducing

$$q_2 \;=\; \mathrm{E}\left(Z_k^{(2)}\right) \tag{5.2.4a}$$

$$q_3 \;=\; \mathrm{E}\left(Z_k^{(3)}\right) \tag{5.2.4b}$$

$$q_4 \;=\; \mathrm{E}\left(Z_k^{(4)}\right) \tag{5.2.4c}$$

and properly restricting the distribution of z_k so that these moments exist (assuming z_k normal will do) he defined it in the following way

DEFINITION 5.4.
THE APPROXIMATING AUGMENTED GARCH(1,1) PROCESS.

$$\ln S_{ks}^{(n)} = \omega(\tilde{h}_{ks}^{(n)})s + \ln S_{(k-1)s}^{(n)} + \sqrt{\tilde{h}_{ks}^{(n)}}\, z_k \sqrt{s} \tag{5.2.5a}$$

$$\begin{aligned}
\phi_{(k+1)s}^{(n)} &= (\gamma_0 + q_4)s + \phi_{ks}^{(n)}\left[1 + (\gamma_1 + \gamma_2 q_2 + \gamma_3 q_3 - 1)s\right] \\
&\quad + \phi_{ks}^{(n)}\left[\gamma_2(Z_k^{(2)} - q_2) + \gamma_3(Z_k^{(3)} - q_3)\right]\sqrt{s} \\
&\quad + (Z_k^{(4)} - q_4)\sqrt{s},
\end{aligned} \tag{5.2.5b}$$

$$\tilde{h}_{ks}^{(n)} = \begin{cases} |\xi\phi_{ks}^{(n)} - \xi + 1|^{1/\xi} & \text{if } \xi > 0, \\ \exp(\phi_{ks}^{(n)} - 1) & \text{if } \xi = 0 \end{cases} \tag{5.2.5c}$$

▲

The approximating augmented GARCH process may seem a bit odd, not at least from a first look. Unfortunately, Duan does not reveal much about the relation between the approximating augmented GARCH process and the augmented GARCH process, so some clarifying remarks are provided below.

THEOREM 5.1.
The approximating augmented GARCH process with $s = 1$ is an augmented GARCH process.

Proof. $\tilde{h}$ and $\ln S$ of Definition 5.4 is readily seen to be identical to their counterparts in definition 5.3 when $s = 1$.

Consider ϕ of Definition 5.4. Setting $s = 1$ and $k = t$ yields

$$\begin{aligned}
\phi_{t+1} &= \gamma_0 + q_4 + \phi_t(1 + \gamma_1 + \gamma_2 q_2 + \gamma_3 q_3 - 1) \\
&\quad + \phi_t\left[\gamma_2\left(Z_t^{(2)} - q_2\right) + \gamma_3\left(Z_t^{(3)} - q_3\right)\right] + Z_t^{(4)} - q_4 \\
&= \gamma_0 + \phi_t(\gamma_1 + \gamma_2 Z_t^{(2)} + \gamma_3 Z_t^{(3)}) + Z_t^{(4)} \tag{5.2.6}
\end{aligned}$$

which is the same updating equation as in Definition 5.3.

The definition of the approximating augmented GARCH model for values of s, $0 < s < 1$ is perhaps best justified by the following remarks.

The goal is to find the limiting model M, $M = \lim_{s \to 0} M_s$, a continuous time stochastic volatility model. It is assumed that this model may be written as

$$d\phi_t = \mu\phi_t dt + \nu_t\phi_t dW_t \tag{5.2.7}$$

where μ is a constant specifying the drift and ν_t specifies the diffusion of ϕ, ν_t being a deterministic function of ϕ_t. Using the Euler discretization scheme (5.1.9) and a discretization step of Δt, (5.2.7) may be discretized like

$$\phi_{t+\Delta t} = \phi_t + \mu\phi_t \Delta t + \nu_t\phi_t \Delta W_t \tag{5.2.8}$$

where $\Delta W_t \in N(0, \Delta t)$. Preferring the standard normal distribution, ΔW_t may be written $\Delta W_t = \sqrt{\Delta t}\Delta N_t$ where $\Delta N_t \in N(0, 1)$. The μ and the ν of (5.2.7) are unknown, but if Equation 5.2.1b of the augmented GARCH(1,1) process (Definition 5.3) is taken to be the discrete time approximation to (5.2.7) when the size of the discretization step is 1, then one may proceed as follows.

Inserting $\Delta t = 1$ into (5.2.8) and substracting ϕ_t from both sides yields

$$\phi_{t+1} \Leftrightarrow \phi_t = \mu\phi_t + \nu_t\phi_t \Delta N_t \tag{5.2.9}$$

Taking the expected value conditional on ϕ_t on both sides, the following expression for $E[\phi_{t+1}|\phi_t]$ is obtained:

$$E[\phi_{t+1} \Leftrightarrow \phi_t|\phi_t] = E[\mu\phi_t|\phi_t] + E[\nu_t\phi_t \Delta N_t|\phi_t] \tag{5.2.10}$$

and as the last term is zero,

$$E[\phi_{t+1} \Leftrightarrow \phi_t|\phi_t] = E[\mu\phi_t|\phi_t] = \mu\phi_t \tag{5.2.11}$$

that is, the drift may be expressed as the expected one-step increment. Inserting this into (5.2.9) and isolating the diffusion term yields

$$\nu_t\phi_t \Delta N_t = \phi_{t+1} \Leftrightarrow \phi_t \Leftrightarrow E[\phi_{t+1} \Leftrightarrow \phi_t|\phi_t]. \tag{5.2.12}$$

The two expressions (5.2.11) and (5.2.12) for the drift and diffusion of ϕ, respectively, are what constitutes the ϕ updating equation of the approximating augmented GARCH process. If the two expressions are inserted into (5.2.8), the following updating equation for ϕ is obtained

$$
\begin{aligned}
\phi_{t+\Delta t} \;=\; & \phi_t + \mathrm{E}[\phi_{t+1} - \phi_t | \phi_t]\Delta t \\
& + \left(\phi_{t+1} - \phi_t - \mathrm{E}[\phi_{t+1} - \phi_t | \phi_t]\right)\sqrt{\Delta t}\,\Delta N_t.
\end{aligned} \tag{5.2.13}
$$

Using Equation 5.2.1b of the definition of the augmented GARCH(1,1) process, it is seen that

$$
\begin{aligned}
\mathrm{E}[\phi_{t+1} - \phi_t | \phi_t] & \\
= \; & \mathrm{E}[\phi_{t+1}|\phi_t] - \phi_t \\
= \; & \mathrm{E}\left[\gamma_0 + \phi_t\left(\gamma_1 + \gamma_2 Z_t^{(2)} + \gamma_3 Z_t^{(3)}\right) + Z_t^{(4)}\Big|\phi_t\right] - \phi_t \\
= \; & \gamma_0 + \gamma_1\phi_t + \gamma_2 q_2\phi_t + \gamma_3 q_3\phi_t + q_4 - \phi_t \\
= \; & \gamma_0 + q_4 + \left(\gamma_1 + \gamma_2 q_2 + \gamma_3 q_3 - 1\right)\phi_t
\end{aligned} \tag{5.2.14}
$$

and that

$$
\begin{aligned}
\phi_{t+1} - \phi_t - & \mathrm{E}[\phi_{t+1} - \phi_t | \phi_t] \\
= \; & \gamma_0 + \phi_t\left(\gamma_1 + \gamma_2 Z_t^{(2)} + \gamma_3 Z_t^{(3)}\right) + Z_t^{(4)} - \phi_t \\
& - \left(\gamma_0 + q_4 + \left(\gamma_1 + \gamma_2 q_2 + \gamma_3 q_3 - 1\right)\phi_t\right) \\
= \; & \left[\gamma_2\left(Z_t^{(2)} - q_2\right) + \gamma_3\left(Z_t^{(3)} - q_3\right)\right]\phi_t + \left(Z_t^{(4)} - q_4\right). \tag{5.2.15}
\end{aligned}
$$

Together with (5.2.13) this means that

$$
\begin{aligned}
\phi_{t+\Delta t} \;=\; & \phi_t + (\gamma_0 + q_4)\Delta t + (\gamma_1 + \gamma_2 q_2 + \gamma_3 q_3 - 1)\phi_t\Delta t \\
& + \left[\gamma_2\left(Z_t^{(2)} - q_2\right) + \gamma_3\left(Z_t^{(3)} - q_3\right)\right]\phi_t\sqrt{\Delta t} \\
& + \left(Z_t^{(4)} - q_4\right)\sqrt{\Delta t}
\end{aligned} \tag{5.2.16}
$$

or, with the notation of Definition 5.4:

$$
\begin{aligned}
\phi_{(k+1)s}^{(n)} \;=\; & (\gamma_0 + q_4)s + \phi_{ks}^{(n)}\left[1 + (\gamma_1 + \gamma_2 q_2 + \gamma_3 q_3 - 1)s\right] \\
& + \phi_{ks}^{(n)}\left[\gamma_2(Z_k^{(2)} - q_2) + \gamma_3(Z_k^{(3)} - q_3)\right]\sqrt{s} \\
& + (Z_k^{(4)} - q_4)\sqrt{s}
\end{aligned} \tag{5.2.17}
$$

which is exactly the updating equation of ϕ in the approximating augmented GARCH process.

Equation 5.2.3 of Definition 5.4 is obtained along the same lines. Assuming that (5.2.3) is the one-step Euler approximation to the solution of a stochastic differential equation, it is readily seen that $\ln S$ of Definition 5.4 is the discretization with step size s.

5.2.2   Diffusion limit of approximating augmented GARCH

Using the approximating augmented GARCH process, an investigation of the limit of the augmented GARCH process is possible.

Define the covariance matrix

$$
V_Z = V\left(z_k, Z_k^{(2)}, Z_k^{(3)}, Z_k^{(4)}\right) = \begin{bmatrix} 1 & \sigma_{12} & \sigma_{13} & \sigma_{14} \\ \sigma_{21} & \sigma_2^2 & \sigma_{23} & \sigma_{24} \\ \sigma_{31} & \sigma_{32} & \sigma_3^2 & \sigma_{34} \\ \sigma_{41} & \sigma_{42} & \sigma_{43} & \sigma_4^2 \end{bmatrix} \tag{5.2.18}
$$

where all the elements by definition are constants, i.e. independent of k.

THEOREM 5.2. The approximating augmented GARCH(1,1) process converges weakly to the unique strong solution of the following diffusion model

$$
d\ln S_t = \omega(\tilde{h}_t)dt + \sqrt{\tilde{h}_t}d\tilde{W}_{1,t} \tag{5.2.19a}
$$

$$
\begin{aligned}
d\phi_t &= \left[\gamma_0 + q_4 + (\gamma_1 + \gamma_2 q_2 + \gamma_3 q_3 \Leftrightarrow 1)\phi_t\right]dt \\
&\quad + v_t\rho_t d\tilde{W}_{1,t} + v_t\sqrt{1 \Leftrightarrow \rho_t^2}d\tilde{W}_{2,t}
\end{aligned} \tag{5.2.19b}
$$

$$
\tilde{h}_t = \begin{cases} |\xi\phi_t \Leftrightarrow \xi + 1|^{1/\xi} & \text{if} \quad \xi > 0, \\ \exp(\phi_t \Leftrightarrow 1) & \text{if} \quad \xi = 0 \end{cases} \tag{5.2.19c}
$$

where

$$
\begin{aligned}
v_t &= \sqrt{\sigma_4^2 + 2(\gamma_2\sigma_{24} + \gamma_3\sigma_{34})\phi_t + (\gamma_2^2\sigma_2^2 + \gamma_3^2\sigma_3^2 + 2\gamma_2\gamma_3\sigma_{23})\phi_t^2} \\
\rho_t &= v_t^{-1}\left[\sigma_{14} + \phi_t(\gamma_2\sigma_{12} + \gamma_3\sigma_{13})\right]
\end{aligned}
$$

and $\tilde{W}_{1,t}$ and $\tilde{W}_{2,t}$ are standard uncorrelated Wiener processes.

Proof. See (Duan 1997).

The proof of Theorem 5.2 in (Duan 1997) is very rigorous and therefore difficult to comprehend. As such, it has been left out. However, an understanding of the theorem may be achieved through a less rigorous, more heuristical analysis.

It is not difficult to see that the updating equation (5.2.19a) of $\ln S$ is the diffusion limit of (5.2.5a). Nor is it difficult to comprehend Equation (5.2.19c) of h as it is identical to (5.2.5c). Remaining is the expression of $d\phi_t$.

The drift of ϕ_t is most easily recognized by relating the equations (5.2.11) and (5.2.14) to one another hence obtaining

$$\mu\phi_t = \gamma_0 + q_4 + (\gamma_1 + \gamma_2 q_2 + \gamma_3 q_3 \Leftrightarrow 1)\,\phi_t, \qquad (5.2.20)$$

the drift of (5.2.19b).

The difficult part is the diffusion of ϕ_t. Reconsider Equation 5.2.12:

$$\nu_t\phi_t\Delta N_t = \phi_{t+1} \Leftrightarrow \phi_t \Leftrightarrow E[\phi_{t+1} \Leftrightarrow \phi_t|\phi_t] \qquad (5.2.21)$$

and apply the conditional variance operator on both sides

$$V\left(\nu_t\phi_t\Delta N_t|\phi_t\right) = V\left(\phi_{t+1} \Leftrightarrow \phi_t \Leftrightarrow E[\phi_{t+1} \Leftrightarrow \phi_t|\phi_t]|\phi_t\right). \qquad (5.2.22)$$

Conditioning on ϕ_t, only ΔN_t of the left hand side and ϕ_{t+1} of the right hand side are stochastic. That is,

$$\nu_t^2\phi_t^2 V\left(\Delta N_t|\,\phi_t\right) = V(\phi_{t+1}|\phi_t). \qquad (5.2.23)$$

and, as $\Delta N_t \in N\left(0,1\right)$,

$$\sqrt{V\left(\phi_{t+1}|\phi_t\right)} = \nu_t\phi_t. \qquad (5.2.24)$$

Inserting (5.2.20) and (5.2.24) into (5.2.7) yields

$$d\phi_t = \gamma_0 + q_4 + (\gamma_1 + \gamma_2 q_2 + \gamma_3 q_3 \Leftrightarrow 1)\,\phi_t + \sqrt{V\left(\phi_{t+1}|\phi_t\right)}dW_t. \quad (5.2.25)$$

Now focus on the Wiener process dW_t driving ϕ_t. Refering to Definition 5.3, ϕ_t is not driven by its "own" innovations, but by z_t through $Z_t^{(2)}$, $Z_t^{(3)}$ and $Z_t^{(4)}$. This means that $\phi_{t+1}|\phi_t$ must be correlated with z_t. Recalling Section 5.1.2.4, this correlation may be incorporated into (5.2.25) like

$$
\begin{aligned}
d\phi_t &= \gamma_0 + q_4 + (\gamma_1 + \gamma_2 q_2 + \gamma_3 q_3 \Leftrightarrow 1)\,\phi_t \\
&\quad + \sqrt{V(\phi_{t+1}|\phi_t)}\,\mathrm{Cor}(z_t, \phi_{t+1}|\phi_t)d\tilde{W}_{1,t} \\
&\quad + \sqrt{V(\phi_{t+1}|\phi_t)}\sqrt{1 \Leftrightarrow \mathrm{Cor}^2(z_t, \phi_{t+1}|\phi_t)}d\tilde{W}_{2,t}
\end{aligned} \tag{5.2.26}
$$

where $\tilde{W}_{1,t}$ and $\tilde{W}_{2,t}$ are two standard uncorrelated Wiener processes, and $\tilde{W}_{1,t}$ is the Wiener process driving $\ln S$, c.f. Theorem 5.2.

Explicit expressions for $V(\phi_{t+1}|\phi_t)$ and $\mathrm{Cor}(z_t, \phi_{t+1}|\phi_t)$ may be derived using the augmented GARCH(1,1) process of Definition 5.3:

$$
\begin{aligned}
V(\phi_{t+1}|\phi_t) &= V\left[\gamma_0 + \left(\gamma_1 + \gamma_2 Z_t^{(2)} + \gamma_3 Z_t^{(3)}\right)\phi_t + Z_t^{(4)}\Big|\phi_t\right] \\
&= V\left[\left(\gamma_1 + \gamma_2 Z_t^{(2)} + \gamma_3 Z_t^{(3)}\right)\phi_t\Big|\phi_t\right] \\
&\quad + V\left[Z_t^{(4)}\Big|\phi_t\right] \\
&\quad + 2\,\mathrm{Cov}\left[\left(\gamma_1 + \gamma_2 Z_t^{(2)} + \gamma_3 Z_t^{(3)}\right)\phi_t, Z_t^{(4)}\Big|\phi_t\right].
\end{aligned} \tag{5.2.27}
$$

With (5.2.18) in mind, explicit expressions of the three parts of this summation may be obtained according to

$$
\begin{aligned}
V\left[\left(\gamma_1 + \gamma_2 Z_t^{(2)} + \gamma_3 Z_t^{(3)}\right)\phi_t\Big|\phi_t\right] & \\
&= \phi_t^2 V\left[\gamma_1 + \gamma_2 Z_t^{(2)} + \gamma_3 Z_t^{(3)}\Big|\phi_t\right] \\
&= \phi_t^2 V\left[\gamma_2 Z_t^{(2)} + \gamma_3 Z_t^{(3)}\Big|\phi_t\right] \\
&= \phi_t^2 \left(\gamma_2^2 \sigma_2^2 + \gamma_3^2 \sigma_3^2 + 2\gamma_2\gamma_3\sigma_{23}\right)
\end{aligned} \tag{5.2.28}
$$

and

$$
V\left[Z_t^{(4)}\Big|\phi_t\right] = \sigma_4^2 \tag{5.2.29}
$$

and

$$
\begin{aligned}
\text{Cov}\left[\left(\gamma_1 + \gamma_2 Z_t^{(2)} + \gamma_3 Z_t^{(3)}\right)\phi_t, Z_t^{(4)}\Big|\phi_t\right] & \\
= \ & \phi_t \,\text{Cov}\left[\left(\gamma_1 + \gamma_2 Z_t^{(2)} + \gamma_3 Z_t^{(3)}\right), Z_t^{(4)}\Big|\phi_t\right] \\
= \ & \phi_t\left(\gamma_2 \sigma_{24} + \gamma_3 \sigma_{34}\right).
\end{aligned}
\tag{5.2.30}
$$

Inserting these expressions into (5.2.27) yields

$$
\begin{aligned}
V(\phi_{t+1}|\phi_t) \ = \ & \phi_t^2\left(\gamma_2^2\sigma_2^2 + \gamma_3^2\sigma_3^2 + 2\gamma_2\gamma_3\sigma_{23}\right) \\
& + \sigma_4^2 + 2\phi_t\left(\gamma_2\sigma_{24} + \gamma_3\sigma_{34}\right)
\end{aligned}
\tag{5.2.31}
$$

so that the diffusion $\sqrt{V(\phi_{t+1}|\phi_t)}$ is actually v_t of Theorem 5.2.

Applying the same technique to the correlation $\text{Cor}(z_t, \phi_{t+1}|\phi_t)$ yields

$$
\text{Cor}(z_t, \phi_{t+1}|\phi_t) = \frac{\text{Cov}(z_t, \phi_{t+1}|\phi_t)}{\sqrt{V(z_t|\phi_t)}\sqrt{V(\phi_{t+1}|\phi_t)}}
\tag{5.2.32}
$$

where $\sqrt{V(z_t|\phi_t)} = 1$, $\sqrt{V(\phi_{t+1}|\phi_t)} = v_t$ and

$$
\begin{aligned}
\text{Cov}&(z_t, \phi_{t+1}|\phi_t) \\
= \ & \text{Cov}\left[z_t, \gamma_0 + \left(\gamma_1 + \gamma_2 Z_t^{(2)} + \gamma_3 Z_t^{(3)}\right)\phi_t + Z_t^{(4)}\Big|\phi_t\right] \\
= \ & \text{Cov}\left[z_t, \left(\gamma_1 + \gamma_2 Z_t^{(2)} + \gamma_3 Z_t^{(3)}\right)\phi_t\Big|\phi_t\right] + \text{Cov}\left[z_t, Z_t^{(4)}\Big|\phi_t\right] \\
= \ & \phi_t\left[\gamma_2 \,\text{Cov}\left(z_t, Z_t^{(2)}\Big|\phi_t\right) + \gamma_3 \,\text{Cov}\left(z_t, Z_t^{(3)}|\phi_t\right)\right] \\
& + \text{Cov}\left(z_t, Z_t^{(4)}|\phi_t\right) \\
= \ & \phi_t\left(\gamma_2\sigma_{12} + \gamma_3\sigma_{13}\right) + \sigma_{14}.
\end{aligned}
\tag{5.2.33}
$$

Hence

$$
\text{Cor}(z_t, \phi_{t+1}|\phi_t) = \frac{\phi_t\left(\gamma_2\sigma_{12} + \gamma_3\sigma_{13}\right) + \sigma_{14}}{v_t}
\tag{5.2.34}
$$

which is identical to ρ_t of Theorem 5.2.

When the achieved expressions for $V(\phi_{t+1}|\phi_t)$ and $\mathrm{Cor}(z_t.\phi_{t+1}|\phi_t)$ are inserted into (5.2.26), $d\phi_t$ emerge as

$$
\begin{aligned}
d\phi_t &= \gamma_0 + q_4 + (\gamma_1 + \gamma_2 q_2 + \gamma_3 q_3 \Leftrightarrow 1)\,\phi_t \\
&\quad + v_t \rho_t d\tilde{W}_{1,t} + v_t \sqrt{1 \Leftrightarrow \rho_t^2}\, d\tilde{W}_{2,t}
\end{aligned}
\tag{5.2.35}
$$

which is (5.2.19b) of Theorem 5.2. The heuristical analysis of the diffusion limit of the augmented GARCH process is hereby concluded.

5.3 Parameter relationships

Now that the diffusion limit of the augmented GARCH process is in place, it is possible to derive algebraic parameter relationships between models fitting into the augmented GARCH(1,1) process and their corresponding diffusion limits. In the same way it is possible to find the discrete time models corresponding to the commonly used continuous time stochastic volatility models of Section 5.1.2.

This section is devoted to the derivation of the continuous time stochastic volatility models corresponding to the GARCH(1,1) and EGARCH(1,1) processes and to the deduction of the discrete time stochastic volatility models corresponding to the Hull-White model of Section 5.1.2.1, the Scott, Stein & Stein and Heston model of Section 5.1.2.2 and the Black, Scholes and Courtadon model of Section 5.1.2.3.

The derivation of the continuous time stochastic volatility model corresponding to the GARCH(1,1) process is included in the thesis as an example of how to derive the continuous time stochastic volatility model that corresponds to a given Arch model, and is thus carefully described. In the same way, a careful description of the deduction of the Arch model that corresponds to the continuous time stochastic volatility model of Black, Scholes and Courtadon is included to exemplify the deduction of the Arch model that corresponds to a given continuous time stochastic volatility model. To save space, the examples of EGARCH(1,1), Hull-White and Scott, Stein & Stein and Heston are treated with less rigour.

Before commencing these derivations and deductions, a minor difficulty has to be circumvented.

5.3.1   The Itô formula

As the reader may or may not have noticed, the models of section 5.1.2 deals with the evolution of the stock price S whereas the diffusion limit of the approximating augmented GARCH process deals with $\ln S$. To change the models of Section 5.1.2 so that they they focus on $\ln S$ instead of S, the Itô formula is needed.

THEOREM 5.3 (THE ITÔ FORMULA).
Let $X(t)$ be a solution to

$$dX_t = \alpha(X_t)dt + \sigma(X_t)dW_t, X_0 = x_0 \tag{5.3.1}$$

and let $\zeta : \Re \mapsto \Re$ be a two times differentiable function applied to X_t

$$Y_t = \zeta(X_t). \tag{5.3.2}$$

Then the following chain rule applies

$$dY_t = \left[\alpha(X_t)\frac{\partial \zeta}{\partial X_t} + \frac{1}{2}\sigma^2(X_t)\frac{\partial^2 \zeta}{\partial X_t^2}\right]dt + \sigma(X_t)\frac{\partial \zeta}{\partial X_t}dW_t \tag{5.3.3}$$

For a more general version of Itô's formula, see (Madsen et al. 1998). The version of Theorem 5.3 suffices for the purposes of this thesis.

Reconsidering (5.1.3) and applying $\zeta(S_t) = \ln S_t$ yields

$$\begin{aligned} d\ln S_t &= \left[\alpha S_t \frac{1}{S_t} + \frac{1}{2}\sigma_t^2 S_t^2 \frac{\Leftrightarrow 1}{S_t^2}\right]dt + \sigma_t S_t \frac{1}{S_t}dW_{1,t} \\ &= \left(\alpha \Leftrightarrow \frac{1}{2}\sigma_t^2\right)dt + \sigma_t dW_{1,t} \end{aligned} \tag{5.3.4}$$

whereas the expression of $dm(\sigma_t)$ of (5.1.3) remains unchanged.

Using (5.3.4), the diffusion models of Section 5.1.2 are easily transformed to the form of the diffusion limit of the approximating augmented GARCH process.

5.3.2   From GARCH(1,1) to diffusion

Recall from Section 3.6.1 that a GARCH(1,1) process is an augmented GARCH(1,1) process with the parameter values listed in Table 5.1

augmented GARCH	ξ	c	δ	γ_0	γ_1	γ_2	γ_3	γ_4	γ_5
Value	1	0	2	α_0	β_1	α_1	0	0	0

Table 5.1: Augmented GARCH(1,1) parameter values yielding a GARCH(1,1) process.

5.3.2.1 Augmented GARCH

Merging the setup of Table 5.1 into the augmented GARCH process of Definition 5.3, the model

$$
\begin{align}
\ln S_t &= \omega(\tilde{h}_t) + \ln S_{t-1} + z_t\sqrt{\tilde{h}_t} \tag{5.3.5a}\\
\phi_t &= \alpha_0 + \left(\beta_1 + \alpha_1 Z_{t-1}^{(2)}\right)\phi_{t-1} + Z_{t-1}^{(4)} \tag{5.3.5b}\\
\tilde{h}_t &= |\phi_t| \tag{5.3.5c}
\end{align}
$$

is obtained. It is noted that $Z_{t-1}^{(3)}$ is not part of the model. The rest of the Z's are (compare with (5.2.2))

$$
\begin{align}
Z_t^{(2)} &= |z_t|^2 = z_t^2 \tag{5.3.6a}\\
Z_t^{(4)} &= 0. \tag{5.3.6b}
\end{align}
$$

meaning that when z_t is standard normal distributed then, according to (5.2.4),

$$
\begin{align}
q_2 &= \mathrm{E}(z_t^2) = 1 \tag{5.3.7a}\\
q_4 &= \mathrm{E}\left(Z_t^{(4)}\right) = 0 \tag{5.3.7b}
\end{align}
$$

5.3.2.2 Approximating augmented GARCH

The approximating GARCH(1,1) process is

$$\ln S_{ks}^{(n)} \;=\; \omega(\tilde h_{ks}^{(n)})s + \ln S_{(k-1)s}^{(n)} + \sqrt{\tilde h_{ks}^{(n)}}\,z_k\sqrt{s} \tag{5.3.8a}$$

$$\phi_{(k+1)s}^{(n)} \;=\; \alpha_0 s + \phi_{ks}^{(n)}\left[1 + (\beta_1 + \alpha_1 \Leftrightarrow 1)s\right]$$
$$\qquad\qquad + \phi_{ks}^{(n)}\alpha_1(z_k^2 \Leftrightarrow 1)\sqrt{s} \tag{5.3.8b}$$

$$\tilde h_{ks}^{(n)} \;=\; |\phi_{ks}^{(n)}| \tag{5.3.8c}$$

and its diffusion limit is

$$d\ln S_t \;=\; \omega(\tilde h_t)dt + \sqrt{\tilde h_t}\,d\tilde W_{1,t} \tag{5.3.9a}$$

$$d\phi_t \;=\; [\alpha_0 + (\beta_1 + \alpha_1 \Leftrightarrow 1)\phi_t]dt$$
$$\qquad\qquad + v_t\rho_t d\tilde W_{1,t} + v_t\sqrt{1 \Leftrightarrow \rho_t^2}\,d\tilde W_{2,t} \tag{5.3.9b}$$

$$\tilde h_t \;=\; |\phi_t| \tag{5.3.9c}$$

where

$$v_t \;=\; \sqrt{\sigma_4^2 + 2\alpha_1\sigma_{24}\phi_t + \alpha_1^2\sigma_2^2\phi_t^2} \tag{5.3.10a}$$

$$\rho_t \;=\; v_t^{-1}[\sigma_{14} + \phi_t\alpha_1\sigma_{12}]. \tag{5.3.10b}$$

Before the diffusion limit is fully specified, explicit expressions for σ_{12}, σ_{14}, σ_2^2, σ_{24} and σ_4^2 have to be computed. As $Z_k^{(4)} = 0$ expressions involving $Z_k^{(4)}$ are all equal to zero. The computation of the remaining expressions goes like this

$$\begin{aligned}
\sigma_{12} \;&=\; \mathrm{Cov}(z_k, z_k^2) \\
&=\; \mathrm{E}\left[(z_k \Leftrightarrow \mathrm{E}(z_k))(z_k^2 \Leftrightarrow E(z_k^2))\right] \\
&=\; \mathrm{E}\left[z_k(z_k^2 \Leftrightarrow 1)\right] \\
&=\; \mathrm{E}(z_k^3 \Leftrightarrow z_k) = \mathrm{E}(z_k^3) \Leftrightarrow \mathrm{E}(z_k) = 0
\end{aligned} \tag{5.3.11a}$$

$$\sigma_2^2 \;=\; \mathrm{V}(z_k^2) = 2 \tag{5.3.11b}$$

where the fact z_k^2 is $\chi^2(1)$ distributed have been utilized. Inserting the σ-expressions into (5.3.10) yields

$$
\begin{align}
v_t &= \sqrt{2\alpha_1^2 \phi_t^2} = \sqrt{2}\alpha_1|\phi_t| \tag{5.3.12a}\\
\rho_t &= 0 \tag{5.3.12b}
\end{align}
$$

and inserting these expressions into (5.3.9b) and (5.3.9c) yields

$$
\begin{align}
d\phi_t &= [\alpha_0 + (\beta_1 + \alpha_1 \Leftrightarrow 1)\phi_t]dt + \sqrt{2}\alpha_1|\phi_t|d\tilde{W}_{2,t} \tag{5.3.13a}\\
\tilde{h}_t &= |\phi_t|, \tag{5.3.13b}
\end{align}
$$

but if the approximating GARCH(1,1) process (5.3.8) is considered along with the GARCH(1,1) parameter restrictions $0 \leq \alpha_1 + \beta_1 < 1$ and the fact that $s < 1$ it is observed that ϕ_t cannot take on negative values. Hence, the resulting diffusion model is as specified in Corollary 5.1

COROLLARY 5.1 (DIFFUSION LIMIT OF GARCH(1,1)).
The augmented GARCH(1,1) process that coincide with the GARCH(1,1) process when $s = 1$ converges weakly to the following diffusion model

$$
d\ln S_t = \omega(\tilde{h}_t)dt + \sqrt{\tilde{h}_t}d\tilde{W}_{1,t} \tag{5.3.14a}
$$

$$
d\tilde{h}_t = [\alpha_0 + (\beta_1 + \alpha_1 \Leftrightarrow 1)\tilde{h}_t]dt + \sqrt{2}\alpha_1\tilde{h}_t d\tilde{W}_{2,t} \tag{5.3.14b}
$$

The result of Corollary 5.1 has an interesting implication.

THEOREM 5.4 (HULL-WHITE).
The diffusion model of Corollary 5.1 is, apart from the parameter α_0, identical to a Hull - White model.

Proof. Recall from Section 5.1.2.1 and Theorem 5.3 that an uncorrelated Hull-White diffusion model may be cast as

$$
\begin{align}
d\ln S_t &= \left(\alpha \Leftrightarrow \frac{1}{2}\sigma_t^2\right)dt + \sigma_t d\tilde{W}_{1,t} \tag{5.3.15a}\\
d\sigma_t^2 &= a\sigma_t^2 dt + b\sigma_t^2 d\tilde{W}_{2,t}. \tag{5.3.15b}
\end{align}
$$

If in the model of Corollary 5.1, the following identities are incorporated

$$\omega\left(\tilde{h}_t\right) = \alpha \Leftrightarrow \frac{1}{2}\tilde{h}_t \tag{5.3.16a}$$

$$\tilde{h}_t = \sigma_t^2 \tag{5.3.16b}$$

$$\alpha_0 = 0 \tag{5.3.16c}$$

$$\beta_1 + \alpha_1 \Leftrightarrow 1 = a \tag{5.3.16d}$$

$$\sqrt{2}\alpha_1 = b \tag{5.3.16e}$$

then a Hull-White model is obtained.

REMARK 5.1. Note that a value of 0 of α_0 is not permitted by the GARCH(1,1) parameter constraints. In the proof of Theorem 5.4, the only purpose of the constraint $\alpha_0 = 0$ is to show that the only thing that distinguishes the diffusion limit of the GARCH(1,1) process from the Hull-White model is α_0. If one insists on the constraint $\alpha_0 = 0$, the NGARCH(1,1) process (Equation 3.5.8) has a correlated Hull-White model as its diffusion limit, c.f. (Duan 1997).   ▼

5.3.3   From EGARCH(1,1) to diffusion

An EGARCH(1,1) process is an augmented GARCH(1,1) process if the parameters of the augmented GARCH(1,1) process are constrained according to Table 5.2.

Par.	ξ	c	δ	γ_0	γ_1	γ_2	γ_3	γ_4	γ_5
Value	0	0	1	$1 + \alpha_0 \Leftrightarrow \beta_1 \Leftrightarrow \alpha_1 \lambda \mathrm{E}\vert z\vert$	β_1	0	0	$2\alpha_1\omega$	$\Leftrightarrow 2\alpha_1\omega$

Table 5.2: Augmented GARCH(1,1) parameter values yielding an EGARCH(1,1) process.

5.3.3.1   Approximating augmented EGARCH

The procedure when deriving the diffusion limit of the augmented GARCH(1,1) process coinciding with an EGARCH(1,1) process is similar to the related procedure of GARCH(1,1) described in the previous section. Thus, there is no need to present the elaborations here. Instead, the diffusion limit is presented, and the reader is referred to (Duan 1997) for a proof.

COROLLARY 5.2 (DIFFUSION LIMIT OF EGARCH(1,1)).
The augmented GARCH(1,1) process that coincide with the EGARCH(1,1) process when $s = 1$ converges weakly to the following diffusion model

$$d \ln S_t = \omega(\tilde{h}_t)dt + \sqrt{\tilde{h}_t}d\tilde{W}_{1,t} \tag{5.3.17a}$$

$$d \ln \tilde{h}_t = \left[\alpha_0 - \alpha_1 \lambda \mathrm{E}|z| + \sqrt{\frac{2}{\pi}}\alpha_1 \omega + (\beta_1 - 1)\ln h_t \right] dt$$

$$\quad - \alpha_1 \omega d\tilde{W}_{1,t} + |\alpha_1 \omega|\sqrt{\frac{\pi - 2}{\pi}}d\tilde{W}_{2,t} \tag{5.3.17b}$$

Proof. Using augmented GARCH parameters, the diffusion limit of $\ln \tilde{h}_t$ is

$$d \ln \tilde{h}_t$$
$$= \left[\gamma_0 + \sqrt{\frac{2}{\pi}}\left(\gamma_4 + \frac{\gamma_5}{2}\right) - \gamma_4 - \gamma_5 + \gamma_1 - 1 + (\gamma_1 - 1)\ln \tilde{h}_t \right] dt$$

$$\quad - \frac{\gamma_5}{2}d\tilde{W}_{1,t} + \left|\gamma_4 + \frac{\gamma_5}{2}\right|\sqrt{\frac{\pi - 2}{\pi}}d\tilde{W}_{2,t} \tag{5.3.18}$$

Incorporating the parameter constraints of Table 5.2 into Equation 5.3.18, the result of Corollary 5.2 is readily obtained. For a proof of (5.3.18), see (Duan 1996) and (Duan 1997).

5.3.4 From diffusion to augmented GARCH

The process of deducing the particular augmented GARCH process that has a given diffusion model as its weak limit is far more difficult than the opposite process. The main sources of difficulty are the facts that ξ, δ and c are unknown thus prohibiting one from knowing the relation between ϕ and $\tilde{h}$ and the exact expressions of the Z's. When deducing the augmented GARCH processes of the Scott, Stein & Stein and Heston model and the Black, Scholes and Courtadon model, one may proceed along the following lines, though.

When reconsidering the diffusion models of sections 5.1.2.2 and 5.1.2.3 it is observed that the diffusion model of σ is specified by an expression of $d\sigma$ and not,

for instance, by an expression of $d\sigma^2$ as was the case with the model of Hull and White. This means that in the diffusion limit of the augmented GARCH process (Theorem 5.2), $\sqrt{h}$ must be a linear function of ϕ. This leads one to suggest $1/2$ as the value of ξ, which again leads to

$$\tilde{h}_t = \left|\frac{1}{2}\phi_t - \frac{1}{2} + 1\right|^2 = \left|\frac{\phi_t + 1}{2}\right|^2 = \left(\frac{\phi_t + 1}{2}\right)^2. \tag{5.3.19}$$

The parameters c and δ regard the correlation between the Wiener processes driving $\ln S_t$ and ϕ, respectively. A simple guess is $c = 0$ and $\delta = 1$ which, when z_k is standard normal distributed, yields

$$Z_k^{(2)} = |z_k| \tag{5.3.20a}$$

$$Z_k^{(3)} = \max(0, -z_k) = \frac{1}{2}(|z_k| - z_k) \tag{5.3.20b}$$

$$\begin{aligned} Z_k^{(4)} &= \gamma_4(|z_k| - 1) + \gamma_5[\max(0, -z_k) - 1] \\ &= \gamma_4|z_k| - \gamma_4 + \gamma_5\left[\frac{1}{2}(|z_k| - z_k) - 1\right] \\ &= \left(\gamma_4 + \frac{1}{2}\gamma_5\right)|z_k| - \frac{1}{2}\gamma_5 z_k - (\gamma_4 + \gamma_5). \end{aligned} \tag{5.3.20c}$$

The expected values of the Z's are

$$\begin{aligned} q_2 &= E|z_k| = \int_{-\infty}^0 \frac{-u}{\sqrt{2\pi}}e^{-\frac{1}{2}u^2}du + \int_0^\infty \frac{u}{\sqrt{2\pi}}e^{-\frac{1}{2}u^2}du \\ &= 2\int_0^\infty \frac{u}{\sqrt{2\pi}}e^{-\frac{1}{2}u^2}du = \sqrt{\frac{2}{\pi}} \end{aligned} \tag{5.3.21a}$$

$$q_3 = E\left[\frac{1}{2}(|z_k| - z_k)\right] = \frac{1}{2}\sqrt{\frac{2}{\pi}} = \frac{1}{\sqrt{2\pi}} \tag{5.3.21b}$$

$$\begin{aligned} q_4 &= E\left[\left(\gamma_4 + \frac{1}{2}\gamma_5\right)|z_k| - \frac{1}{2}\gamma_5 z_k - (\gamma_4 + \gamma_5)\right] \\ &= \left(\gamma_4 + \frac{1}{2}\gamma_5\right)\sqrt{\frac{2}{\pi}} - (\gamma_4 + \gamma_5) \end{aligned} \tag{5.3.21c}$$

and the elements of the covariance matrix V_Z of (5.2.18) are

$$\sigma_{12} = \text{Cov}(z_k, |z_k|) = 0 \tag{5.3.22}$$

$$
\begin{aligned}
\sigma_{13} &= \text{Cov}\left[z_k, \frac{1}{2}(|z_k| - z_k)\right] \\
&= \frac{1}{2}\text{Cov}(z_k, |z_k|) - \frac{1}{2}V(z_k) = -\frac{1}{2}
\end{aligned}
\tag{5.3.23}
$$

$$
\begin{aligned}
\sigma_{14} &= \text{Cov}\left[z_k, \left(\gamma_4 + \frac{1}{2}\gamma_5\right)|z_k| - \frac{1}{2}\gamma_5 z_k - (\gamma_4 + \gamma_5)\right] \\
&= -\frac{1}{2}\gamma_5 V(z_k) = -\frac{1}{2}\gamma_5
\end{aligned}
\tag{5.3.24}
$$

$$
\begin{aligned}
\sigma_2^2 &= V|z_k| = E\left(|z_k|^2\right) - [E|z_k|]^2 \\
&= E\left(z_k^2\right) - q_2^2 \\
&= 1 - \frac{2}{\pi} = \frac{\pi - 2}{\pi}
\end{aligned}
\tag{5.3.25}
$$

$$
\begin{aligned}
\sigma_{23} &= \text{Cov}\left[|z_k|, \frac{1}{2}(|z_k| - z_k)\right] \\
&= \frac{1}{2}V|z_k| = \frac{1}{2}\frac{\pi - 2}{\pi} = \frac{\pi - 2}{2\pi}
\end{aligned}
\tag{5.3.26}
$$

$$
\begin{aligned}
\sigma_{24} &= \text{Cov}\left[|z_k|, \left(\gamma_4 + \frac{1}{2}\gamma_5\right)|z_k| - \frac{1}{2}\gamma_5 z_k - (\gamma_4 + \gamma_5)\right] \\
&= \left(\gamma_4 + \frac{1}{2}\gamma_5\right)V|z_k| = \left(\gamma_4 + \frac{1}{2}\gamma_5\right)\frac{\pi - 2}{\pi}
\end{aligned}
\tag{5.3.27}
$$

$$
\begin{aligned}
\sigma_3^2 &= V\left[\frac{1}{2}(|z_k| + z_k)\right] \\
&= \frac{1}{4}[V|z_k| - V(z_k)] = \frac{1}{4}\left(\frac{\pi - 2}{\pi} + 1\right)
\end{aligned}
\tag{5.3.28}
$$

$$
\begin{aligned}
\sigma_{34} &= \operatorname{Cov}\left[\frac{1}{2}|z_k| - \frac{1}{2}z_k,\right.\\
&\qquad \left.\left(\gamma_4 + \frac{1}{2}\gamma_5\right)|z_k| - \frac{1}{2}\gamma_5 z_k - (\gamma_4 + \gamma_5)\right]\\
&= \frac{1}{2}\left(\gamma_4 + \frac{1}{2}\gamma_5\right)\operatorname{V}|z_k| + \frac{1}{4}\gamma_5\operatorname{V}(z_k)\\
&= \left(\gamma_4 + \frac{1}{2}\gamma_5\right)\frac{\pi - 2}{2\pi} + \frac{1}{4}\gamma_5 \qquad (5.3.29)
\end{aligned}
$$

$$
\begin{aligned}
\sigma_4^2 &= \operatorname{V}\left[\left(\gamma_4 + \frac{1}{2}\gamma_5\right)|z_k| - \frac{1}{2}\gamma_5 z_k - (\gamma_4 + \gamma_5)\right]\\
&= \left(\gamma_4 + \frac{1}{2}\gamma_5\right)^2 \operatorname{V}|z_k| + \frac{1}{4}\gamma_5^2 \operatorname{V}(z_k)\\
&= \left(\gamma_4 + \frac{1}{2}\gamma_5\right)^2 \frac{\pi - 2}{\pi} + \frac{1}{4}\gamma_5^2 \qquad (5.3.30)
\end{aligned}
$$

The above findings leads to the following expression of $d\phi_t$:

$$
\begin{aligned}
d\phi_t &= \left[\gamma_0 + \left(\gamma_4 + \frac{\gamma_5}{2}\right)\sqrt{\frac{2}{\pi}} - (\gamma_4 + \gamma_5)\right.\\
&\qquad \left. + \left(\gamma_1 + \gamma_2\sqrt{\frac{2}{\pi}} + \frac{\gamma_3}{\sqrt{2\pi}} - 1\right)\phi_t\right]dt\\
&\quad + v_t\rho_t d\tilde{W}_{1,t} + v_t\sqrt{1 - \rho_t^2}\, d\tilde{W}_{2,t} \qquad (5.3.31)
\end{aligned}
$$

where

$$
\begin{aligned}
v_t &= \Bigg\{ \left(\gamma_4 + \frac{\gamma_5}{2}\right)^2 \frac{\pi-2}{\pi} + \frac{\gamma_5^2}{4} \\
&\quad + 2\left[\gamma_2\left(\gamma_4 + \frac{\gamma_5}{2}\right)\frac{\pi-2}{\pi}\right. \\
&\quad + \gamma_3\left(\left(\gamma_4 + \frac{\gamma_5}{2}\right)\frac{\pi-2}{2\pi} + \frac{\gamma_5}{4}\right)\Bigg]\phi_t \\
&\quad + \left[\gamma_2^2\frac{\pi-2}{\pi} + \frac{\gamma_3^2}{4}\left(\frac{\pi-2}{\pi}+1\right) + 2\gamma_2\gamma_3\frac{\pi-2}{2\pi}\right]\phi_t^2\Bigg\}^{\frac{1}{2}} \quad (5.3.32\text{a})
\end{aligned}
$$

$$
\rho_t = v_t^{-1}\left(-\frac{\gamma_5}{2}-\phi_t\frac{\gamma_3}{2}\right) = \frac{-\gamma_5-\gamma_3\phi_t}{2v_t}. \quad (5.3.32\text{b})
$$

This is a far as one can get when working with *both* the model of Black, Scholes and Courtadon and the Scott, Stein & Stein and Heston model.

5.3.4.1 Black, Scholes and Courtadon

Deferring the Scott, Stein & Stein and Heston model to a later subsection, the parameters leading to the Black, Scholes and Courtadon model may be identified by comparing (5.3.31) and (5.3.32) to the correlated Black, Scholes and Courtadon model from (5.1.6) which is restated here for the sake of simplicity:

$$
d\ln S_t = \left(\alpha - \frac{1}{2}\sigma_t^2\right)dt + \sigma_t d\tilde{W}_{1,t} \quad (5.3.33\text{a})
$$

$$
d\sigma_t = \kappa(\beta - \sigma_t)dt + \rho\theta\sigma_t d\tilde{W}_{1,t} + \sqrt{1-\rho^2}\,\theta\sigma_t d\tilde{W}_{2,t}. \quad (5.3.33\text{b})
$$

Unfortunately, the comparison is inhibited by Equation 5.3.19 by the way it transforms ϕ before it is used as a diffusion. To remedy this difficulty, ϕ is transformed using the transformation

$$
\phi_t^* = \frac{\phi_t + 1}{2} \qquad \Leftrightarrow \qquad \phi_t = 2\phi_t^* - 1 \quad (5.3.34)
$$

which enables one to write

$$
\tilde{h}_t = \phi_t^{*2} \quad (5.3.35)
$$

and as $\sqrt{\tilde{h}_t}$ is the diffusion of $\ln S_t$, then so is ϕ_t^* if it is required that $\phi_t^* > 0$. It is readily seen that

$$\frac{d\phi_t^*}{d\phi_t} = \frac{1}{2} \qquad \Leftrightarrow \qquad d\phi_t^* = \frac{d\phi_t}{2} \tag{5.3.36}$$

as the diffusion dependent part of the drift of the Itô formula (Theorem 5.3) vanishes since $\frac{d^2\phi_t^*}{d\phi_t^2} = 0$. Inserting (5.3.34) and (5.3.36) into (5.3.31) yields the following expression for $d\phi_t^*$:

$$
\begin{aligned}
d\phi_t^* &= \left[\frac{\gamma_0}{2} + \frac{2\gamma_4 + \gamma_5}{4}\sqrt{\frac{2}{\pi}} \Leftrightarrow \frac{\gamma_4 + \gamma_5}{2} \right.\\
&\quad + \left. \left(\gamma_1 + \gamma_2\sqrt{\frac{2}{\pi}} + \frac{\gamma_3}{\sqrt{2\pi}} \Leftrightarrow 1 \right) \left(\phi_t^* \Leftrightarrow \frac{1}{2} \right) \right] dt \\
&\quad + \frac{v_t\rho_t}{2}d\tilde{W}_{1,t} + \frac{v_t\sqrt{1 \Leftrightarrow \rho_t^2}}{2}d\tilde{W}_{2,t} \\
&= \left[\frac{\gamma_0}{2} + \frac{2\gamma_4 + \gamma_5}{4}\sqrt{\frac{2}{\pi}} \Leftrightarrow \frac{\gamma_4 + \gamma_5}{2} \Leftrightarrow \frac{\gamma_1}{2} \Leftrightarrow \frac{\gamma_2}{\sqrt{2\pi}} \Leftrightarrow \frac{\gamma_3}{2\sqrt{2\pi}} + \frac{1}{2} \right.\\
&\quad + \left. \left(\gamma_1 + \gamma_2\sqrt{\frac{2}{\pi}} + \frac{\gamma_3}{\sqrt{2\pi}} \Leftrightarrow 1 \right) \phi_t^* \right] dt \\
&\quad + \frac{v_t\rho_t}{2}d\tilde{W}_{1,t} + \frac{v_t\sqrt{1 \Leftrightarrow \rho_t^2}}{2}d\tilde{W}_{2,t} \tag{5.3.37}
\end{aligned}
$$

By comparing this huge expression to (5.3.33b) it is observed that $\frac{1}{2}v_t\rho_t = \rho\theta\sigma_t$, i.e. that $\frac{1}{2}v_t\rho_t$ must be proportional to σ_t. Using (5.3.32),

$$\frac{v_t\rho_t}{2} = \Leftrightarrow\frac{\gamma_5 + \gamma_3\phi_t}{4} = \Leftrightarrow\frac{\gamma_5 + \gamma_3(2\phi_t^* \Leftrightarrow 1)}{4} = \Leftrightarrow\frac{\gamma_5 \Leftrightarrow \gamma_3 + 2\gamma_3\phi_t^*}{4} \tag{5.3.38}$$

and from here it is obvious that if this expression is to be proportional to σ_t, the restriction $\gamma_5 = \gamma_3$ have to be imposed.

By implementing this restriction and by using an approach identical to the one above, the fact that $\frac{v_t\sqrt{1-\rho_t^2}}{2}$ must proportional to σ_t introduces yet another parameter restriction. This is recognized as

$$\frac{v_t\sqrt{1 \Leftrightarrow \rho_t^2}}{2} \propto \sigma_t \qquad \Leftrightarrow \qquad \frac{v_t^2 \Leftrightarrow v_t^2\rho_t^2}{4} \propto \phi_t^{*2}, \tag{5.3.39}$$

$$v_t^2 \rho_t^2 = \gamma_3^2 \phi_t^{*2} \tag{5.3.40}$$

and, by using (5.3.32)

$$
\begin{aligned}
v_t^2 = {} & \left(\gamma_4 + \frac{\gamma_3}{2}\right)^2 \frac{\pi - 2}{\pi} + \frac{\gamma_3^2}{4} \\
& + 2\left[\gamma_2\left(\gamma_4 + \frac{\gamma_3}{2}\right)\frac{\pi - 2}{\pi}\right. \\
& \left. + \gamma_3\left(\left(\gamma_4 + \frac{\gamma_3}{2}\right)\frac{\pi - 2}{2\pi} + \frac{\gamma_3}{4}\right)\right](2\phi_t^* - 1) \\
& + \left[\gamma_2^2\frac{\pi - 2}{\pi} + \frac{\gamma_3^2}{4}\left(\frac{\pi - 2}{\pi} + 1\right) + 2\gamma_2\gamma_3\frac{\pi - 2}{2\pi}\right](2\phi_t^* - 1)^2
\end{aligned}
\tag{5.3.41}
$$

and by grouping the expressions as

$$
\begin{aligned}
v_t^2 &- v_t^2 \rho_t^2 \\
= {} & \left(\gamma_4 + \frac{\gamma_3}{2}\right)^2 \frac{\pi - 2}{\pi} + \frac{\gamma_3^2}{4} \\
& - 2\left[\gamma_2\left(\gamma_4 + \frac{\gamma_3}{2}\right)\frac{\pi - 2}{\pi} + \gamma_3\left(\left(\gamma_4 + \frac{\gamma_3}{2}\right)\frac{\pi - 2}{2\pi} + \frac{\gamma_3}{4}\right)\right] \\
& + \gamma_2^2\frac{\pi - 2}{\pi} + \frac{\gamma_3^2}{4}\left(\frac{\pi - 2}{\pi} + 1\right) + 2\gamma_2\gamma_3\frac{\pi - 2}{2\pi} \\
& + 4\left[\gamma_2\left(\gamma_4 + \frac{\gamma_3}{2}\right)\frac{\pi - 2}{\pi} + \gamma_3\left(\left(\gamma_4 + \frac{\gamma_3}{2}\right)\frac{\pi - 2}{2\pi} + \frac{\gamma_3}{4}\right)\right. \\
& \left. - \gamma_2^2\frac{\pi - 2}{\pi} - \frac{\gamma_3^2}{4}\left(\frac{\pi - 2}{\pi} + 1\right) - 2\gamma_2\gamma_3\frac{\pi - 2}{2\pi}\right]\phi_t^* \\
& + \left\{4\left[\gamma_2^2\frac{\pi - 2}{\pi} + \frac{\gamma_3^2}{4}\left(\frac{\pi - 2}{\pi} + 1\right) + 2\gamma_2\gamma_3\frac{\pi - 2}{2\pi}\right] - \gamma_3^2\right\}\phi_t^{*2}.
\end{aligned}
\tag{5.3.42}
$$

If this is to be proportional to ϕ_t^{*2}, the constant and the ϕ_t^*-factor must evaluate to zero. Focusing on the ϕ_t^*-factor this means that

$$
\begin{aligned}
& \gamma_2\left(\gamma_4 + \frac{\gamma_3}{2}\right)\frac{\pi - 2}{\pi} + \gamma_3\left(\left(\gamma_4 + \frac{\gamma_3}{2}\right)\frac{\pi - 2}{2\pi} + \frac{\gamma_3}{4}\right) \\
& = \gamma_2^2\frac{\pi - 2}{\pi} + \frac{\gamma_3^2}{4}\left(\frac{\pi - 2}{\pi} + 1\right) + 2\gamma_2\gamma_3\frac{\pi - 2}{2\pi}
\end{aligned}
\tag{5.3.43}
$$

which is equivalent to

$$\gamma_2\gamma_4\frac{\pi \Leftrightarrow 2}{\pi} + \gamma_2\gamma_3\frac{\pi \Leftrightarrow 2}{2\pi} + \gamma_3\gamma_4\frac{\pi \Leftrightarrow 2}{2\pi} + \frac{\gamma_3^2}{4}\frac{\pi \Leftrightarrow 2}{\pi} + \frac{\gamma_3^2}{4}$$

$$= \gamma_2^2\frac{\pi \Leftrightarrow 2}{\pi} + \frac{\gamma_3^2}{4}\frac{\pi \Leftrightarrow 2}{\pi} + \frac{\gamma_3^2}{4} + \gamma_2\gamma_3\frac{\pi \Leftrightarrow 2}{\pi} \qquad (5.3.44)$$

which again is equivalent to

$$\gamma_2\gamma_4\frac{\pi \Leftrightarrow 2}{\pi} + \gamma_2\gamma_3\frac{\pi \Leftrightarrow 2}{2\pi} + \gamma_3\gamma_4\frac{\pi \Leftrightarrow 2}{2\pi}$$

$$= \gamma_2^2\frac{\pi \Leftrightarrow 2}{\pi} + \gamma_2\gamma_3\frac{\pi \Leftrightarrow 2}{\pi} \qquad (5.3.45)$$

and

$$\gamma_2\gamma_4 + \frac{\gamma_2\gamma_3}{2} + \frac{\gamma_3\gamma_4}{2} \;\; = \;\; \gamma_2^2 + \gamma_2\gamma_3$$

$$\Updownarrow$$

$$\gamma_4\left(\gamma_2 + \frac{\gamma_3}{2}\right) \;\; = \;\; \gamma_2\left(\gamma_2 + \frac{\gamma_3}{2}\right) \qquad (5.3.46)$$

The least restrictive way to assure this identity is to impose the restriction $\gamma_4 = \gamma_2$.
When this restriction is combined with the constant part of (5.3.42), as in

$$\gamma_2^2\frac{\pi \Leftrightarrow 2}{\pi} + \gamma_2\gamma_3\frac{\pi \Leftrightarrow 2}{\pi} + \frac{\gamma_3^2}{4}\frac{\pi \Leftrightarrow 2}{\pi} + \frac{\gamma_3^2}{4}$$

$$\Leftrightarrow 2\left[\gamma_2^2\frac{\pi \Leftrightarrow 2}{\pi} + \gamma_2\gamma_3\frac{\pi \Leftrightarrow 2}{2\pi} + \gamma_2\gamma_3\frac{\pi \Leftrightarrow 2}{2\pi} + \frac{\gamma_3^2}{4}\frac{\pi \Leftrightarrow 2}{\pi} + \frac{\gamma_3^2}{4}\right]$$

$$+\gamma_2^2\frac{\pi \Leftrightarrow 2}{\pi} + \frac{\gamma_3^2}{4}\frac{\pi \Leftrightarrow 2}{\pi} + \frac{\gamma_3^2}{4} + \gamma_2\gamma_3\frac{\pi \Leftrightarrow 2}{\pi} \qquad (5.3.47)$$

it is noted that the restriction causes the constant to vanish.

The remains of $v_t^2 - v_t^2\rho_t^2$ are

$$
\begin{aligned}
v_t^2 &- v_t^2\rho_t^2 \\
&= \left[4\left(\gamma_2^2 \frac{\pi-2}{\pi} + \frac{\gamma_3^2}{4}\frac{\pi-2}{\pi} + \frac{\gamma_3^2}{4} + \gamma_2\gamma_3\frac{\pi-2}{\pi} \right) - \gamma_3^2 \right] \phi_t^{*2} \\
&= 4\frac{\pi-2}{\pi}\left(\gamma_2^2 + \frac{\gamma_3^2}{4} + \gamma_2\gamma_3 \right)\phi_t^{*2} \\
&= \frac{\pi-2}{\pi}\left(4\gamma_2^2 + \gamma_3^2 + 4\gamma_2\gamma_3 \right)\phi_t^{*2} \\
&= \frac{\pi-2}{\pi}\left(2\gamma_2 + \gamma_3 \right)^2 \phi_t^{*2} \quad (5.3.48)
\end{aligned}
$$

which is the same as

$$
v_t\sqrt{1 - \rho_t^2} = \sqrt{\frac{\pi-2}{\pi}}\,|2\gamma_2 + \gamma_3|\,\phi_t^{*2}. \quad (5.3.49)
$$

The stationarity conditions from Theorem 3.3 ($\gamma_2 \geq 0$ and $\gamma_2 + \gamma_3 \geq 0$) implies that $2\gamma_2 + \gamma_3 \geq 0$ which means that

$$
v_t\sqrt{1 - \rho_t^2} = \sqrt{\frac{\pi-2}{\pi}}\,(2\gamma_2 + \gamma_3)\,\phi_t^{*2}. \quad (5.3.50)
$$

Finally, it is possible to write down an expression for $d\phi_t^*$ that matches the expression for $d\sigma_t$ from the Black, Scholes and Courtadon model:

$$
\begin{aligned}
d\phi_t^* &= \left[\frac{\gamma_0}{2} + \frac{2\gamma_2+\gamma_3}{4}\sqrt{\frac{2}{\pi}} - \frac{\gamma_2+\gamma_3}{2} - \frac{\gamma_1}{2} - \frac{\gamma_2}{\sqrt{2\pi}} - \frac{\gamma_3}{2\sqrt{2\pi}} + \frac{1}{2} \right. \\
&\quad \left. + \left(\gamma_1 + \gamma_2\sqrt{\frac{2}{\pi}} + \frac{\gamma_3}{\sqrt{2\pi}} - 1 \right)\phi_t^* \right] dt \\
&\quad + \frac{v_t\rho_t}{2}d\tilde{W}_{1,t} + \frac{v_t\sqrt{1-\rho_t^2}}{2}d\tilde{W}_{2,t} \\
&= \left[\frac{\gamma_0 - \gamma_1 - \gamma_2 - \gamma_3 + 1}{2} + \left(\gamma_1 + \frac{2\gamma_2+\gamma_3}{\sqrt{2\pi}} - 1 \right)\phi_t^* \right] dt \\
&\quad - \frac{\gamma_3}{2}\phi_t^* d\tilde{W}_{1,t} + \frac{2\gamma_2+\gamma_3}{2}\sqrt{\frac{\pi-2}{\pi}}\,\phi_t^* d\tilde{W}_{2,t} \quad (5.3.51)
\end{aligned}
$$

When this expression is compared to the diffusion part of the Black, Scholes and Courtadon model

$$d\sigma_t = \kappa(\beta \Leftrightarrow \sigma_t) + \rho\theta\sigma_t d\tilde{W}_{1,t} + \sqrt{1 \Leftrightarrow \rho^2}\theta\sigma_t d\tilde{W}_{2,t}, \qquad (5.3.52)$$

it is observed that the two diffusion processes are identical. If in the approximating augmented GARCH model $\omega(\tilde{h}_t) = \alpha \Leftrightarrow \frac{1}{2}h_t$ is inserted, the expression for $d\ln S_t$ is

$$d\ln S_t = \left(\alpha \Leftrightarrow \frac{1}{2}\tilde{h}_t\right) dt + \sqrt{\tilde{h}_t}d\tilde{W}_{1,t}, \qquad (5.3.53)$$

which is identical to

$$d\ln S_t = \left(\alpha \Leftrightarrow \frac{1}{2}\sigma_t^2\right) dt + \sigma_t d\tilde{W}_{1,t} \qquad (5.3.54)$$

of the Black, Scholes and Courtadon model as $\sqrt{\tilde{h}_t} = \phi_t^* = \sigma_t$.

The above findings constitutes the proof of

THEOREM 5.5 (BLACK, SCHOLES AND COURTADON).
The diffusion limit of the approximating GARCH(1,1) process with $\xi = 1/2$, $c = 0$, $\delta = 1$, $\gamma_5 = \gamma_3$, $\gamma_4 = \gamma_2$, $\omega(\tilde{h}_t) = \alpha \Leftrightarrow \frac{1}{2}\tilde{h}_t$ and $\gamma_1 \geq 0$, $\gamma_2 \geq 0$, $\gamma_2 + \gamma_3 \geq 0$ and $\gamma_0 > \gamma_1 + \gamma_2 + \gamma_3 \Leftrightarrow 1$ is a Black, Scholes and Courtadon model.

REMARK 5.2. The inequality constraints imposed on the parameters are to ensure stationarity and non-negativity of ϕ^*.   ▼

Though (5.3.51) suffices for most purposes, it is sometimes necessary to know the exact relationship between augmented GARCH model parameters and the parameters of the Black, Scholes and Courtadon model:

$$dS_t = \alpha S_t dt + \sigma_t S_t d\tilde{W}_{1,t} \qquad (5.3.55a)$$
$$d\sigma_t = \kappa(\beta \Leftrightarrow \sigma_t)dt + \rho\theta\sigma_t d\tilde{W}_{1,t} + \sqrt{1 \Leftrightarrow \rho^2}\theta\sigma_t d\tilde{W}_{2,t} \qquad (5.3.55b)$$

This relation is provided in Theorem 5.6.

THEOREM 5.6 (BLACK, SCHOLES AND COURTADON PARAMETERS).
The relationship between augmented GARCH(1,1) and Black, Scholes and Courtadon parameters,

$$\kappa \;=\; 1 \Leftrightarrow \gamma_1 \Leftrightarrow \frac{2\gamma_2 + \gamma_3}{\sqrt{2\pi}} \tag{5.3.56a}$$

$$\beta \;=\; \frac{\gamma_0 \Leftrightarrow \gamma_1 \Leftrightarrow \gamma_2 \Leftrightarrow \gamma_3 + 1}{2\left(1 \Leftrightarrow \gamma_1 \Leftrightarrow \frac{2\gamma_2+\gamma_3}{\sqrt{2\pi}}\right)} \tag{5.3.56b}$$

$$\rho \;=\; \frac{\Leftrightarrow \gamma_3}{2\sqrt{\frac{\gamma_3^2}{4} + \left(\frac{2\gamma_2+\gamma_3}{2}\right)^2 \frac{\pi-2}{\pi}}} \tag{5.3.56c}$$

$$\theta \;=\; \sqrt{\frac{\gamma_3^2}{4} + \left(\frac{2\gamma_2 + \gamma_3}{2}\right)^2 \frac{\pi \Leftrightarrow 2}{\pi}} \tag{5.3.56d}$$

Proof. The proof is based on a comparison of (5.3.51) to (5.3.52).

κ. By comparison,

$$\Leftrightarrow \kappa = \gamma_1 + \frac{2\gamma_2 + \gamma_3}{\sqrt{2\pi}} \Leftrightarrow 1. \tag{5.3.57}$$

β. By comparison,

$$\kappa\beta = \frac{\gamma_0 \Leftrightarrow \gamma_1 \Leftrightarrow \gamma_2 \Leftrightarrow \gamma_3 + 1}{2}. \tag{5.3.58}$$

Isolating β from this equation and inserting the expression for κ yields the desired result.

ρ and θ. By comparison,

$$\rho\theta = \frac{\Leftrightarrow \gamma_3}{2} \tag{5.3.59}$$

and

$$\sqrt{1 \Leftrightarrow \rho^2}\,\theta = \frac{2\gamma_2 + \gamma_3}{2}\sqrt{\frac{\pi \Leftrightarrow 2}{\pi}}. \tag{5.3.60}$$

Note that by the second equation it may be deduced that $\theta > 0$ as $2\gamma_2 + \gamma_3 \geq 0$ by the stationarity conditions. Isolating ρ from the first equation yields

$$\rho = \frac{-\gamma_3}{2\theta} \tag{5.3.61}$$

and inserting into the second equation yields

$$\sqrt{1 - \rho^2}\,\theta = \sqrt{1 - \frac{\gamma_3^2}{4\theta^2}}\,\theta = \sqrt{\theta^2 - \frac{\gamma_3^2}{4}} \tag{5.3.62}$$

and by comparison,

$$\sqrt{\theta^2 - \frac{\gamma_3^2}{4}} = \frac{2\gamma_2 + \gamma_3}{2}\sqrt{\frac{\pi - 2}{\pi}} \tag{5.3.63}$$

or

$$\theta^2 = \frac{(2\gamma_2 + \gamma_3)^2}{4}\frac{\pi - 2}{\pi} + \frac{\gamma_3^2}{4}. \tag{5.3.64}$$

The θ value of Theorem 5.6 follows from this expression and the fact that $\theta > 0$. The value of ρ is obtained by inserting the value of θ into (5.3.61).

Discrete time counterpart

To find out which Arch process that matches the parameter constraints of Theorem 5.5, the constraints are incorporated into the augmented GARCH(1,1) process of Definition 5.3. This maneuver results in the following expression of ϕ_t:

$$\begin{aligned}
\phi_t ={}& \gamma_0 + [\gamma_1 + \gamma_2|z_{t-1}| + \gamma_3 \max(0, -z_{t-1})]\phi_{t-1} \\
& + \gamma_2(|z_{t-1}| - 1) + \gamma_3[\max(0, z_{t-1}) - 1]
\end{aligned} \tag{5.3.65}$$

Incorporation the transformation of Equation 5.3.34 into this expression, one can write

$$
\begin{aligned}
\phi_t^* &= \frac{\gamma_0 + 1}{2} + \left[\gamma_1 + \gamma_2 |z_{t-1}| + \gamma_3 \max(0, -z_{t-1})\right]\left(\phi_{t-1}^* - \frac{1}{2}\right) \\
&\quad + \frac{\gamma_2}{2}\left(|z_{t-1}| - 1\right) + \frac{\gamma_3}{2}\left[\max(0, -z_{t-1}) - 1\right] \\
&= \frac{\gamma_0 + 1}{2} - \frac{\gamma_1}{2} - \frac{\gamma_2}{2}|z_{t-1}| - \frac{\gamma_3}{2}\max(0, z_{t-1}) \\
&\quad + \frac{\gamma_2}{2}|z_{t-1}| - \frac{\gamma_2}{2} + \frac{\gamma_3}{2}\max(0, -z_{t-1}) - \frac{\gamma_3}{2} \\
&\quad + \left[\gamma_1 + \gamma_2|z_{t-1}| + \gamma_3 \max(0, z_{t-1})\right]\phi_{t-1}^* \\
&= \frac{\gamma_0 - \gamma_1 - \gamma_2 - \gamma_3 + 1}{2} + \gamma_1 \phi_{t-1}^* + \gamma_2 |z_{t-1}| \phi_{t-1}^* \\
&\quad + \gamma_3 \max(0, -z_{t-1}) \phi_{t-1}^*.
\end{aligned} \tag{5.3.66}
$$

Since $\phi_t^* = \sqrt{\tilde{h}_t}$, this is the same as

$$
\begin{aligned}
\sqrt{\tilde{h}_t} &= \frac{\gamma_0 - \gamma_1 - \gamma_2 - \gamma_3 + 1}{2} + \gamma_1 \sqrt{\tilde{h}_{t-1}} + \gamma_2 \sqrt{\tilde{h}_{t-1}}|z_{t-1}| \\
&\quad + \gamma_3 \sqrt{\tilde{h}_{t-1}} \max(0, -z_{t-1}),
\end{aligned} \tag{5.3.67}
$$

which is identical to the TGARCH(1,1) process of (3.5.10)

$$
\sqrt{h_t} = \alpha_0 + \alpha_1^{(1)} |\varepsilon_{t-1}| + \alpha_1^{(2)} \max(0, -\varepsilon_{t-1}) + \beta_1 \sqrt{h_{t-1}}, \tag{5.3.68}
$$

if the following parameter relationships are introduced:

$$
\begin{aligned}
\alpha_0 &= \frac{\gamma_0 - \gamma_1 - \gamma_2 - \gamma_3 + 1}{2} & \text{(5.3.69a)} \\
\alpha_1^{(1)} &= \gamma_2 & \text{(5.3.69b)} \\
\alpha_1^{(2)} &= \gamma_3 & \text{(5.3.69c)} \\
\beta_1 &= \gamma_1. & \text{(5.3.69d)}
\end{aligned}
$$

It is also noted that inequality constraints of the TGARCH parameters of Section 3.5.5 are matching the inequality constraints of Theorem 5.5.

Thus we have

THEOREM 5.7 (TGARCH AND BLACK, SCHOLES AND COURTADON).
The augmented GARCH(1,1) process that coincide with the TGARCH(1,1) process when $s = 1$ converges weakly to the Black, Scholes and Courtadon diffusion process.

From Black, Scholes and Courtadon to TGARCH(1,1)

It is, of course, also possible to establish a direct relationship between the parameters of the Black, Scholes and Courtadon model and the TGARCH(1,1) parameters, without the diversion via the augmented GARCH(1,1) parameters. From Equation 5.3.58 we have

$$\kappa\beta = \frac{\gamma_0 \Leftrightarrow \gamma_1 \Leftrightarrow \gamma_2 \Leftrightarrow \gamma_3 + 1}{2} \tag{5.3.70}$$

which is exactly α_0 of TGARCH. From Equation 5.3.59 the relationship

$$\rho\theta = \frac{\Leftrightarrow \gamma_3}{2} \tag{5.3.71}$$

is obtained, and since $\alpha_1^{(2)}$ of TGARCH is equal to γ_3 of augmented GARCH, it is readily seen that

$$\alpha_1^{(2)} = \Leftrightarrow 2\rho\theta. \tag{5.3.72}$$

It is a bit more difficult to deduce an expression for $\alpha_1^{(1)}$ of TGARCH which is equal to γ_2 of augmented GARCH, but, by Equation 5.3.60,

$$\sqrt{1 \Leftrightarrow \rho^2}\theta = \frac{2\gamma_2 + \gamma_3}{2}\sqrt{\frac{\pi \Leftrightarrow 2}{\pi}}, \tag{5.3.73}$$

and since it is established that $\Leftrightarrow \gamma_3/2 = \rho\theta$, one can write

$$\alpha_1^{(1)} = \rho\theta + \sqrt{\frac{\pi}{\pi \Leftrightarrow 2}}\sqrt{1 \Leftrightarrow \rho^2}\theta. \tag{5.3.74}$$

Finally, an expression for $\beta_1 = \gamma_1$ is deduced from Equation 5.3.57:

$$
\begin{aligned}
\beta_1 &= 1 - \kappa + \frac{2\gamma_2 + \gamma_3}{\sqrt{2\pi}} \\
&= 1 - \kappa + \frac{1}{\sqrt{2\pi}} 2\sqrt{\frac{\pi}{\pi - 2}} \sqrt{1 - \rho^2}\,\theta \\
&= 1 - \kappa + \sqrt{\frac{2}{\pi - 2}} \sqrt{1 - \rho^2}\,\theta.
\end{aligned}
\tag{5.3.75}
$$

Observe that since θ and α_1 are non-negative, ρ is non-positive.

From TGARCH(1,1) to Black, Scholes and Courtadon

The above formulae (5.3.70), (5.3.72), (5.3.74) and (5.3.75) make it possible to compute the corresponding values of the TGARCH(1,1) parameters from the Black, Scholes and Courtadon parameters. This may be useful in some situations, model comparisons for example, but the purpose of relating discrete time Arch models to continuous time stochastic volatility models is to be able to compute the parameters of the continuous time model from the discrete time model parameters. Thus, it is important to have formulas for this purpose. Fortunately, these formulas are very easy to obtain. Just merge (5.3.69) and (5.3.56) to obtain

$$
\kappa = 1 - \beta_1 - \frac{2\alpha_1^{(1)} + \alpha_1^{(2)}}{\sqrt{2\pi}}
\tag{5.3.76a}
$$

$$
\beta = \frac{\alpha_0}{1 - \beta_1 - \frac{2\alpha_1^{(1)} + \alpha_1^{(2)}}{\sqrt{2\pi}}}
\tag{5.3.76b}
$$

$$
\rho = \frac{-\alpha_1^{(2)}}{\sqrt{\alpha_1^{(2)2} + \left(2\alpha_1^{(1)} + \alpha_1^{(2)}\right)^2 \frac{\pi - 2}{\pi}}}
\tag{5.3.76c}
$$

$$
\theta = \sqrt{\frac{\alpha_1^{(2)2}}{4} + \left(\frac{2\alpha_1^{(1)} + \alpha_1^{(2)}}{2}\right)^2 \frac{\pi - 2}{\pi}}
\tag{5.3.76d}
$$

5.3.4.2   Scott, Stein & Stein and Heston

The model of Scott, Stein & Stein and Heston stated by use of only uncorrelated Wiener processes is

$$d \ln S_t = \left(\alpha \Leftrightarrow \frac{1}{2}\sigma_t^2 \right) + \sigma_t d\tilde{W}_{1,t} \tag{5.3.77a}$$

$$d\sigma_t = \kappa(\beta \Leftrightarrow \sigma_t)dt + \rho\theta\, d\tilde{W}_{1,t} + \sqrt{1 \Leftrightarrow \rho^2}\theta d\tilde{W}_{2,t}. \tag{5.3.77b}$$

As previously announced, the deduction of the discrete time Arch model corresponding to the continuous time stochastic volatility model of Scott, Stein & Stein and Heston will be treated with less rigour than the deduction of the Arch model corresponding to the Black, Scholes and Courtadon model of the last subsection, because the general idea should be clear by now. In this subsection, the augmented GARCH parameter restrictions leading to the Scott, Stein & Stein and Heston model will be stated instead of deduced.

THEOREM 5.8 (SCOTT, STEIN & STEIN AND HESTON).
The diffusion limit of the approximating augmented GARCH process with $\xi = 1/2$, $c = 0$, $\delta = 1$ and $\gamma_2 = \gamma_3 = 0$, $\omega(\tilde{h}_t) = \alpha \Leftrightarrow \frac{1}{2}\tilde{h}_t$ is a Scott, Stein & Stein and Heston model.

Proof. The parameter constraints imposed on ξ, c and δ are identical to those of the Black, Scholes and Courtadon model. Thus, the values of the Z's of (5.3.20), the q's of (5.3.21) and V_Z of (5.2.18) may be used here as well as the auxiliary variable ϕ_t^* of (5.3.34).

Incorporating the parameter constraints into the diffusion limit of the approximating augmented GARCH process of Theorem 5.2 yields

$$
\begin{aligned}
v_t &= \sqrt{\sigma_4^2 + 2(\gamma_2\sigma_{24} + \gamma_3\sigma_{34})\phi_t + (\gamma_2^2\sigma_2^2 + \gamma_3^2\sigma_3^2 + 2\gamma_2\gamma_3\sigma_{23})\phi_t^2} \\
&= \sqrt{\sigma_4^2} = \sqrt{\left(\gamma_4 + \frac{\gamma_5}{2}\right)^2 \frac{\pi \Leftrightarrow 2}{\pi} + \frac{\gamma_5^2}{4}}
\end{aligned} \tag{5.3.78}
$$

and

$$
\begin{aligned}
\rho_t &= v_t^{-1}\left[\sigma_{14} + \phi_t(\gamma_2\sigma_{12} + \gamma_3\sigma_{13})\right] \\
&= v_t^{-1}\sigma_{14} = \Leftrightarrow\frac{\gamma_5}{2}v_t^{-1}.
\end{aligned} \tag{5.3.79}
$$

so that

$$\rho_t v_t = \Leftrightarrow \frac{\gamma_5}{2} \tag{5.3.80}$$

and

$$
\begin{aligned}
v_t\sqrt{1 \Leftrightarrow \rho_t^2} &= \sqrt{v_t^2 \Leftrightarrow \rho_t^2 v_t^2} = \sqrt{\left(\gamma_4 + \frac{\gamma_5}{2}\right)^2 \frac{\pi \Leftrightarrow 2}{\pi} + \frac{\gamma_5^2}{4} \Leftrightarrow \frac{\gamma_5^2}{4}} \\
&= \left|\frac{2\gamma_4 + \gamma_5}{2}\right| \sqrt{\frac{\pi \Leftrightarrow 2}{\pi}}.
\end{aligned}
\tag{5.3.81}
$$

Using these expressions, an explicit expression for $d\phi_t$ is available as

$$
\begin{aligned}
d\phi_t &= [\gamma_0 + q_4 + (\gamma_1 + \gamma_2 q_2 + \gamma_3 q_3 \Leftrightarrow 1)\phi_t]\, dt \\
&\quad + v_t \rho_t d\tilde{W}_{1,t} + v_t\sqrt{1 \Leftrightarrow \rho_t^2}\, d\tilde{W}_{2,t} \\
&= \left[\gamma_0 + \left(\gamma_4 + \frac{\gamma_5}{2}\right)\sqrt{\frac{2}{\pi}} \Leftrightarrow (\gamma_4 + \gamma_4) + (\gamma_1 \Leftrightarrow 1)\phi_t\right] dt \\
&\quad \Leftrightarrow \frac{\gamma_5}{2}d\tilde{W}_{1,t} + \left|\frac{2\gamma_4 + \gamma_5}{2}\right|\sqrt{\frac{\pi \Leftrightarrow 2}{\pi}}d\tilde{W}_{2,t}
\end{aligned}
\tag{5.3.82}
$$

and if this is expressed using ϕ^* instead of ϕ it appears as

$$
\begin{aligned}
d\phi_t^* &= \left[\frac{\gamma_0 \Leftrightarrow \gamma_1 + \left(\gamma_4 + \frac{\gamma_5}{2}\right)\sqrt{\frac{2}{\pi}} \Leftrightarrow \gamma_4 \Leftrightarrow \gamma_5 + 1}{2} + (\gamma_1 \Leftrightarrow 1)\phi_t^*\right] dt \\
&\quad \Leftrightarrow \frac{\gamma_5}{4}d\tilde{W}_{1,t} + \frac{|2\gamma_4 + \gamma_5|}{4}\sqrt{\frac{\pi \Leftrightarrow 2}{\pi}}d\tilde{W}_{2,t}.
\end{aligned}
\tag{5.3.83}
$$

Since $\sqrt{\tilde{h}_t} = \phi_t^* = \sigma_t$ if ϕ_t^* is non-negative, it is by the expression for $d\phi_t^*$ seen that for correctly chosen parameter values, $d\phi_t^*$ is identical to $d\sigma_t$ of the Scott, Stein & Stein and Heston model. This concludes the proof of Theorem 5.8.

The relationship between the parameters of the Scott, Stein & Stein and Heston process and the parameters of the approximating augmented GARCH(1,1) process is obtained by relating the expression of $d\phi_t^*$ to the expression of $d\sigma_t$ in (5.3.77).

THEOREM 5.9 (SCOTT, STEIN & STEIN AND HESTON PARAMETERS).
The relationship between augmented GARCH(1,1) and Scott, Stein & Stein and Heston parameters is

$$\kappa = 1 \Leftrightarrow \gamma_1 \tag{5.3.84a}$$

$$\beta = \frac{\gamma_0 \Leftrightarrow \gamma_1 + \left(\gamma_4 + \frac{\gamma_5}{2}\right)\sqrt{\frac{2}{\pi}} \Leftrightarrow \gamma_4 \Leftrightarrow \gamma_5 + 1}{2(1 \Leftrightarrow \gamma_1)} \tag{5.3.84b}$$

$$\rho = \frac{\Leftrightarrow \gamma_5}{\sqrt{\left(\frac{2\gamma_4 + \gamma_5}{2}\right)^2 \frac{\pi-2}{\pi} + \frac{\gamma_5^2}{4}}} \tag{5.3.84c}$$

$$\theta = \sqrt{\left(\frac{2\gamma_4 + \gamma_5}{4}\right)^2 \frac{\pi \Leftrightarrow 2}{\pi} + \frac{\gamma_5^2}{16}} \tag{5.3.84d}$$

Proof.

κ. By comparison, it is readily seen that

$$\Leftrightarrow \kappa = \gamma_1 \Leftrightarrow 1 \tag{5.3.85}$$

and by isolating κ from this equation, the results of the theorem is obtained.

β. The constant part of the drifts of the Scott, Stein & Stein and Heston model and ϕ_t^* is

$$\kappa\beta = \frac{\gamma_0 \Leftrightarrow \gamma_1 + \left(\gamma_4 + \frac{\gamma_5}{2}\right)\sqrt{\frac{2}{\pi}} \Leftrightarrow \gamma_4 \Leftrightarrow \gamma_5 + 1}{2} \tag{5.3.86}$$

and by isolating β and inserting the expression of κ, the result of the theorem is available.

ρ and θ. The coefficients of the Wiener processes are

$$\rho\theta = \Leftrightarrow \frac{\gamma_5}{4} \tag{5.3.87}$$

and

$$\sqrt{1 \Leftrightarrow \rho^2}\theta = \sqrt{\theta^2 \Leftrightarrow \rho^2\theta^2} = \frac{|2\gamma_4 + \gamma_5|}{4}\sqrt{\frac{\pi \Leftrightarrow 2}{\pi}}, \tag{5.3.88}$$

from where it is readily seen that $\theta > 0$. Combining the to expressions,

$$
\begin{aligned}
\sqrt{\theta^2 - \frac{\gamma_5^2}{16}} &= \frac{|2\gamma_5 + \gamma_5|}{4}\sqrt{\frac{\pi-2}{\pi}} && \Leftrightarrow \\[2mm]
\theta^2 - \frac{\gamma_5^2}{16} &= \left(\frac{2\gamma_4+\gamma_5}{4}\right)^2 \frac{\pi-2}{\pi} && \Leftrightarrow \\[2mm]
\theta^2 &= \left(\frac{2\gamma_4+\gamma_5}{4}\right)^2 \frac{\pi-2}{\pi} + \frac{\gamma_5^2}{16} && \Leftrightarrow \\[2mm]
\theta &= \sqrt{\left(\frac{2\gamma_4+\gamma_5}{4}\right)^2 \frac{\pi-2}{\pi} + \frac{\gamma_5^2}{16}} && (5.3.89)
\end{aligned}
$$

which proofs the theorem for θ. By inserting the θ-expression in the expression for $\rho\theta$, the proof of Theorem 5.9 is completed.

REMARK 5.3. Note that it has been assumed that ϕ_t^* is non-negative, but that the non-negativity has not been accounted for. However, since the obtained diffusion model is identical to the model of Scott, Stein & Stein and Heston, the properties of the two models are also identical. Thus, if σ_t of the Scott, Stein & Stein and Heston model is assured to be non-negative, then so is ϕ_t^*, and if not, the framework of approximating augmented GARCH and its diffusion limit will surely not add to the lack of mathematical rigour. ▼

Discrete time counterpart

To find out which discrete time Arch model that matches the augmented GARCH(1,1) process of Definition 5.3 with the parameter constraints of Theorem 5.8, the constraints are imposed on the augmented GARCH process. Doing this, the following expression for ϕ_t is obtained

$$
\phi_t = \gamma_0 + \gamma_1\phi_{t-1} + \gamma_4|z_{t-1}| + \gamma_5 \max(0, -z_{t-1}) - (\gamma_4 + \gamma_5). \qquad (5.3.90)
$$

Transforming this equation to an equation of ϕ^* instead of ϕ yields

$$
\begin{aligned}
\phi_t^* &= \frac{\gamma_0 + 1}{2} + \gamma_1\left(\phi_{t-1}^* - \frac{1}{2}\right) + \frac{\gamma_4}{2}|z_{t-1}| + \frac{\gamma_5}{2}\max(0, -z_{t-1}) \\
&\quad - \frac{\gamma_4 + \gamma_5}{2} \\
&= \frac{\gamma_0 - \gamma_1 - \gamma_4 - \gamma_5 + 1}{2} + \frac{\gamma_4}{2}|z_{t-1}| + \frac{\gamma_5}{2}\max(0, -z_{t-1}) \\
&\quad + \gamma_1\phi_{t-1}^*,
\end{aligned}
\tag{5.3.91}
$$

and since $\phi_t^* = \sqrt{\tilde{h}_t}$, this is the same as

$$
\sqrt{\tilde{h}_t} = \alpha_0 + \alpha_1^{(1)}|z_{t-1}| + \alpha_2^{(2)}\max(0, -z_{t-1}) + \beta_1\sqrt{\tilde{h}_{t-1}},
\tag{5.3.92}
$$

when the α's and the β are chosen to be

$$
\begin{aligned}
\alpha_0 &= \frac{\gamma_0 - \gamma_1 - \gamma_4 - \gamma_5 + 1}{2} & \text{(5.3.93a)} \\
\alpha_1^{(1)} &= \frac{\gamma_4}{2} & \text{(5.3.93b)} \\
\alpha_1^{(2)} &= \frac{\gamma_5}{2} & \text{(5.3.93c)} \\
\beta_1 &= \gamma_1. & \text{(5.3.93d)}
\end{aligned}
$$

This model is not recognized as an existing Arch model!

None the less, it fits under the Arch framework in the subclass of models where $\sigma_t = \sqrt{h_t}$. Using the initials of the corresponding continuous time stochastic volatility model, it will be referred to by SSSH-GARCH for the rest of this thesis.

Note that to assure strict stationarity of the SSSH-GARCH process, the constraint $\beta_1 \leq 1$ has to be imposed, c.f. Theorem 3.3 and to assure positivity of $\sqrt{\tilde{h}}$, the constraints $\alpha_0 > 0$, $\alpha_1^{(1)} \geq 0$, $\alpha_1^{(1)} + \alpha_1^{(2)} \geq 0$ and $\beta_1 \geq 0$ are needed.

From SSSH-GARCH to Scott, Stein & Stein and Heston

As with the Black, Scholes and Courtadon model, it is useful to know how to compute the values of the Scott, Stein & Stein and Heston parameters directly from SSSH-GARCH parameter estimates. These relations are obtained by combining (5.3.93) with (5.3.84) to obtain

$$\kappa = 1 \Leftrightarrow \beta_1 \tag{5.3.94a}$$

$$\beta = \frac{\alpha_0 + \frac{2\alpha_1^{(1)} + \alpha_1^{(2)}}{\sqrt{2\pi}}}{1 \Leftrightarrow \beta_1} \tag{5.3.94b}$$

$$\rho = \frac{\Leftrightarrow \alpha_1^{(2)}}{\sqrt{\left(2\alpha_1^{(1)} + \alpha_1^{(2)}\right)^2 \frac{\pi-2}{\pi} + \alpha_1^{(2)2}}} \tag{5.3.94c}$$

$$\theta = \sqrt{\left(\alpha_1^{(1)} + \frac{\alpha_1^{(2)}}{2}\right)^2 \frac{\pi \Leftrightarrow 2}{\pi} + \frac{\alpha_1^{(2)2}}{4}} \tag{5.3.94d}$$

The formulas for the reverse computations, from Scott, Stein & Stein and Heston parameters to SSSH-GARCH parameters, will not be derived, as they are of no use in this thesis. It should be said, however, that the problem has more than one solution due to the squared subexpressions of ρ and θ of (5.3.84).

5.4 Estimation study

As examples of the idea of estimating continuous time stochastic volatility model parameters using discrete time Arch model parameter estimates, this section estimates the parameters of the Black, Scholes and Courtadon and the Scott, Stein & Stein and Heston continuous time stochastic volatility models by fitting the TGARCH and SSSH-GARCH models to the four well-known return series of Hewlett-Packard, Sony, Mobil and Pepsi.

5.4.1 TGARCH and Black, Scholes and Courtadon

The discrete time Arch model to estimate is

$$R_t = \alpha \Leftrightarrow \frac{1}{2}h_t + \varepsilon_t \qquad\qquad (5.4.1a)$$

$$\sqrt{h_t} = \alpha_0 + \alpha_1^{(1)}|\varepsilon_{t-1}| + \alpha_1^{(2)}\max(0, \Leftrightarrow\varepsilon_{t-1}) + \beta_1\sqrt{h_{t-1}} \qquad (5.4.1b)$$

where $\alpha_0 > 0$, $\alpha_1^{(1)} \geq 0$, $\alpha_1^{(1)} + \alpha_1^{(2)} \geq 0$ and $\beta_1 \geq 0$, and where ε_t is Gaussian distributed with zero mean and variance h_t.

Because of the way h_t enters the drift of R_t, the model (5.4.1) does not allow for RPLR estimation. This is because the RPLR method requires that the model can be formulated as $\hat{Y}_{t|t-1} = X_t^\top \theta$ in such a way that the observations Y does not depend on the parameters, and clearly, this is not possible for (5.4.1).

The Maximum Likelihood method, however, is perfectly suitable for estimating the model parameters. Recalling Section 4.2.1, the log-likelihood of (5.4.1) is

$$L_T = \sum_{t=1}^{T} l_t(\theta) \qquad\qquad (5.4.2)$$

where the conditional log-likelihood is

$$l_t(\theta) = \Leftrightarrow\frac{1}{2}\ln(2\pi) \Leftrightarrow \frac{1}{2}\ln(h_t) \Leftrightarrow \frac{1}{2}\frac{\varepsilon_t^2}{h_t} \qquad\qquad (5.4.3)$$

where h_t is computed by (5.4.1b) and where ε_t is computed from the observed return R_t by (5.4.1a).

The Maximum Likelihood estimates of the TGARCH model are presented in Table 5.3 along with their standard deviations and the estimated value of $\sqrt{h_0}$. In line with the findings in Chapter 4, the Maximum Likelihood estimation have included $\sqrt{h_0}$ in order to avoid estimation bias caused by an incorrectly specified h_0.

Using the transformation formulas of (5.3.76), the Black, Scholes and Courtadon model parameter values of Table 5.4 are obtained.

The standard errors of the Black, Scholes and Courtadon model parameters are not available. Theoretically, the standard errors of the Black, Scholes and Courtadon parameter estimates can be computed from the standard error of the TGARCH parameter estimates, but the transformation formulas of (5.3.76) are highly unlinear, a fact that make the computations cumbersome, and because of this it has not been possible to include the formulas in this thesis.

	HP	Sony	Mobil	Pepsi
α	1.4216×10^{-3}	8.0462×10^{-4}	6.4095×10^{-4}	4.4597×10^{-4}
	5.2342×10^{-4}	3.5592×10^{-4}	2.7983×10^{-4}	3.3645×10^{-4}
α_0	1.9743×10^{-4}	1.6284×10^{-4}	4.8634×10^{-4}	1.5471×10^{-3}
	9.2116×10^{-5}	7.6166×10^{-5}	1.6416×10^{-4}	5.4702×10^{-4}
$\alpha_1^{(1)}$	1.0112×10^{-2}	5.9884×10^{-2}	4.4708×10^{-2}	9.9291×10^{-2}
	6.9158×10^{-3}	1.0093×10^{-2}	1.4855×10^{-2}	2.3982×10^{-2}
$\alpha_1^{(2)}$	3.1797×10^{-2}	6.1106×10^{-3}	3.2260×10^{-2}	3.8616×10^{-3}
	1.2252×10^{-2}	5.9542×10^{-3}	1.7725×10^{-2}	6.0415×10^{-3}
β_1	9.7082×10^{-1}	9.4541×10^{-1}	9.1007×10^{-1}	8.2336×10^{-1}
	8.0634×10^{-3}	9.8076×10^{-3}	2.0585×10^{-2}	4.7349×10^{-2}
$\sqrt{h_0}$	2.0185×10^{-2}	1.1175×10^{-2}	1.4878×10^{-2}	8.0363×10^{-3}
	3.7397×10^{-3}	3.2751×10^{-3}	4.6981×10^{-3}	4.2191×10^{-3}

Table 5.3: Maximum Likelihood parameter estimates of TGARCH(1,1) applied to return series.

	HP	Sony	Mobil	Pepsi
α	1.4216×10^{-3}	8.0462×10^{-4}	6.4095×10^{-4}	4.4597×10^{-4}
κ	8.4270×10^{-3}	4.3735×10^{-3}	4.1391×10^{-2}	9.5874×10^{-2}
β	2.3428×10^{-2}	3.7234×10^{-2}	1.1750×10^{-2}	1.6137×10^{-2}
ρ	$\Leftrightarrow 7.1200 \times 10^{-1}$	$\Leftrightarrow 8.0269 \times 10^{-2}$	$\Leftrightarrow 4.0261 \times 10^{-1}$	$\Leftrightarrow 3.1627 \times 10^{-2}$
θ	2.2330×10^{-2}	3.8063×10^{-2}	4.0064×10^{-2}	6.1048×10^{-2}

Table 5.4: Parameter estimates of the Black, Scholes and Courtadon model computed from the TGARCH(1,1) parameter estimates.

5.4.1.1 Validation

In order to decide how the above estimation procedure performs relative to traditional continuous time stochastic volatility model estimation procedures, some way of validating the obtained parameter estimates is needed. The literature in the field of continuous time stochastic volatility models often suggest (i) to measure the ability of the estimated model to describe the return series characteristics, (ii) to compute the fair price of some option on the stocks that were used to estimate the parameters and then compare the obtained option price to prices obtained using other well-known estimation procedures or (iii) to measure to ability of the procedure to estimate parameters of simulated data series. Unfortunately, none of these measures are suitable here.

The first one requires the estimated model to be discretized and simulated before the standardized residuals can be computed and investigated for presence of return series characteristics such as autocorrelation. But when the estimated model is discretized, it is converted from continuous time to discrete time, and because of this, the only model validated using this measure is the discrete time model originally estimated, not the continuous time model in focus.

The second measure is perhaps more suitable, but option pricing formulas are outside the scope of this thesis and besides, this measure requires that alternative estimation procedures are applied to the exact same stock series. In addition, it will be very difficult to decide which estimation procedure to rely on in case the procedures disagree on the option price and difficult to conclude if the two procedures agree. In the latter case, though, ones preference may be decided by other matters such as the amount of required computer resources.

The last measure suffers from the same problem as measure number one. In order to simulate a continuous time stochastic volatility model it has to be discretizised, and then the only thing being measured is the discrete time model.

Because of the above mentioned problems related to model validation, the obtained continuous time stochastic volatility model parameter estimates will not be validated. Instead, the parameter estimates are presented, perhaps to be validated at a later time when a suitable validation procedure has been developed.

5.4.2 SSSH-GARCH and Scott, Stein & Stein and Heston

In the case of Scott, Stein & Stein and Heston, the discrete time Arch model to estimate is

$$R_t = \alpha \Leftrightarrow \frac{1}{2} h_t + \varepsilon_t \tag{5.4.4a}$$

$$\sqrt{h_t} = \alpha_0 + \alpha_1^{(1)} |z_{t-1}| + \alpha_1^{(2)} \max(0, \Leftrightarrow z_{t-1}) + \beta_1 \sqrt{h_{t-1}} \tag{5.4.4b}$$

where $\alpha_0 > 0$, $\alpha_1^{(1)} \geq 0$, $\alpha_1^{(1)} + \alpha_1^{(2)} \geq 0$ and $\beta_1 \geq 0$, where z_t is Gaussian distributed with zero mean and variance 1 and where $\varepsilon_t = z_t \sqrt{h_t}$.

As in the case of TGARCH, this model can not be estimated by RPLR, only by Maximum Likelihood. The log-likelihood of the model is again as defined by (5.4.2) and (5.4.3).

Though the SSSH-GARCH model above may look very similar to the TGARCH model of (5.4.1), the two models are in fact very different due to the fact the $\sqrt{h_t}$ is driven by z_{t-1} in the SSSH-GARCH case and ε_{t-1} in the TGARCH case. Recall from Chapter 4 that the conditional variances are in the neighborhood of 10^{-4} such that eps_t has a mean of zero and a variance 10^{-4} where as z_t has a mean of zero and variance 1, 10000 times larger than ε_t. Because of this, the estimated values of $\alpha_1^{(1)}$ and $\alpha_1^{(2)}$ of SSSH-GARCH are expected to be a lot smaller that the corresponding values of TGARCH. This assumption is confirmed by Table 5.5 where the estimation results are presented.

	HP	Sony	Mobil	Pepsi
α	1.4144×10^{-3}	7.8771×10^{-4}	6.5215×10^{-4}	4.5712×10^{-4}
α_0	$\Leftrightarrow 1.6656 \times 10^{-4}$	$\Leftrightarrow 4.0912 \times 10^{-4}$	$\Leftrightarrow 2.1848 \times 10^{-5}$	5.5258×10^{-4}
$\alpha_1^{(1)}$	1.6292×10^{-4}	8.2776×10^{-4}	5.2024×10^{-4}	1.3976×10^{-3}
$\alpha_1^{(2)}$	6.6642×10^{-4}	1.3927×10^{-4}	3.2823×10^{-4}	8.7237×10^{-5}
β_1	9.8996×10^{-1}	9.8577×10^{-1}	9.5515×10^{-1}	8.9319×10^{-1}
$\sqrt{h_0}$	2.0239×10^{-2}	1.0019×10^{-2}	6.3095×10^{-4}	9.2947×10^{-3}

Table 5.5: Maximum Likelihood parameter estimates of SSSH-GARCH applied to return series.

Note that the parameter estimates of Hewlett-Packard, Sony and Mobil violates the constraint $\alpha_0 > 0$. This fact makes it theoretically possible that $\sqrt{h}$ may become non-positive, but investigations of the trajectories of $\sqrt{h_t}$ (not shown) reveals that this does not happen. The negative values of α_0 should not be taken as indications of incorrect parameter estimates, recall that the Scott, Stein & Stein and Heston model does not ensure positivity of σ_t.

Constraint maxization ($\alpha_0 > 0$) of the log-likelihood is not possible because the maximum of the constraint likelihood is on the boundary $\alpha = 0$.

Unfortunately, standard errors of the SSSH-GARCH parameter estimates are not available as it has not been possible to obtain a reliable of the Hessian.

Because the TGARCH and the SSSH-GARCH models have essentially the same likelihood function, the log likelihood of the two models are comparable. These are as displayed in Table 5.6.

As seen from the table, the two models fits equally well, expect perhaps for Sony where the SSSH-GARCH model provides the best description of the returns. This suggests that the returns of Sony are best modelled by the Scott, Stein & Stein and Heston model. While comparing the two models, it should be said that the param-

	HP	Sony	Mobil	Pepsi
TGARCH	3.111	3.099	3.874	3.510
SSSH-GARCH	3.111	3.501	3.874	3.510

Table 5.6: Values of the log-likelihood ($\times 10^{-3}$) of the TGARCH and SSSH-GARCH models applied to the four return series..

eters of the SSSH-GARCH model are much easier to estimate than the parameter of the TGARCH model. The numerical routine used to maximize the likelihood converges much faster in the SSSH-GARCH case.

The Scott, Stein & Stein and Heston parameter values corresponding to the SSSH-GARCH parameter estimates are presented in Table 5.7.

	HP	Sony	Mobil	Pepsi
α	1.4144×10^{-3}	7.8771×10^{-4}	6.5215×10^{-4}	4.5712×10^{-4}
κ	1.0035×10^{-2}	1.4232×10^{-2}	4.4854×10^{-2}	1.0681×10^{-1}
β	2.2849×10^{-2}	2.1564×10^{-2}	1.1687×10^{-2}	1.5940×10^{-2}
ρ	$\Leftrightarrow 4.9595 \times 10^{-4}$	$\Leftrightarrow 1.7781 \times 10^{-5}$	$\Leftrightarrow 1.2133 \times 10^{-4}$	$\Leftrightarrow 4.3745 \times 10^{-6}$
θ	4.4775×10^{-4}	5.4542×10^{-4}	4.4398×10^{-4}	8.6986×10^{-4}

Table 5.7: Parameter estimates of the Scott, Stein & Stein and Heston model computed from the SSSH-GARCH(1,1) parameter estimates.

Note that the estimates of α of Table 5.7 are close the the corresponding estimates in Table 5.4 as expected.

5.5   Summary

This chapter has illustrated an alternative way of estimating parameters of continuous time stochastic volatility models using estimated parameter values of discrete time Arch models. This is interesting as it is much easier to estimate parameters of discrete time models than parameters of bivariate continuous time stochastic volatility models.

Deducing the discrete time model corresponding to a given continuous time stochastic volatility model is only possible through a lot of elaborations and computations. Fortunately, the elaborations only have to be done once for each model.

This chapter has estimated the parameters of the Black, Scholes and Courtadon

model and Scott, Stein & Stein and Heston model using parameter estimates from discrete time Arch model, but it has not validated the obtained parameter estimates. Because of this, a formal evaluation of the estimation procedure is needed before it can be applied in practical applications. Such an evaluation is hereby suggested as an idea for a future research project.

Ideas for future research projects also include investigations of the diffusion limits of Arch models with non-Gaussian (e.g. t-distributed) innovations as well the deductions of the discrete time models corresponding to more of often used the continuous time stochastic volatility models.

Chapter 6

Conclusions

This final chapter provides a summary of the thesis, some concluding remarks about financial time series and the discrete time stochastic volatility models that has been investigated and suggestions for further research.

Returns of financial time series possess a number of well-documented characteristics different from those of Gaussian White Noise. The distribution of the returns is fat tailed, and the returns are linearly uncorrelated, though the returns are not independent as the squared returns are significantly correlated. Most important for the use of volatility models is that the returns exhibit changing variance, or volatility, over time, that is they are heteroskedastic.

The stochastic volatility models investigated in this thesis are developed to capture the characteristics of the return series. An in-depth analysis of the particular models known as GARCH and EGARCH have revealed that these models are perfectly able to capture the autocorrelations of the squared returns, but that the models need some adjustment before they are able to capture the heavy tails of the return series distributions.

The task of estimating the parameters of GARCH and EGARCH models is a difficult task and yet it has not received much attention in the literature of mathematical finance. In Chapter 4 an in-depth investigation of the properties of Maximum Likelihood estimation of GARCH and EGARCH models was presented, and an alternative recursive estimation procedure was proposed, analysed and compared to Maximum Likelihood. The analysis of Chapter 4 have established that the parameter estimates, regardless of estimation method, are influenced by a number of factors, factors whose influence it is not always possible to assess and control.

185

Traditionally, the models used for option pricing are continuous time stochastic volatility models, not discrete time models. The parameters of continuous time stochastic volatility models are, however, very difficult to estimate and because of this, a number of researchers have proposed to use estimates of discrete time model parameters in the process of estimating the parameters of continuous time stochastic volatility models. In Chapter 5, an approach where algebraic relationships between parameters of discrete and continuous time models are established by considering the discrete time models as discretized versions of the continuous time models was analysed. Continuous time models corresponding to the discrete time GARCH and EGARCH models were derived, and discrete time models corresponding to three well-known continuous time stochastic volatility models were deduced, a process that resulted in the discovery of a new GARCH model, a model that was named SSSH-GARCH.

For lists of ideas for further research, see the last pages of Chapter 4 and Chapter 5.

Appendix A

Software

As mentioned in the introduction, all estimations, simulations and computations in general have been performed using MatLab, and all needed computer programs not available as integrated parts of MatLab have been written by myself.

It would clearly be impossible to include all computer code in the thesis, but to give an idea of what has been implemented, the tables below provides an overview.

Function name	Purpose
ARX	Simulation of ARX process
ARMA	Simulation of ARMA process
tgsim	Simulation of TGARCH process
gsim	Simulation of GARCH process with zero mean
garch_state	Simulation of GARCH state-model with zero mean
gtsim	Simulation of GARCH process (t-distributed innovations) with zero mean
egsim	Simulation of EGARCH process with zero mean
bscgsim	Simulation of SSSH-GARCH process
eghsim	Simulation of EGARCH process with mean zero
Eulerbsc	Simulates Black, Scholes & Courtadon SDE using Euler scheme

Table A.1: Functions for simulation purposes.

187

Function name	Purpose
tdist	Calculates value of t-distribution with zero mean, variance sigma and n degrees of freedom in x
tford	Calculates value of cumulated t-distribution with zero mean, variance sigma and n degrees of freedom in x
trand	Returns random number drawn from a t-distribution with zero mean and variance sigma

Table A.2: Functions dealing with probability distributions.

Function name	Purpose
l	Computes value of pseudo-likelihood for GARCH series
lh	Computes value of pseudo-likelihood for GARCH series, h_0 included in estimation
liket	Computes value of pseudo-likelihood for GARCH series, t-distributed innovations
liketh	Computes value of pseudo-likelihood for GARCH series, t-distributed innovations, h_0 included in estimation
lsssh	Computes value of pseudo-likelihood for SSSH-GARCH series
ltrue_garch	Calculates true value of GARCH-likelihood
legarch	Computes value of pseudo-likelihood for EGARCH series
lhegarch	Computes value of pseudo-likelihood for EGARCH series, h_0 included in estimation
legarch2	Computes value of pseudo-likelihood for EGARCH series, only identifiable parameters
lhegarch2	Computes value of pseudo-likelihood for EGARCH series, h_0 included in estimation, only identifiable parameters

Table A.3: Likelihood functions.

Function name	Purpose
rplr_garch	Estimates GARCH(p,q) parameters using RPLR algorithm
rplr_garch_state	Estimates GARCH-state model parameters using
	RPLR algorithm
rplr_egarch	Estimates EGARCH(p,q) parameters using RPLR algorithm
rplr_garch_gain	Estimates GARCH(p,q) parameters using
	RPLR algorithm with gain
Autocorr	Estimates autocorrelation function of time series
Kernel	Non-parametric estimation of conditional variance
Kernelfunc	Gaussian kernel

Table A.4: Functions for estimation purposes.

Function name	Purpose
jarquebera	Calculates jarque-bera normality test
Ljung-Box	Calculates Ljung-Box test statistic
lmtest	Calculates test statistic of Lagrange Multiplier Test
	for heteroskedasticity

Table A.5: Test statistics

Function name	Purpose
dldthetahat	Approximates first derivative of pseudo log-likelihood
garchautocorr	Calculates autocorrelation of squared returns
	for a GARCH(p,q) model
Hessian	Approximates Hessian matrix
bscdisc	Discretised version of Black, Scholes, Courtadon model

Table A.6: Numerical approximations.

Function name	Purpose
garch_sqe	Calculates GARCH(p,q) squared errors
Spawn	Makes n copies of vector for use with RPLR
dldtheta	Computes the first derivative of pseudo log-likelihood
tgarch2bsc	Computes Black-Scholes-Courtadon model parameters from TGARCH parameters
sssh2sssh	Computes Scott, Stein & Stein and Heston model parameters from SSSH-GARCH parameters
m2tex	Converts MATLAB matrix to tex table
Testmin	Verified maximum of likelihood
bsc2tgarch	Computes TGARCH parameters from continuous time Black, Scholes and Courtadon parameters
Showmin	Plots likelihood in the proximity of the maximum
Showmin2	Plots likelihood in the proximity of the maximum, one dimension only
lambdafunc	Calculates forgetting factor and gain for observation at time t
arith	Transforms data series according to the definition of the arithmetic rate of return
geometric	Transforms data series according to the definition of the geometric rate of return
Garch_state_pars	Calculates dual GARCH-state and GARCH(2,2) parameters
G22tostate	Computes GARCH-state parameters from GARCH(2,2) parameters

Table A.7: Auxiliary functions.

Bibliography

Bera, A. K. & Higgins, M. L. (1993), 'A survey of arch models: Properties, estimation and testing', *Journal of Economic Surveys* **7**.

Berndt, E. K., Hall, B. H., Hall, R. E. & Hausman, J. A. (1974), 'Estimation inference in nonlinear strctural models', *Annals of Economic and Social Measurement* **4**, 653–665.

Bisgaard, P. (1998), Stochastic Modelling of Market Risk, Master's thesis, IMM, Technical University of Denmark, Lyngby.

Black, F. (1976), Studies of stock price volatility changes, *in* 'Proceedings of the 1976 Meeting of the American Statistical Association', pp. 177–181.

Black, F. & Scholes, M. (1973), 'The pricing of options and corporate liabilities', *Journal of Politicial Economy* **81**, 637–654.

Blair, B. J. (1995), Stochastic models in finance, Master's thesis, Victoria University of Wellington, United Kingdom.

Bollerslev, T. (1986), 'Generalized Autoregressive Conditional Heteroskedasticity', *Journal of Econometrics* **31**, 307–327.

Bollerslev, T. (1987), 'A conditionally heteroskedastic time series model for speculative prices and rates of return', *Review of Economics and Statistics* **69**, 542–547.

Bollerslev, T. (1988), 'On the correlation structure of the generalized autoregressive conditional heteroscedastic process', *Journal of Time Series Analysis* **9**(2), 12–31.

Bollerslev, T., Chou, R. Y. & Kroner, K. F. (1992), 'ARCH modelling in finance: A review of the theory and empirical evidence', *Journal of Econometrics* **52**, 5–59.

Breusch, T. S. & Pagan, A. R. (1979), 'A simple test for heteroscedasticity and random coefficient variation', *Journal of Econometrics* **47**, 1287–1294.

Burghardt, G. & Hanweck Jr., G. A. (1993), 'Calendar-adjusted volatilities', *Journal of Derivatives* .

Cao, C. Q. & Nelson, D. B. (1992), 'Inequality constraints in the univariate GARCH model', *Journal of Business and Economic Statistics* **10**, 229–235.

Conradsen, K. (1995), *En Introduktion til Statistik Bind 1, 6. udgave*, Institute of Mathematical Modelling, Lyngby, Denmark.

Davis, R. A. & Resnick, S. I. (1996), 'Limit theory for bilinear processes with heavy-tailed noise', *The Annals of Applied Probability* **6**, 1191–1210.

Drost, F. C. & Nijman, T. E. (1993), 'Temporal aggregation of GARCH processes', *Econometrica* **61**, 909–927.

Drost, F. C. & Werker, B. J. M. (1996), 'Closing the GARCH gap: Continuous time GARCH modeling', *Journal of Econometrics* **74**, 31–57.

Duan, J.-C. (1996), A unified theory of option pricing under stochastic volatility – from GARCH to diffusion. Working paper.

Duan, J.-C. (1997), 'Augmented GARCH(p,q) process and its diffusion limit', *Journal of Econometrics* **79**, 97–127.

Engle, R. F. (1982), 'Autoregressive conditional heteroscedasticity with estimates of the united kingdom inflation', *Econometrica* **50**(4), 987–1008.

Engle, R. F., Granger, C. W. J. & Kraft, D. F. (1984), 'Combining competing forecasts of inflation using a bivariate ARCH model', *Journal of Economic Dynamics and Control* **8**, 151–165.

Engle, R. F., Lilien, D. M. & Robins, R. P. (1987), 'Estimating time varying risk premia in the term structure: The arch-m model', *Econometrica* **55**(2), 391–407.

Engle, R. F. & Ng, V. K. (1993), 'Measuring and Testing the Impact of News on Volatility', *Journal of Finance* **48**(5), 1749–78.

Geweke, J. & Pantula, S. (1986), 'Modeling the persistence of conditional variances: a comment', *Econometric Review* **5**.

Glosten, L., Jagannathan, R. & Runkle, D. (1993), 'Relationship between the expected value and the volatility of the nominal excess return on stocks', *Journal of Finance* **48**, 1779–1801.

Gourieroux, C., Monfort, A. & Renault, E. (1993), 'Indirect inference', *Journal of Applied Econometrics* **8**, 85–118.

He, C. & Teräsvirta, T. (1999), 'Properties of moments of a family of GARCH processes', *Journal of Econometrics* **92**, 173–192.

Heston, S. (1993), 'A Closed-Form Solution for Options with Stochastic Volatility with Applications to Bond and Currency Options', *Review of Financial Studies* **6**, 327–343.

Hull, J. & White, A. (1987), 'The Pricing of Options on Assets with Stochastic Volatilities', *Journal of Finance* **42**(2), 281–300.

Jarque, C. M. & Bera, A. K. (1980), 'Efficient tests for normality, homoscedasticity and serial independence of regression residuals', *Economics Letters* **6**, 255–259.

Kloeden, P. E. & Platen, E. (1995), *Numerical Solutions of Stochastic Differential Equations, Second Edition*, Springer-Verlag, Heidelberg.

Kotz, S., ed. (1983), *Encyclopedia of Statistical Sciences*, Vol. 4, John Wiley & sons, pp. 456–458.

Ljung, L. (1987), *System Identification: Theory for the User*, Prentice-Hall, New York.

Ljung, L. & Söderström, T. (1983), *Theory and Practice of Recursive Estimation*, MIT Press, Cambridge, Massachusetts.

Madsen, H. (1995), *Tidsrækkeanalyse*, 2 edn, IMM, DTU, DK-2800 Lyngby.

Madsen, H. & Holst, J. (1996), *Modelling Non-linear and non-stationary time series*, IMM-DTU.

Madsen, H., Nielsen, J. N. & Baadsgaard, M. (1998), *Statistics in Finance - Lecture notes*, Institute of Mathematical Modelling, Lyngby.

Mandelbrot, B. (1963), 'The variation of certain speculative prices', *Journal of Business* **36**, 394–419.

McLeod, A. I. & Li, W. K. (1983), 'Diagnostic checking ARMA time series models using squared-residual autocorrelations', *Journal of Time Series Analysis* **4**(4), 269–273.

Nelson, D. B. (1991), 'Conditional heteroskedasticity in asset returns: A new approach', *Econometrica* **59**(2), 347–370.

Press, W. H., Teukolsky, S. A., Vetterling, W. T. & Flannery, B. P. (1992), *Numerical Recipes in C*, second edn, Cambrigde University Press, chapter 5, pp. 186–189.

Scott, L. O. (1987), 'Option Pricing when the Variance Changes Randomly: Theory, Estimation, and an Application', *Journal of Financial and Quantitative Analysis* **22**(4), 419–438.

Shephard, N. (1996), Statistical aspects of ARCH and stochastic volatility, *in* D. Cox, D. Hinkley & O. Barndorff-Nielsen, eds, 'Time Series Models: In econometrics, finance and other fields', Chapman & Hall, pp. 1–67.

Stein, E. M. & Stein, J. C. (1991), 'Stock price distributions with stochastic volatility – an analytic approach', *Review of Financial Studies* **4**, 727–752.

Taylor, S. J. (1986), *Modelling Financial Time Series*, Chicester.

Taylor, S. J. (1994), 'Modeling stochastic volatility: A review and comparative study', *Mathematical Finance* **4**(2), 183–204.

Zakoian, J.-M. (1994), 'Threshold heteroskedatstic models', *Journal of Economiic Dynamics and Control* **18**, 931–955.

Eksamensprojekter forår 2000 ved IMM.

1.	Ebbe Sørensen	Automated seamless mosaicking of retinal fundus images. (Ikke mangfoldiggjort).
2.	Torben Christiansen	Clustering positron emission tomography time series.
3.	Kim Spetzler Petersen	Signalseparation med uafhængig komponent analyse (ICA).
4.	Niels Wotetmann Andersen	Acoustic echoreduction.
5.	Anders Jensen	Slotted waveguides.
6.	Sverrir Gúdmundsson	Crustal deformation mapped by combined GPS and InSAR.
7.	Kåre Lynge Jensen	Image analysis on knitted textiles.
8.	Jesper Leedgaard Kim Herholdt Mortensen	Optimering af mandskabsplanægning i detail-branchen.
9.	Denis Fjeld Jakobsen	Automatisk differentiation til løsning af DAE.
10.	Per Hove Nielsen	Unit commitment i kraft-varme produktion.
11.	Erik Engström	Business integration strategies. - Analysis for B2B e-commerce. (Ikke mangfoldiggjort).
12.	Ágústa Haflidadóttir	Estimation of IPP parameters based on counts.
13.	Brian Kallehauge	Lagrange dualitet og ikke-differentiabel opti-mering. - Anvendt i rutelægning.
14.	Jacob Lemming	Multi-period portfolio management.
15.	Lars Jensen	Estimating functions for stochastic pro-cesses.
16.	Jon Kierkegaard	Estimation of nonlinear stochastic processes.