Minimum Distance Estimation of GARCH(1,1)-models

Submitted to
the Department of Statistics
of the University od Dortmund
In fulfilment of the Requirements for the Degree of
Doctor of Natural Sciences

by

Baudouin Tameze Azamo

Supervisor: Prof. Dr. Walter Krämer
Co-Supervisor: PD. Dr. Rafael Weißbach
Date of the Oral examination: October 19, 2007

Dortmund 2007

Contents

Preface

The following presents my research under the supervision of Prof Krämer. It has been a great experience to work with him. I particularly thank him for introducing me to this subject and suggesting challenging research questions. His patience, motivating comments and constructive criticism have strengthened this work and my academic education.

Other people have been a source of motivation and encouragement, providing helpful discussions and recommendations. I sincerely thank my fellow doctoral students at the Ruhr Graduate School in Economics, in particular Jan Brenner, Pavel Stoimenov and Christoph Hanck for their support in difficult periods. To my colleagues at the institute of Business and Economic Statistics and to our secretary Heide Aßhoff, I would like to express my gratitude. I am grateful to my teachers who, acknowledging my mathematical background have tried to give me the economic intuition behind different concepts and formulas presented in our course work.

I acknowledge the financial support of the Ruhr Graduate School in Economics (RGS Econ) through a generous research grant and the sponsorship to participate in various international conferences. It is the place here to thank the board of management and the coordinator of the RGS Econ for giving me the chance to participate in this program.

Finally, I thank my family and especially my mother for the sacrifices she made in keeping me out of the street and forcing me to get an education.

Chapter 1

Introduction

In applied econometrics, one is interested in quantifying how much one variable changes in response to a change in another variable. This is easily done in the ordinary least squared framework where it is assumed that the expected value of the squared error terms is the same at any given point. This is the constant variance assumption also called homoskedasticicy. Financial data however are known to be conditionally heteroskedastic. At some point in time the conditional variance is greater than at some other points in time. The family of ARCH models introduced by Engle (1982) and generalized into GARCH by Bollerslev (1986) are the main focus of this conditional heteroskedasticicy. They do not consider this conditional heteroskedasticity to be a problem as such, rather as a variance to be modeled. At the end, not only weaknesses of least squares are corrected but a forecast for the variance of the error term is computed.

In the particular case of financial times series (e.g daily data), a stylized fact is the so called "volatility clustering". Some periods are riskier than others and these risky times are not scattered randomly across the data. There is instead a degree of autocorrelations. Mandelbrot (1963) said " *...Large changes tend to* *follow large changes - of either sign - and small changes by small changes...".* In financial applications where the dependent variable is generally the return on an asset or portfolio, and the variance of the return represents the risk level of those returns, the prediction of the latter turns out to be of great interest. Investors require higher returns for holding riskier assets and option pricing theory uses empirically estimated volatilities in the Black-Scholes formula. It is therefore fundamental to carefully and properly model conditional heteroskedasticy for financial time series. The ARCH and GARCH models which stand for Autoregressive Conditional Heteroskecaticity and Generalized Autoregressive Conditional Heteroskedasticity are designed to take care of these problems.

Estimating ARCH and GARCH models is mainly done through (quasi)maximum likelihood techniques where a vector is built up of all model parameters and a likelihood function is constructed depending on this vector. An iterative search procedure is then used to find the parameters in the model that maximize the likelihood function. It is usually assumed that the conditional distribution of the returns is normal.

Previous research on financial market data have described the behavior of the autocorrelations of the squared and absolute returns series, see Dacorogna, Muller, R.J, Olsen, and Pictet (1993), Ding and Granger (1996) and Muller, Dacorogna, R.J, Olsen, and Pictet (1997). They have expressed the desire to construct a model that closely replicates the autocorrelations of the squared returns. Furthermore, on figure 1 in Jacquier, Polson, and Rossi (1994), there are some discrepancies between the autocorrelations of transformations of fitted returns from MLE and the autocorrelations of the actual fitted returns. Another reason for using this estimator is that the true data generating process can possess extreme nonnormality, in this case applying Maximum Likelihood methods do not always produce asymptotically efficient parameter estimates (see Baillie and Chung (2001), page 632).

As a distribution-free alternative, numerous authors have proposed minimum distance estimators of the GARCH parameters using either the autocorrelations or the autocovariances of the squared returns. In our particular case, the parameters of the GARCH(1,1)-model will be estimated from the autocorrelations of the squared process. This method applies a minimum distance estimator (MDE[1]) to the empirical autocorrelations of the GARCH squared process. In general, in cases where it is difficult to numerically estimate GARCH models from extreme non normal densities[2], the minimum distance estimator can be an interesting alternative. Furthermore, the MDE generates a model where the autocorrelations of the fitted squared values are closed to the population autocorrelations (see e.g. Baillie and Chung (2001), section 5).

In this work, we take this new alternative of distribution free estimation of GARCH(1,1)-models as given and study its implications in various issues previously addressed in the GARCH framework. Issues like small sample bias of estimated GARCH parameters, structural changes and market risk measurements have previously been addressed in the quasi maximum likelihood estimated GARCH models. We investigate these issues in the specific case in which the GARCH parameters are directly estimated from the autocorrelations of the squared process. This approach uses the minimum distance estimator which estimates the GARCH parameters by minimizing the Mahanalobis generalized distance of a vector of empirical autocorrelations from the corresponding population autocorrelations.

[1]will be defined later

[2] Soosung and Pereira (2006) discusses the convergence errors of the MLE estimators under the Bollerslev conditions defined in equation (4.2) .

In the following, after the introduction in chapter 1, chapter 2 addresses the issue of existence of solutions to the GARCH processes. Chapter 3 introduces the distribution free minimum distance estimation of the simple and often used GARCH(1,1)-model. Chapter 4 deals with the small sample bias of the estimated persistence in the GARCH(1,1)-model. Chapter 5 addresses the issue of lag choice in minimum distance estimation of a GARCH(1,1)-model. Chapter 6 looks at the structural change and the estimated persistence, particularly, it analyzes the effect a growing size deterministic structural breaks in the constant term of the conditional mean equation on the estimated persistence. In chapter 7, once again, the issue of structural breaks is addressed, in the context of a fixed break size and growing sample sizes. In the same chapter, we extend the investigation into the context of stochastic changes in the mean. We study these specific types of structural changes and its extensions to artificial long memory. In chapter 8, we look at applications of the GARCH estimates in risk management. The Value at Risk (VaR) and its coherent alternative, the expected tail loss (ETL) are described. We discuss their calculations in the minimum distance estimation framework. Chapter 9 concludes by recapping the main findings in this thesis and suggests further research questions.

Chapter 2

GARCH models

2.1 Introduction

Uncertainty as measured by risk or volatility is central in any type of financial analysis. In option pricing theory, the most determinant factor is the volatility associated with the price of the underlying asset. When calculating standard market risk measures such as the Value at Risk, we are mostly interested in the current levels of volatilities. We are namely assessing possible changes in the value of the portfolio over a very short period of time. In the process of valuing derivatives, a forecast of volatilities over the whole life of the derivatives is usually required.

One important stylized fact of financial returns series is that their conditional volatility changes over time. Figure 2.1 shows the price level and Figure 2.2 shows the innovations, conditional standard deviations and returns of the German Deutsche Bank stock and one easily realizes that its volatility is not constant over time. These concepts will be defined in the following section. In particular one observes in the second figure periods of large movements in

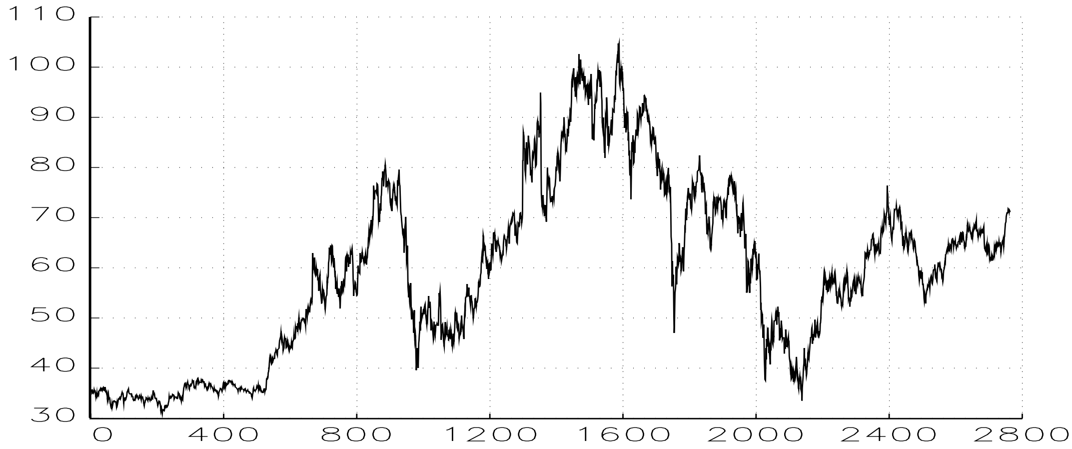


FIGURE 2.1: DEUTSCHE BANK STOCK PRICE FROM 01/01/95 TILL 04/08/2005 (2764 OBSERVATIONS)

prices alternating with periods in which prices hardly change. This is termed "volatility clustering". In the presence of this changing volatility, ARCH and GARCH are the most used tools in financial risk management. These models are so popular because not only do they account for volatility clustering but even more, they account for certain other characteristics such as pronounced excess kurtosis and fat-tailedness.

2.2 Assumptions

The logarithmic return r_t from an asset with price S_t at time t is defined as

$$\begin{aligned} r_t &= ln(S_{t+1}) - ln(S_t) \quad (2.1) \\ &= \mathbb{E}\big(r_t|\mathcal{F}_{t-1}\big) + \epsilon_t \quad \text{where } \epsilon_t \text{ is the error term} \\ &= \mu(b) + \epsilon_t, \quad t = 1, 2, 3, ..., N. \end{aligned}$$

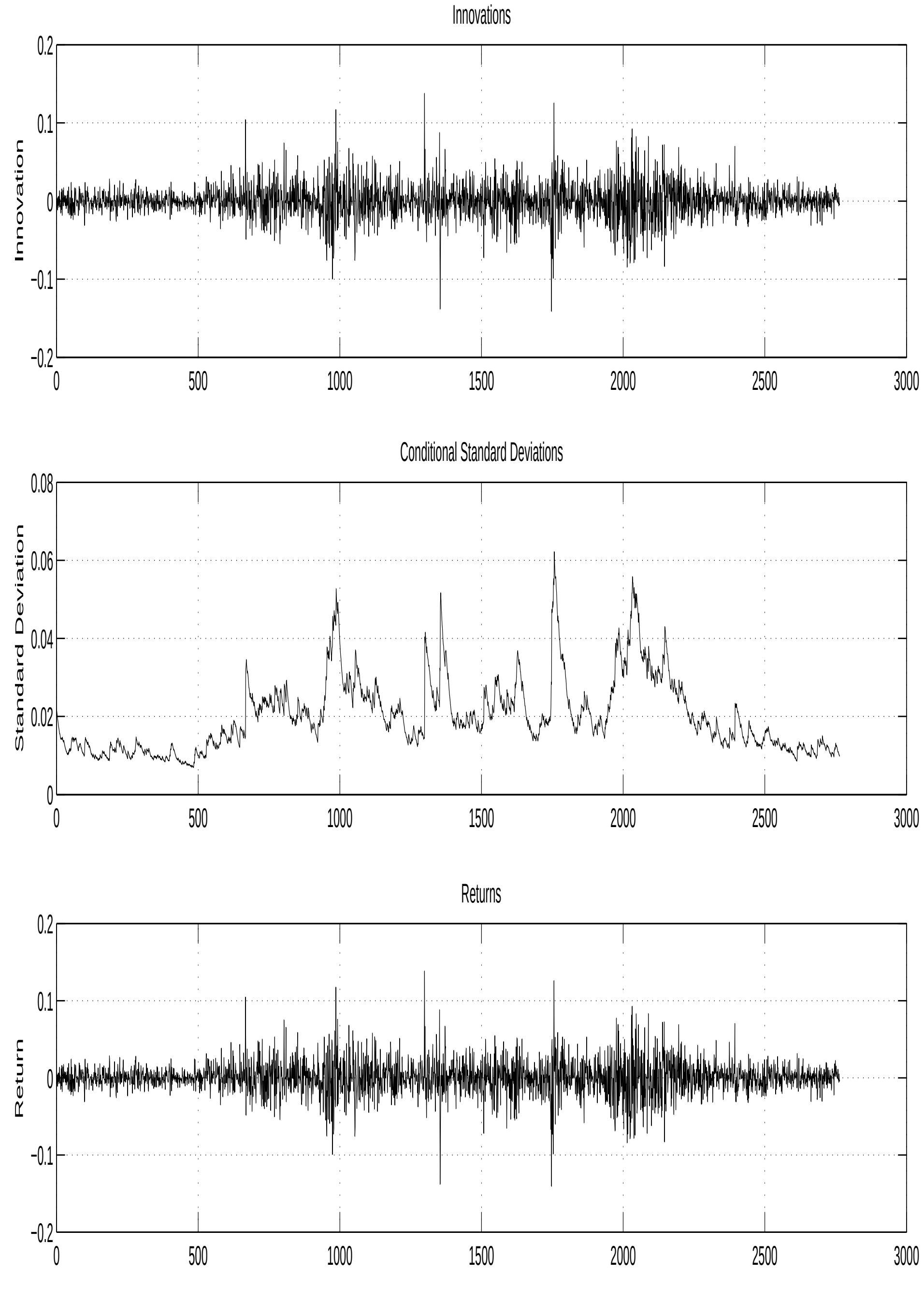


FIGURE 2.2: DEUTSCHE BANK STOCK RETURNS FROM 01/01/95 TILL 04/08/2005 (2764 OBSERVATIONS)

The σ-field $\mathcal{F}_t = \sigma\big(\epsilon_k, k \leq t\big)$ denotes the filtration modeling the information set. The GARCH model conditioned on such an information set is heteroskedastic (see e.g Greene (2003), page 241). μ is the conditional mean function with argument b, for example in a regression $\mu(b) = z_t'b$, where z_t denotes a set of independent variables. We shall consider this conditional mean function to be constant to ease our discussion, in particular on structural changes in chapter 6 and 7. The error term ϵ_t, also called disturbances or innovations, is the quantity of interest. Financial analysts and risk managers are mainly preoccupied by what makes it vary and how large it can be.

We model this error term as

$$\epsilon_t = \eta_t \sigma_t, \tag{2.2}$$

where η_t is a zero mean unit variance process.

The conditional variance equation σ_t is assumed to follow the difference equation

$$\sigma_t^2 = \omega + \sum_{i=1}^{q} \alpha_i \epsilon_{t-i}^2 + \sum_{i=1}^{p} \beta_i \sigma_{t-i}^2 \text{ with } \beta_i \geq 0 \text{ and } \alpha_j \geq 0. \tag{2.3}$$

This is the general GARCH(p,q) specification of Bollerslev (1986), which is the extension of the ARCH(q)-model of Engle (1982). It allows the conditional variance to depend not only on the past squared residuals but additionally on its own past realizations. The quantity $\delta = \sum_{i=1}^{q} \alpha_i + \sum_{i=1}^{p} \beta_i$ is defined as the persistence of this linear GARCH(p,q)-model. It is the most important parameter and the reason why we at all estimate GARCH model. Depending on its value for example, multiple periods forecasts of the volatility can be made. The following section discusses the existence of solutions of the GARCH process equation.

2.3 Existence of weakly stationary solutions of the linear GARCH process equation

This section treats weak stationarity of GARCH models. We start by defining the concept of weak stationarity.

Definition 1. *A stochastic process x_t is weakly stationary (or second order stationary or covariance stationary) if each x_t is squared integrable and if for all $t, m \in \mathbb{Z}, \mathbb{E}(x_t)$ and $cov(x_t, x_{t+m})$ are independent of t.*

Now we state the weak stationarity theorem for GARCH models:

Theorem 1. *(Bollerslev, 1986)*
When $\omega > 0$, the GARCH(p,q) model has a weakly stationary solution if and only if

$\sum_{i=1}^{p} \beta_i + \sum_{j=1}^{q} \alpha_j < 1$, *with* $\beta_i \geq 0$ *and* $\alpha_j \geq 0$.

Proof:

We start with

$$\epsilon_t = \eta_t \sigma_t \text{ with } \eta_t \sim i.i.d.\ \mathcal{N}(0,1). \tag{2.4}$$

Then subsequent substitution into the conditional variance (equation (2.3)) yields

$$\begin{aligned}
\sigma_t^2 &= \omega + \sum_{j=1}^{q} \alpha_j \eta_{t-j}^2 \sigma_{t-j}^2 + \sum_{i=1}^{p} \beta_i \sigma_{t-i}^2 \\
&= \omega + \sum_{j=1}^{q} \alpha_j \eta_{t-j}^2 \big(\omega + \sum_{i=1}^{q} \alpha_i \eta_{t-j-i}^2 \sigma_{t-j-i}^2 + \sum_{i=1}^{p} \beta_i \sigma_{t-i-j}^2\big) \\
&\quad + \sum_{j=1}^{p} \beta_j \big(\omega + \sum_{i=1}^{q} \alpha_i \eta_{t-j-i}^2 \sigma_{t-j-i}^2 + \sum_{i=1}^{p} \beta_i \sigma_{t-i-j}^2\big) \\
&= \ldots = \omega + \sum_{k=0}^{\infty} M(t,k).
\end{aligned}$$

where $M(t,k)$ contains all the terms of the form

$$\Pi_{i=1}^{q}{\alpha_i}^{a_i}\ \Pi_{j=1}^{p}{\beta_j}^{b_j}\ \Pi_{l=1}^{n}\eta_l^2 - S_l$$

for

$$\sum_{j=1}^{q} a_i + \sum_{i=1}^{p} b_j = k, \quad \sum_{j=1}^{q} a_i = n \tag{2.5}$$

and $S_l,\ \ 1 \leq l \leq n$ is a sequence of numbers satisfying

$$1 \leq S_1 < S_2 < ... < S_n \leq max(kq, (k-1)q + p).$$

So, it follows

$$M(t,0) = 1,$$

$$M(t,1) = \sum_{j=1}^{q} \alpha_j \eta_{t-j}^2 + \sum_{i=1}^{p} \beta_i,$$

$$M(t,2) = \sum_{j=1}^{q} \alpha_j \eta_{t-j}^2 \big(\sum_{j=1}^{q} \alpha_j \eta_{t-i-j}^2 + \sum_{i=1}^{p} \beta_i\big) + \sum_{i=1}^{p} \beta_i \big(\sum_{j=1}^{q} \alpha_j \eta_{t-i-j}^2 + \sum_{i=1}^{p} \beta_i\big)$$

and generally,

$$M(t,k+1) = \sum_{j=1}^{q} \alpha_j \eta_{t-j}^2 M(t-j,k) + \sum_{i=1}^{p} \beta_i M(t-i,k). \tag{2.6}$$

Since η_t^2 is $i.i.d.$, the moments of $M(t,k)$ do not depend on t, and in particular

$$\mathbb{E}(M(t,k)) = \mathbb{E}(M(s,k)) \text{ for all } \ k, s, t. \tag{2.7}$$

From (2.6) and (2.7), we can deduce

$$\begin{aligned}
\mathbb{E}(M(t,k+1)) &= \big(\sum_{j=1}^{q} \alpha_j + \sum_{i=1}^{p} \beta_i\big)\ \mathbb{E}M(t,k) \\
&\ . \\
&\ . \\
&\ . \\
&= \big(\sum_{j=1}^{q} \alpha_j + \sum_{i=1}^{p} \beta_i\big)^{k+1}\ \mathbb{E}M(t,0) \\
&= \big(\sum_{j=1}^{q} \alpha_j + \sum_{i=1}^{p} \beta_i\big)^{k+1}.
\end{aligned}$$

By combining the above results (in particular (2.4), (2.5) and (2.8)), we immediately obtain

$$\begin{aligned}\mathbb{E}(\epsilon_t^2) &= \omega + \mathbb{E}\big(\sum_{k=0}^{\infty} M(t,k)\big) \\ &= \omega + \sum_{k=0}^{\infty} \mathbb{E}\big(M(t,k)\big).\end{aligned}$$

The geometric series

$$\sum_{k=0}^{\infty} \mathbb{E}\big(M(t,k) \tag{2.8}$$

with

$$\mathbb{E}(M(t,k)) = \big(\sum_{j=1}^{q} \alpha_j + \sum_{i=1}^{p} \beta_i\big)^k$$

converges if and only if

$$\sum_{j=1}^{q} \alpha_j + \sum_{i=1}^{p} \beta_i < 1. \tag{2.8}$$

In this case then, we have

$$\mathbb{E}(\epsilon_t^2) = \frac{\omega}{1 - \big(\sum_{j=1}^{q} \alpha_j + \sum_{i=1}^{p} \beta_i\big)} \tag{2.9}$$

Indeed, weak stationarity in GARCH(p,q)-models is equivalent to

$$\sum_{j=1}^{q} \alpha_j + \sum_{i=1}^{p} \beta_i < 1 \tag{2.10}$$

As opposed to weak stationarity, strict stationarity is a stronger concept. We address the existence of strictly stationary solutions for the general GARCH process next.

2.4 Existence of strictly stationary solutions of the general GARCH process equation

In this section, we present necessary and sufficient conditions for the existence of a strictly stationary solution of the GARCH time series equation. For this, we recall some useful definitions. We will consider a probability measure λ on $\mathbb{R}$ with zero mean and unit variance.

Definition 2. *A process x_t is strictly stationary if for all $t, m \in \mathbb{Z}$ the law of $\big(x_t, x_{t+1}, ..., x_{t+m}\big)$ is independent of t.*

Definition 3. *The top Lyapounov exponent associated to a sequence $\mathbf{A_t}, t \in \mathbb{Z}$ of i.i.d. random matrices is given by*

$$\gamma = inf\big\{\mathbb{E}\big(\frac{1}{t+1}\ln\|\mathbf{A_0}\mathbf{A_{-1}}...\mathbf{A_{-t}}\|\big), t \in \mathbb{N}\big\} \tag{2.11}$$

$$when\ \mathbb{E}(max(\ln\|\mathbf{A_0}\|, 0) < \infty, \tag{2.12}$$

where $\|\ \|$ is a matrix norm on $M \in \mathbb{M}(d)$, the set of $d \times d$ matrices. $\|\ \|$ is defined as

$$\|M\| = sup\{\|Mx\|/\|x\|; x \in \mathbb{R}^d, x \neq 0\}. \tag{2.13}$$

In the previous section, we characterized the existence of weakly stationary solutions for the linear GARCH(p,q)-process. Bollerslev (1987) however, among others has found that some financial time series, especially daily data, show parameters which are not in the weak stationarity region. Even if these series are not squared integrable, they are strictly stationary.

For convenience, in the linear GARCH(p,q)-process, we will always suppose $p, q \geq 2$. We can always make the redundant α_j's (resp. β_i's) equal to zero when the corresponding p (resp. q) is smaller than two. With this we can define the following matrix:

$$\mathbf{A_t} = \begin{pmatrix} \beta_1 + \alpha_1 \frac{\epsilon_t}{\sigma_t} & \beta_2 & \dots & \beta_{p-2} & \beta_{p-1} & \beta_p & \alpha_2 & \alpha_3 & \dots & \alpha_{q-2} & \alpha_{q-1} & \alpha_q \\ 1 & 0 & \dots & 0 & 0 & 0 & 0 & 0 & \dots & 0 & 0 & 0 \\ 0 & 1 & \dots & 0 & 0 & 0 & 0 & 0 & \dots & 0 & 0 & 0 \\ 0 & 0 & \ddots & 1 & 0 & 0 & 0 & 0 & \dots & 0 & 0 & 0 \\ 0 & 0 & \dots & 0 & 1 & 0 & 0 & 0 & \dots & 0 & 0 & 0 \\ \frac{{\epsilon_t}^2}{{\sigma_t}^2} & 0 & \dots & 0 & 0 & 0 & 0 & 0 & \dots & 0 & 0 & 0 \\ 0 & 0 & \dots & 0 & 0 & 0 & 1 & 0 & \dots & 0 & 0 & 0 \\ 0 & 0 & \dots & 0 & 0 & 0 & 0 & 1 & \dots & 0 & 0 & 0 \\ 0 & 0 & \dots & 0 & 0 & 0 & 0 & 0 & \ddots & 1 & 0 & 0 \\ 0 & 0 & \dots & 0 & 0 & 0 & 0 & 0 & \dots & 0 & 1 & 0 \end{pmatrix}.$$

$\mathbf{A_t}$ contains two identity matrices of size $p-1$ and $p-2$ respectively. We can easily see that it is a $(p+q-1) \times (p+q-1)$ matrix. The random matrices are $i.i.d.$, and all coefficients of these matrices are integrable. So $\mathbb{E}(max(\ln \|\mathbf{A_0}\|, 0)$ is finite. The top Lyapounov exponent is therefore well defined. We have all the ingredients to present the following theorem, due to Bougerol and Picard (1992) that characterizes the existence of strictly stationary solutions of the general GARCH process equation.

Theorem 2. *(Bougerol and Picard, 1992)*
When $\omega > 0$, the GARCH(p,q) model has a strictly stationary solution if and only if the Lyapounov exponent γ associated with the matrices $\mathbf{A_t}, t \in \mathbb{Z}$ is strictly negative. Moreover, this solution is ergodic. Its is the only stationary solution when the η_t's are given.

The proof of this central theorem requires four steps. First we state the lemma ensuring the strict negativity of the top Lyapounov exponent given a sequence of $i.i.d.$ stochastic matrices. We then move on to the two steps require to prove any mathematical equivalence and conclude with the proof of the unicity of the solution.

Let us define the following multivariate model:

Definition 4. *A generalized autoregressive equation with non negative i.i.d. coefficients is*

$$x_{t+1} = \mathbf{A_{t+1}} x_t + \mathbf{B_t},\ t \in \mathbb{Z}, \tag{2.14}$$

where $\{(\mathbf{A_t}, \mathbf{B_t})\ t \in \mathbb{Z}\}$ *is a given sequence of independent, identically distributed, random variables with values in* $\mathbb{M}(d) \times \mathbb{R}^d$ *and* x_t *is in* $\mathbb{R}^d$.

There is tight connection between this multivariate model and the GARCH process. This connection plays an essential role in the proof of theorem 2. We start by noting this general fact:
Consider the process $x_t,\ t \in \mathbb{Z}$ defined by

$$x_t = (\sigma^2_{t+1}, \sigma^2_t, ..., \sigma^2_{t-p+2}, \epsilon^2_t, \epsilon^2_{t-1}, ..., \epsilon^2_{t-q+2})', \tag{2.15}$$

and define the $p + q - 1$ real vector B as

$$B = (\omega, 0, ..., 0)'. \tag{2.16}$$

It is easily seen that ϵ_t solves the GARCH process equation if and only if x_t is a solution of

$$x_{t+1} = A_{t+1} x_t + B. \tag{2.17}$$

The following lemma characterizes the negativity of the top Lyapounov exponent.

Lemma 1. *(Bougerol and Picard, 1992)*
Let $A_n, n \in \mathbb{Z}$ *be a sequence of independent, identically distributed, random matrices such that* $\mathbb{E}(max(\ln \|\mathbf{A_0}\|, 0) < \infty$.
If $lim_{t \longrightarrow \infty} \|\mathbf{A_0 A_{-1} ... A_{-t}}\| = 0$, *then the top Lyapounov exponent associated with this sequence is strictly negative.*

Proof : See e.g Bougerol (1987) or Bougerol and Picard (1990) for a complete proof and related results.

Here we are going to suppose that there exists a strictly stationary solution $\epsilon_t, t \in \mathbb{Z}$ of the GARCH process and prove that the top Lyapounov exponent associated with the matrices $\mathbf{A_t}$ is strictly negative. This constitutes the $\Rightarrow$ part of the equivalence.

Using (2.17), we can write for $t > 0$;

$$
\begin{aligned}
x_0 &= \mathbf{A_0}x_{-1} + \mathbf{B} && (2.18)\\
&= \mathbf{A_0A_{-1}}x_{-2} + \mathbf{B} + \mathbf{A_0B} && (2.19)\\
&= \mathbf{A_0A_{-1}A_{-2}}x_{-3} + \mathbf{B} + \mathbf{A_0}B + \mathbf{A_0A_{-1}B} && (2.20)\\
&= \mathbf{A_0A_{-1}A_{-2}}...\mathbf{A_{-t}}x_{-t-1} + \mathbf{B} + \sum_{k=0}^{t-1}\mathbf{A_0A_{-1}}...A_{-k}\mathbf{B}. && (2.21)
\end{aligned}
$$

The coefficients of $\mathbf{A_t}$, x_t and $\mathbf{B}$ are non negative. So for any $t > 0$, it holds

$$\sum_{k=0}^{t-1}\mathbf{A_0A_{-1}}...\mathbf{A_{-k}B} \leq x_0. \tag{2.22}$$

This shows that the series $\sum_{k=0}^{t-1}\mathbf{A_0A_{-1}}...\mathbf{A_{-k}B}$ converges almost surely. This never happens unless $\mathbf{A_0A_{-1}}...\mathbf{A_{-t}B}$ converges almost surely to zero as $t \longrightarrow \infty$.

Next, we prove that

$$lim_{t\longrightarrow\infty}\mathbf{A_0A_{-1}}...\mathbf{A_{-t}}e_l = 0, \tag{2.23}$$

for all $1 \leq l \leq p+q-1$.

Consider $(e_1, e_2, ..., e_{p+q-1})$ a canonical base of $\mathbb{R}^{p+q-1}$.

The case $l = 1$:

$B = \omega e_1$ and since $\omega \neq 0$ (ω is strictly positive),

$$lim_{t\longrightarrow\infty}\ \mathbf{A_0A_{-1}}...\mathbf{A_{-t}}e_1 = 0. \tag{2.24}$$

Since $\mathbf{A}_{-t}e_p = \beta_p e_1$, we have

$$lim_{t\longrightarrow\infty}\ \mathbf{A_0A_{-1}...A_{-t+1}A_{-t}}e_p \quad = \tag{2.25}$$

$$\beta_p\ lim_{t\longrightarrow\infty}\ \mathbf{A_0A_{-1}...A_{-t+1}}e_1 \quad = \quad 0. \tag{2.26}$$

Now we use induction (backward recursion) to show that (2.23) holds for all the $l \leq j$ given any j so that $2 < j \leq p$.
Suppose that for such a fixed j,

$$lim_{t\longrightarrow\infty}\ \mathbf{A_0A_{-1}...A_{-t+1}A_{-t}}e_j = 0. \tag{2.27}$$

Then using $\mathbf{A_{-t}}e_{j-1} = \beta_{j-1}e_1 + e_j$,

$$lim_{t\longrightarrow\infty}\ \mathbf{A_0A_{-1}...A_{-t+1}A_{-t}}e_{j-1} = \tag{2.28}$$

$$lim_{t\longrightarrow\infty}\ \mathbf{A_0A_{-1}...A_{-t+1}}(\beta_{j-1}e_1 + e_j) = \tag{2.29}$$

$$\beta_{j-1}\ lim_{t\longrightarrow\infty}\ \mathbf{A_0A_{-1}...A_{-t+1}}e_1 +\ lim_{t\longrightarrow\infty}\ \mathbf{A_0A_{-1}...A_{-t+1}}e_j = 0. \tag{2.30}$$

(2.28) shows that (2.23) holds as well for $j-1$. So (2.23) holds as well for all the $j \leq p$. Using the same reasoning and arguments as above for $e_{p+q-1}, ..., e_{p+1}$ and the relations

$$\mathbf{A_t}e_{p+q-1} = \alpha_q e_1 \quad \text{and} \quad \mathbf{A_t}e_{p+j-1} = \alpha_j e_1 + e_{p+j} \tag{2.31}$$

for $2 \leq j \leq q-1$, we conclude that (2.23) holds for all the e_l. So

$$lim_{t\longrightarrow\infty}\ \mathbf{A_0A_{-1}...A_{-t+1}A_{-t}} = 0\ a.s. \tag{2.32}$$

We now use Lemma 1 to conclude that the top Lyapounov exponent associated with the matrices $\mathbf{A_t}$ is strictly negative. This proves the $\Rightarrow$ part of the theorem.

We suppose now that the top Lyapounov exponent associated with the matrices $\mathbf{A_t}$ is strictly negative. We want to prove that the GARCH model then has a strictly stationary solution. This shall constitute the $\Longleftarrow$ part of the theorem. The sub-additive ergodic theorem proven for example in Kingman (1973) says that the top Lyapounov exponent almost surely can be rewritten as

$$\gamma = \lim_{t \longrightarrow \infty} = \frac{1}{t} \ln \|\mathbf{A_0 A_{-1}} ... A_{-t}\|. \tag{2.33}$$

Norms are equivalent in $\mathbb{R}^n$ so this Lyapounov exponent is independent of the norm. The Lyapounov exponent being strictly negative, (2.33) implies that the series

$$\sum_{k=0}^{\infty} \mathbf{A_t A_{t-1} ... A_{t-k} B} \tag{2.34}$$

converges almost surely for any t. Let construct the following $x_t,\ t \in \mathbb{Z}$ as

$$x_t = \mathbf{B} + \sum_{k=0}^{\infty} \mathbf{A_t A_{t-1} ... A_{t-k} B}. \tag{2.35}$$

This sequence is non negative and fulfills

$$x_{t+1} = \mathbf{A_{t+1}} x_t + \mathbf{B}. \tag{2.36}$$

Pose $\sigma_t = \sqrt{x^1_{t-1}}$ where x^1_{t-1} is the first component of the vector x_{t-1}.
The process $\epsilon_t = \sigma_t \eta_t$ is a solution of the GARCH model where η_t is any zero mean unit variance process whose distribution conditional on $\mathcal{F}_{t-1}$ follows the law λ. Then the process $\{\mathbf{A_t}, \eta_t,\ t \in \mathbb{Z}\}$ is strictly stationary and ergodic. For some measurable function F independent of t, we can write

$$\epsilon_t \quad = \quad F(\eta_t, \mathbf{A_n}, \mathbf{A_{n-1}}, \mathbf{A_{n-2}}, ...). \tag{2.37}$$

So $\{\epsilon_t,\ t \in \mathbb{Z}\}$ is a strictly stationary and ergodic process, solution of the GARCH process equation. This proves the $\Longleftarrow$ of the theorem.

In the last step of this proof, we are concerned about the unicity of this strictly stationary solution.

Let $z_t,\ t \in \mathbb{Z}$ be another solution of the GARCH process.
Then for $t \geq 0$, we have

$$\|x_0 - z_0\| = \|\mathbf{A_0 A_{-1} ... A_{-t}}(x_{-t-1} - z_{-t-1})\| \quad (2.38)$$

$$\leq \|\mathbf{A_0 A_{-1} ... A_{-t}}\| \ \|x_{-t-1} - z_{-t-1}\|. \quad (2.39)$$

$\|\mathbf{A_0 A_{-1} ... A_{-t}}\|$ converges to zero a.s., x_t and z_t are strictly stationary so is $x_{-t-1} - z_{-t-1}$. This means that $x_{-t-1} - z_{-t-1}$ is independent of t. This leads to $x_0 - z_0$ equals to zero in probability. By the same reasonning, one easilly obtains $x_t = z_t$ for $t > 0$ a.s. This shows that when the η_t are given, then the GARCH process has a unique solution.

2.5 General GARCH(1,1)-models

We shall restrict ourselves in this work to a very simple but very useful formulation of the conditional variance equation

$$\sigma_t^2 = \omega + \alpha\epsilon_{t-1}^2 + \beta\sigma_{t-1}^2. \quad (2.40)$$

Equations (2.2) and (2.40) form the GARCH(1,1)-model which is the most used tool in finance in the presence of conditional heteroskedaticity. α is the ARCH parameter, β is the GARCH parameter and $\delta = \alpha + \beta$ is the persistence parameter. As explained in Campbell, Lo, and MacKinlay (1997) page 483, the persistence parameter is important in constructing multi-period forecasts of volatility. When the persistence is smaller than 1, the unconditional variance of the GARCH process ϵ_t or equivalently the unconditional expectation of σ_t^2 is

$$\frac{\omega}{1 - (\alpha + \beta)} \ . \quad (2.41)$$

In fact, recursive substitution in (2.40) and the law of iterated expectation at time t, yield the following k-periods ahead conditional expected volatility[1]

$$\mathbb{E}(\sigma_{t+k}^2) = (\alpha+\beta)^k \big(\sigma_t^2 - \frac{\omega}{1-(\alpha+\beta)}\big) + \frac{\omega}{1-(\alpha+\beta)} \ . \tag{2.42}$$

So the multi-period volatility forecast reverts to its unconditional mean at rate $\alpha + \beta$. In the case $\alpha + \beta = 1$, the conditional expected volatility k periods ahead at time t, is

$$\mathbb{E}(\sigma_{t+k}^2) = \sigma_t^2 + k\omega \ . \tag{2.43}$$

In this specific case, the GARCH(1,1)-model has a unit autoregressive root so today's volatility affects forecasts of the volatility into the indefinite future.

The ARMA(1,1) representation of (2.2) and (2.40) is given by

$$\epsilon_t^2 = \omega + (\alpha+\beta)\epsilon_{t-1}^2 + u_t - \beta u_{t-1}, \tag{2.44}$$

where

$$u_t = \epsilon_t^2 - \mathbb{E}\big(\epsilon_t^2 | \epsilon_{t-1}^2, \epsilon_{t-2}^2, ...\big) = \epsilon_t^2 - \sigma_t^2 \tag{2.45}$$

is white noise uncorrelated with past ϵ^2s. In order to insure the positivity of σ_t^2 we require $\omega > 0$, $\alpha \geq 0$ and $\beta \geq 0$. In the literature, these are called "the Bollerslev non-negativity condition". We also require weak stationarity by imposing $\alpha + \beta < 1$.

In general one differentiates between linear and non-linear GARCH models. The GARCH(1,1)-model introduced is the simplest and standard formulation in the GARCH family of models. Other linear GARCH models include the Integrated GARCH (IGARCH), the Fractionally Integrated GARCH (to be presented in details in chapter 7) and the GARCH in mean model.

[1] As mentioned previously, the expected volatility is conditional on the filtration $\mathcal{F}_t$ modeling information set at time t.

The IGARCH is the GARCH model where the parameters α and β sum up to exactly 1. This fact implies a unit root in its ARMA(1,1)-formulation. (2.44) becomes

$$\epsilon_t^2 - \epsilon_{t-1}^2 = \omega + u_t - \beta u_{t-1}. \tag{2.46}$$

In this case, the unconditional variance of ϵ_t is not finite and the IGARCH is not covariance-stationary. Still, Nelson (1990) shows that it is strictly stationary. The autocorrelations of ϵ_t^2 in the case of the IGARCH model are not defined properly but Ding and Granger (1996) show that one can approximate autocorrelations at a given lag g by

$$\rho_g = \frac{1}{3}(1 + 2\alpha)(1 + 2\alpha^2)^{-\frac{g}{2}}. \tag{2.47}$$

The GARCH in mean model introduced by Engle, Lilien, and Robins (1987) was designed to capture the relationship between the return and the time varying conditional variance. It considers the $\mu(b)$ in (2.1) as $\kappa\sigma_t^2$ such that the return is written as

$$x_t = \kappa\sigma_t^2 + \epsilon_t. \tag{2.48}$$

The conditional variance equation stays the same. Under this formulation, the autocorrelation function of the GARCH(1,1)-M at a given lag g as derived in Hong (1991) is

$$\rho_g = (\alpha + \beta)\rho_{g-1} \tag{2.49}$$

$$= (\alpha + \beta)^{g-1}\rho_1, \tag{2.50}$$

where

$$\rho_1 = \frac{(\alpha + \beta)(2\alpha^2\kappa^2\omega)}{(2\alpha^2\kappa^2\omega) + (1 - \alpha - \beta)(1 - \beta^2 - 2\alpha\beta - 3\alpha^2)}. \tag{2.51}$$

Bollerslev, Chou, and Kroner (1992) offer an overview of the applications of GARCH in mean models to stock returns, interest rates and foreign exchange rates.

Among the rich class of nonlinear GARCH models, exponential GARCH, GJR-GARCH, quadratic GARCH and Markov-switching GARCH are used in modern finance (see Franses and Dijk (2000), Hentschel (1995) and Bollerslev, Chou, and Kroner (1992) among others). The conditional variance in standard GARCH models depends on the square of the shock, so positive and negative shocks of the same magnitude will have the same effect. Volatile periods in stock markets are initiated by large negative shocks. When the stock falls, the debt-to-equity ratio(also called leverage) increases leading to an increase of the volatility. Positive and negative shocks have different impact on the conditional volatility of the following observations as recognized by Black (1976). This property is called the leverage effect. These nonlinear GARCH models have been designed to account for the effects of positive and negative shocks or other types of asymmetries.

The exponential GARCH (EGARCH) model of Nelson (1991) was the first variant of GARCH models to address the issues of asymmetries. In this model the conditional variance is

$$\ln(\sigma_t^2) = \omega + \alpha\eta_{t-1} + \gamma(\mid \eta_{t-1} \mid - \mathbb{E}(\mid \eta_{t-1} \mid)) + \beta \ln(\sigma_t^2). \tag{2.52}$$

As seen, the EGARCH differs from the standard GARCH models in being formulated in terms of the log of the conditional variance. Nonpositive variances need not be prevented and this simplifies the estimation. As wanted, negative shocks have a different impact than positive.

The GJR-GARCH of Glosten, Jagannathan, and Runkle (1993) is an alternative method that accounts for asymmetries. In this particular model, the coefficients of ϵ_t^2 depends on the sign of the shock. The conditional variance is written as

$$\sigma_t^2 = \omega + \alpha\epsilon_{t-1}^2(1 - \mathbb{I}[\epsilon_{t-1} > 0]) + \gamma\epsilon_{t-1}^2\mathbb{I}[\epsilon_{t-1} > 0] + \beta\sigma_{t-1}^2 \ , \tag{2.53}$$

where $\mathbb{I}$ is the indicator function.

The condition for the nonnegativity of the conditional variance are

$$\omega > 0, \quad \alpha + \gamma/2 \geq 0 \text{ and } \quad \beta > 0. \tag{2.54}$$

The condition for covariance stationarity is

$$\alpha + \gamma/2 + \beta \leq 1. \tag{2.55}$$

GJR-GARCH and EGARCH can be considered as alternative models for asymmetries in the same series. It is however difficult to develop a criteria to distinguish between them (see e.g Franses and Dijk (2000) page 151).

The quadratic GARCH (QGARCH) is another GARCH that takes are of the asymmetries, see Sentana (1995). Its conditional variance is written as

$$\sigma_t^2 = \omega + \gamma\epsilon_{t-1} + \alpha\epsilon_{t-1}^2 + \beta\sigma_{t-1}^2. \tag{2.56}$$

The term $\gamma\epsilon_{t-1}$ enables positive and negative shocks to have different effects on σ_t^2. If $\gamma < 0$, the effect positive shocks will be smaller than the effect of negative shocks of the same magnitude. Apart from the asymmetry, QGARCH and GARCH are very similar. They have the same unconditional variance, and the condition for covariance stationarity and existence of the unconditional fourth moment are the same. The kurtosis as expected is different because QGARCH is built up to account for it whereas standard GARCH does not.

In these nonlinear models so far, the parameter in the model change with respect to the sign and the size of the lagged shock, which is observable. Instead of letting the model parameter change according to the sign of these observable shocks, we can assume that parameters change according to an unobservable Markov process s_t. A general Markov Switching model has its conditional variance equation written as

$$\sigma_t^2 = \left[\omega + \alpha\epsilon_{t-1}^2 + \beta\sigma_{t-1}^2\right]\mathbb{I}[s_t = 1] + \left[\zeta + \gamma\epsilon_{t-1}^2 + \iota\sigma^2{}_{,t-1}\right]\mathbb{I}[s_t = 2] \tag{2.57}$$

where s_t is a two state Markov chain with given transition probabilities and $\mathbb{I}$ is the indicator function. This form is considered in Klaassen (2002). Dueker (1997), Cai (1989) and Hamilton and Susmel (1994) among others have applied different forms of the Markov Switching GARCH.

Existence of higher order moments in the GARCH(1,1)-model

This section studies the necessary and sufficient conditions for the existence of higher order moments of the GARCH(1,1)-model. It gives as well a close formula for its computation.

Theorem 3. *(Bollerslev, 1986)*

For the GARCH(1,1)-model as previously defined, a necessary and sufficient condition for existence of the 2mth moment is

$$\sum_{j=0}^{m} \binom{m}{j} a_j \ \alpha^j \ \beta^{m-j} < 1, \tag{2.58}$$

where

$$a_0 = 1, \ a_j = \Pi_{i=1}^{j}(2j-1), \ j = 1, ..., m \tag{2.59}$$

The 2mth moment can be expressed by the recursive formula

$$\mathbb{E}(\epsilon_t^{2m}) = a_m \big[\sum_{n=0}^{m-1} a_n^{-1} \mathbb{E}(\epsilon_t^{2n}) \alpha^{m-n} \binom{m}{m-n} \sum_{j=0}^{n} \binom{n}{j} a_j \ \alpha^j \ \beta^{n-j} \big] \tag{2.60}$$

$$\times \ \big[1 - \sum_{j=0}^{m} \binom{m}{j} a_j \ \alpha^j \ \beta^{m-j} \big]^{-1}. \tag{2.61}$$

Proof: By normality,

$$\mathbb{E}(\epsilon_t^{2m}) = a_m \ \mathbb{E}(\sigma_t^{2m}), \tag{2.62}$$

where a_m is defined as in the theorem. Using the binomial formula, we can write

$$\sigma_t^{2m} = (\omega + \alpha\epsilon_{t-1}^2 + \beta\sigma_{t-1}^2)^m \tag{2.63}$$

$$= \sum_{n=0}^{m} \binom{m}{n} \omega^{m-n} \sum_{j=0}^{n} \binom{n}{j} \alpha\ \beta^{n-j}\ \epsilon_{t-1}^{2j} \sigma_{t-1}^{2(n-j)}. \tag{2.64}$$

Now we use the equality

$$\mathbb{E}(\epsilon_{t-1}^{2j}\sigma_{t-1}^{2(n-j)}|\mathcal{F}_{t-2}) = a_j \sigma_{t-1}^{2n} \tag{2.65}$$

to obtain

$$\mathbb{E}(\sigma_t^{2m}|\mathcal{F}_{t-2}) = \sum_{n=0}^{m} \sigma_{t-1}^{2n} \binom{m}{n} \omega^{m-n} \sum_{j=0}^{n} \binom{n}{j}\ a_j\ \alpha^j\ \beta^{n-j}. \tag{2.66}$$

Let's define $z_t = (\sigma_t^{2m}, \sigma_t^{2(m-1)}, ..., \sigma_t^2)$. Then by (2.66),

$$\mathbb{E}(z_t|\mathcal{F}_{t-2}) = d + \mathbb{C}z_{t-1}, \tag{2.67}$$

where $\mathbb{C}$ is an $m \times m$ matrix with diagonal elements

$$\sum_{j=0}^{i} \binom{i}{j}\ a_j\ \alpha^j\ \beta^{i-j} \quad \text{for } i = 1, 2, ..., m. \tag{2.68}$$

Replacing in (2.67) yields

$$\mathbb{E}(z_t|\mathcal{F}_{t-k-1}) = (\mathbb{I} + \mathbb{C} + \mathbb{C}^2 + ... + \mathbb{C}^{k-1})d + \mathbb{C}^k z_{t-k}\ . \tag{2.69}$$

We assume that the process has started far away in the past with finite $2m$ moments, the limit as k goes to infinity exists and does not depend on t if and only if all the eigenvalues of $\mathbb{C}$ lie inside the unit circle,

$$\mathbb{E}(z_t|\mathcal{F}_{t-k-1}) = (\mathbb{I} + \mathbb{C})^{-1}d = \mathbb{E}(z_t). \tag{2.70}$$

Because $\mathbb{C}$ is upper triangular, the eigenvalues are equal to the diagonal elements as given in (2.68). Intense and straightforward calculations (see e.g

Bollerslev (1986) page 325 and onward) show that $\sum_{j=0}^{i} \binom{i}{j} a_j \alpha^j \beta^{i-j} < 1$ implies $\sum_{j=0}^{i-1} \binom{i-1}{j} a_j \alpha^j \beta^{i-1-j} < 1$ for $\alpha+\beta \leq 1$ and $\sum_{j=0}^{m} \binom{m}{j} a_j \alpha^j \beta^{m-j} < 1$ is enough for the $2mth$ moment to exist.

Finally we rearrange (2.62) and (2.66) to obtain

$$\begin{aligned} \mathbb{E}(\epsilon_t^{2m}) &= a_m \big[\sum_{n=0}^{m-1} a_n^{-1} \mathbb{E}(\epsilon_t^{2n}) \alpha^{m-n} \binom{m-n}{m} \sum_{j=0}^{n} \binom{n}{j} a_j \alpha^j \beta^{n-j} \big] \quad (2.71) \\ &\times \big[1 - \sum_{j=0}^{m} \binom{m}{j} a_j \alpha^j \beta^{m-j} \big]^{-1}. \quad (2.72) \end{aligned}$$

According to this theorem, the fourth order moment exists if and only if $3\alpha^2 + 2\alpha\beta + \beta^2 < 1$. Furthermore,

$$\mathbb{E}(\epsilon_t^2) = \frac{\omega}{1-(\alpha+\beta)} , \quad (2.73)$$

and

$$\mathbb{E}(\epsilon_t^4) = \frac{3\alpha^2(1+\alpha+\beta)}{(1-\alpha-\beta)(1-3\alpha^2-2\alpha\beta-\beta^2)} . \quad (2.74)$$

We easily calculate the coefficient of kurtosis as

$$\begin{aligned} \kappa &= \frac{\mathbb{E}(\epsilon_t^4) - 3\mathbb{E}(\epsilon_t^2)^2}{\mathbb{E}(\epsilon_t^2)^2} \quad (2.75) \\ &= \frac{6\alpha^2}{1-\beta^2-2\alpha\beta-3\alpha^2} \quad (2.76) \end{aligned}$$

and this quantity is greater than zero by assumption. So the GARCH(1,1)-process is heavy tailed.

2.6 Invariance of the estimated MLE GARCH parameters to the choice of $\mu(b)$ and ω

We consider the GARCH(1,1) - model:

$$r_t = \epsilon_t + \mu(b) \ , \tag{2.77}$$

$$\epsilon_t = \eta_t \sigma_t \ , \tag{2.78}$$

$$\sigma_t^2 = \omega + \alpha \epsilon_{t-1}^2 + \beta \sigma_{t-1}^2 . \tag{2.79}$$

In any estimation of GARCH(1,1)-models, the important parameters are the ARCH and GARCH parameters. The following three results[2] show that scaling the data by a constant, changing $\mu(b)$ or altering ω do not change the estimated ARCH parameter $\hat{\alpha}$ and estimated GARCH parameter $\hat{\beta}$. We stress here that this issue is totally different from the structural changes one where the change of the parameters occur within the time serie, inducing an increase of the estimated persistence (see chapter 6 and 7). In general, a change of any of the above three parameters might produce a shift in the whole time series but not in different blocks of the time series.

Result 1. *Multiplying the GARCH data by a constant k will not change the value of $\hat{\alpha}$ and $\hat{\beta}$ but it will scale the value of $\mu(b)$ by k and the value of ω by k_0^2.*

Let's define $r'_t = k r_t$. Then from (2.77) and (2.79), it comes

$$r'_t = k(\epsilon_t + \mu(b)) \ , \tag{2.80}$$

$$\epsilon'_t = k \eta_t \sigma_t \ , \tag{2.81}$$

$$\sigma_t'^2 = k^2 \sigma_t^2 \tag{2.82}$$

$$= k^2 \omega + \alpha k^2 \epsilon_{t-1}^2 + \beta k^2 \sigma_{t-1}^2 \tag{2.83}$$

$$= \omega' + \alpha \epsilon_{t-1}'^2 + \beta \sigma_{t-1}'^2 . \tag{2.84}$$

(2.85)

[2]We follow in this section Lumsdaine (1995).

The scaled model is of a form similar to (2.77) and (2.79) where $\mu'(b) = k\mu(b)$ and $\omega' = k^2\omega$ and α and β remain unchanged. The maximum likelihood of the scaled model will be constructed as previously. The estimated conditional variances and the maximum likelihood estimates will be scaled in the same manner as the true parameters. The value of the likelihood function will decrease by $ln(k)$ but this is only a constant and does not affect the ranking of the function. Therefore the estimated α and β will remain unchanged.

Result 2. *If we change $\mu(b)$, this will not change the value of the values of $\hat{\alpha}$ and $\hat{\beta}$.*

Let $\mu(b)$ suffer a shift k. We called

$$\mu'(b) = \mu(b) + k. \tag{2.86}$$

A new process r'_t will be built as

$$r'_t = \mu'(b) + \epsilon_t. \tag{2.87}$$

This amount to a shift in the distribution of r_t but not on its shape. The estimated $\mu(b)$ will change by a shift of magnitude k,

$$\mu'(b) - \mu(b) = \hat{\mu}'(b) - \hat{\mu}(b). \tag{2.88}$$

The quantities ϵ_t and σ_t do not change, so the estimated likelihood conditional on the data being generated with this new parameter value will not change either so this change does not alter the estimated ARCH and GARCH parameters.

Result 3. *If we change the value of ω, this will not change the value of $\hat{\alpha}$ and $\hat{\beta}$. It will result in different values for the estimated likelihood, $\hat{\mu}(b) - \mu$, and $\hat{\omega} - \omega$. In particular*[3], doubling ω will decrease the likelihood by $\frac{\ln 2}{2}$, will multiply $\hat{\mu}(b) - \mu$ by $\sqrt{2}$ and will double $\hat{\omega} - \omega$.

[3] A detailed proof of these results can be found in Lumsdaine (1995) page 9 and 10.

2.7 Quasi maximum likelihood estimation of GARCH(1,1)-models

We have assumed so far that the returns were conditionally normal and based our maximum likelihood function on this important assumption. However, one can never be sure that the specified distribution is the correct one and there is more and more evidence in the literature that returns are not conditionally normally distributed, see Rachev, C, and Fabozzi (2005), Embrechts, McNeil, and Frey (2005) among others. The practical approach is to ignore this problem and still base the likelihood on the normal distribution assumption. This is usually referred to in the literature as quasi maximum likelihood[4]. Table 2.1 shows the quasi maximum likelihood computation and GARCH estimates for the Deutsche bank returns considered in figure 2.2. The best parameter are the one that maximizes the likelihood function for N given observations

$$\prod_{t}^{N}[\frac{1}{\sqrt{2\pi\sigma_t^2}}exp(\frac{r_t^2}{2\sigma_t^2})]. \tag{2.89}$$

In general GARCH models (see e.g. Franses and Dijk (2000)), this still yields consistent and asymptotically normal estimates, provided that the models for the conditional mean and the conditional variance are correctly specified. In the specific case of the GARCH(1,1)-model, Lumsdaine (1996) and Lee and Hansen (1991) proved the consistency and asymptotic normality of the quasi maximum likelihood estimator. The results of Lumsdaine (1996) hold not only for the ARCH parameter α and the GARCH parameter β, but also for the other parameters μ and ω. Another finding in that research article is that in contrast to the case of the unit root in the conditional mean, the presence of unit root in conditional variance does not affect the limiting distribution of the estimators when the returns are normally distributed. Even more interesting, consistency

[4]To be more precise, this is Gaussian quasi maximum likelihood as opposed to non Gaussian QMLE such as the student t or chi square distribution of the returns.

Day t	r_t	σ_t^2	$-2ln(\sigma_t) - r_t^2/\sigma_t^2$
1			
2	-0.007869	0.0000619	——
3	0.003661	0.00000638	9.44866
4	0.001124	0.00000620	9.66783
⋮	⋮	⋮	⋮
2761	-0.005750	0.0001100	8.814167
2762	0.004771	0.0001061	8.936126
2763	-0.004349	0.0001018	9.006677
			19396.302

Table 2.1: QMLE of the GARCH(1,1) model with data from figure 2.2. Final estimates $\omega = 0.0000019$, $\alpha = 0.074$ and $\beta = 0.923$ for a persistence of $\delta = 0.997$ (2767 observations).

and asymptotic normality of the quasi maximum likelihood estimates do not require that the parameters of the GARCH(1,1)-model satisfy the covariance stationarity condition $\alpha + \beta < 1$.

2.8 Conclusion

In this chapter, we have introduced GARCH models in general, and have given following Bougerol and Picard (1992), necessary and sufficient conditions for the existence of a strictly stationary solution. Bollerslev (1986) discusses the existence of weakly stationary solutions. Nelson (1990) solved the problem of the existence of strictly stationary solution for the GARCH(1,1)-model and indicated that the theory of products of random matrices should be the appropriate technique to handle the general case. Bougerol and Picard (1992) followed his advice to solve the problem for general GARCH models. They characterized the existence of the solution by the strict negativity of the Lyapounov exponent associated to the GARCH corresponding random matrices.

In the following, we consider the simple but still very useful linear GARCH(1,1)-model. We will assume at least weak stationarity and ergodicity. The results obtained in the case of the GARCH(1,1)-model are fairly easily generalized to linear GARCH(p,q)-models and provide insights for nonlinear models.

Chapter 3

Minimum Distance Estimation of GARCH(1,1)-models

3.1 Introduction

The Minimum Distance Estimation (MDE) of a GARCH(1,1)-model to be introduced in this section minimizes the Mahalanobis generalized distance of a vector of empirical autocorrelations from the corresponding population autocorrelations. The attraction of this estimator compared to maximum likelihood estimator is that it does not require any strong distributional assumption on the disturbances of the process. The MDE as such is very similar to the generalized method of moment (GMM) estimation in the sense that, as we will see in the next pages, it is obtained by the minimization of a quadratic criterion function.

3.2 The autocorrelation function of the GARCH(1,1) squared process

We consider a covariance-stationary GARCH(1,1)-model.The long run or unconditional variance of ϵ_t as said previously is

$$\sigma^2 = \frac{\omega}{1-\alpha-\beta} \ . \tag{3.1}$$

We therefore can rewrite the conditional variance equation as

$$\sigma_t^2 = \sigma^2(1-\alpha-\beta) + \alpha\epsilon_{t-1}^2 - \beta\sigma_{t-1}^2. \tag{3.2}$$

We rearrange the above equation to obtain

$$\epsilon_t^2 - \sigma^2 = (\alpha+\beta)(\epsilon_t^2 - \sigma^2) - \beta\sigma_{t-1}^2(\eta_{t-1}^2 - 1) + \sigma_t^2(\eta_t^2 - 1). \tag{3.3}$$

By multiplying both sides of (3.3) by $(\epsilon_{t-1}^2 - \sigma^2)$ and taking the expectations, we obtain

$$\mathbb{E}(\epsilon_t^2 - \sigma^2)(\epsilon_{t-1}^2 - \sigma^2) = (\alpha+\beta)\mathbb{E}(\epsilon_t^2 - \sigma^2)^2 - 2\beta\mathbb{E}\sigma_{t-1}^4. \tag{3.4}$$

We easily identify

$$\gamma_1 = \mathbb{E}(\epsilon_t^2 - \sigma^2)(\epsilon_{t-1}^2 - \sigma^2) \tag{3.5}$$

as the covariance between ϵ_t^2 and ϵ_{t-1}^2 and

$$\gamma_0 = \mathbb{E}(\epsilon_{t-1}^2 - \sigma^2)^2 \tag{3.6}$$

as the variance of ϵ_{t-1}^2. We further ensure that the fourth moment of the ϵ_t exists by imposing that[1]

$$3\alpha^2 + 2\alpha\beta + \beta^2 < 1. \tag{3.7}$$

Then we can divide both sides of (3.4) by γ_0 and obtain[2]

$$\rho_1 = \frac{\gamma_1}{\gamma_0} = \alpha + \beta - 2\beta\frac{\mathbb{E}\sigma_{t-1}^4}{\gamma_0}. \tag{3.8}$$

[1]This is a result already seen in Chap 2 as an application of Theorem 3.
[2]γ_0 is finite.

By definition we have

$$\gamma_0 = \mathbb{E}(\epsilon_t^2 - \sigma^2)^2 = 3\mathbb{E}\sigma_t^4 - \sigma^4, \tag{3.9}$$

which yields

$$\mathbb{E}\sigma_t^4 = \frac{\gamma_0 + \sigma^4}{3} \ . \tag{3.10}$$

Substituting (3.10) into (3.8) gives

$$\rho_1 = \alpha + \beta - 2\beta\frac{\gamma_0 + \sigma^4}{3} = \alpha + \frac{1}{3}\beta - \frac{2}{3}\beta\frac{\sigma^4}{\gamma_0} \ . \tag{3.11}$$

From the conditional variance equation we derive

$$\mathbb{E}\sigma_{t-1}^4 = \frac{\sigma^4(1-\alpha-\beta)(1+\alpha+\beta)}{1-(3\alpha^2+2\alpha\beta+\beta^2)} \ , \tag{3.12}$$

and substituting this into (3.9), we have

$$\gamma_0 = 3\mathbb{E}\sigma_t^4 - \sigma^4 = \sigma^4\frac{2(1-2\alpha\beta-\beta^2)}{1-(3\alpha^2+2\alpha\beta+\beta^2)} \ , \tag{3.13}$$

so that

$$\frac{\sigma^4}{\gamma_0} = \frac{1-(3\alpha^2+2\alpha\beta+\beta^2)}{2(1-2\alpha\beta-\beta^2)} \ . \tag{3.14}$$

Substituting this into (3.11) and simplifying further gives

$$\rho_1 = \alpha + \frac{\alpha^2\beta}{1-2\alpha\beta-\beta^2} \ . \tag{3.15}$$

Bollerslev (1988) discusses the correlation structure for the GARCH-model and concludes that the autocorrelations functions for ϵ_t^2 for GARCH(p,q) is given by

$$\rho_g = \Sigma_{i=1}^m \phi_i \rho_{g-i} \quad g \geq q+1, \tag{3.16}$$

where $m = max(p,q)$ and $\phi_i = \alpha_i + \beta_i$ for $i = 1, 2, ..., m$ $\alpha_i = 0$ for $i > p$, and $\beta_i = 0$ for $i > q$.

For $p = q = 1$,

$$\rho_g = \phi_1\rho_{g-1} = (\alpha+\beta)\rho_{g-1} \ \text{ for } \quad g \geq 2. \tag{3.17}$$

Combining (3.15) and (3.17) gives the autocorrelation[3] functions of the GARCH(1,1) squared process as

$$\rho_g = (\alpha + \frac{\alpha^2\beta}{1 - 2\alpha\beta - \beta^2})(\alpha + \beta)^{g-1} \quad \text{for} \quad g \geq 2. \tag{3.18}$$

The conditional normality and the covariance stationarity assumptions can be relaxed, still similar results can be derived. These specific issues are detailed in Ding and Granger (1996).

3.3 Minimum Distance Estimation of GARCH(1,1)-models

We consider the GARCH(1,1) - model defined as in (2.77) and (2.79). It is well known that the empirical autocorrelations of this GARCH(1,1)-process are very small. Empirical autocorrelations of the squared process are however significantly different from zero, even for large lags, see Mikosch and Starica (2000) and Ding and Granger (1996). Figure 3.1 plots the first 50 autocorrelations of the Deutsche Bank returns data already presented in figure 2.2. The generalized Mahalanobis distance between the empirical autocorrelations and the theoretical autocorrelations of the squared process has its minimum at a vector whose coordinates are the estimated ARCH and GARCH parameters of the GARCH(1,1)-process.

Specifically, this estimator is based on the ARMA(1,1) - representation of ϵ_t^2 as given by (2.44) and (2.45). The basic idea is to exploit the fact that, because of (2.44), the theoretical autocorrelations of ϵ_t^2 as derived in the previous section (see (3.15) and (3.17)) are known functions of α and β.

The empirical ρ_g are then estimated by

[3] We shall write $\rho_g = \rho_g(\alpha, \beta)$ to highlight the fact that the theoretical autocorrelation depends solely on α and β at a given lag g.

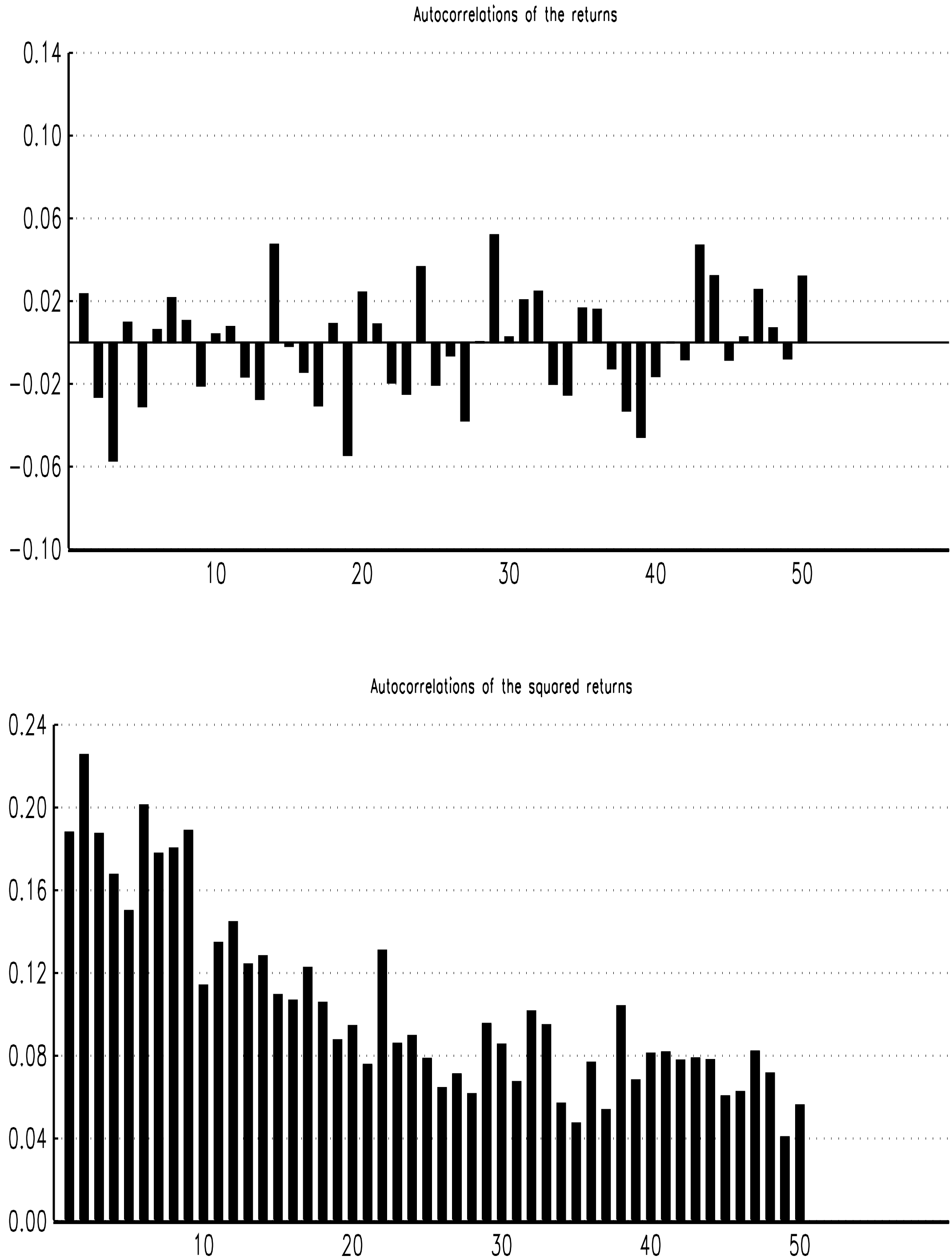


Figure 3.1: Autocorrelations of Deutsche Bank stock returns and squared returns

$$\hat{\rho}_g = \frac{\sum_{t=1}^{n-g}(\tilde{\epsilon}_t^2 - \bar{\tilde{\epsilon}}^2)(\tilde{\epsilon}_{t+g}^2 - \bar{\tilde{\epsilon}}^2)}{\sum_{t=1}^{n}(\tilde{\epsilon}_t^2 - \bar{\tilde{\epsilon}}^2)^2}, \tag{3.19}$$

where $\tilde{\epsilon}_t := r_t - \bar{r}$, and the Minimum Distance Estimators $\hat{\alpha}$ and $\hat{\beta}$ for α and β are obtained as

$$\arg\min_{\alpha,\beta}[\hat{\rho} - \rho(\alpha,\beta)]'\mathbb{W}[\hat{\rho} - \rho(\alpha,\beta)], \tag{3.20}$$

where the $g \times g$ matrix $\mathbb{W}$ is some suitable positive definite weighting matrix, $\hat{\rho} = (\hat{\rho}_1, \ldots, \hat{\rho}_g)'$ and where $\rho(\alpha,\beta) = (\rho_1(\alpha,\beta), \ldots, \rho_g(\alpha,\beta))'$ is a vector defined in (3.17) and (3.18) for $\alpha > 0,\ \beta \geq 0$ and $\alpha + \beta \leq 1$.

3.4 Consistent estimation of $\mathbb{W}$

Standard covariance matrix estimations are concerned with processes whose innovations are *i.i.d.*. This is clearly not the case in (2.44). Baillie and Chung (2001), Domowitz and White (1982) and White (1984) discuss consistent estimation of the covariance matrices when the *i.i.d.* assumption does not hold. They consistently estimate a covariance matrix using the Newey and West (1987) procedure and obtain feasible Minimum Distance Estimators. The following discussion follow Baillie and Chung (2001).

First, we construct the sample autocovariance and autocorrelation of the squared process as:

$$\tilde{\gamma}_k = T^{-1}\sum_{t=k+1}^{T}({\epsilon^2}_t - \bar{\epsilon^2})({\epsilon^2}_{t-k} - \bar{\epsilon^2}) \tag{3.21}$$

and

$$\tilde{\rho}_k = \frac{\tilde{\gamma}_k}{\tilde{\gamma}_0}\ . \tag{3.22}$$

Next, we construct the robust covariance matrix estimator following Domowitz and White (1982) and White (1984):

$$\sqrt{T}(\tilde{\rho}_g - \rho_g) = T^{\frac{-1}{2}}\big(\frac{1}{\tilde{\gamma}_0}\big)\sum_{t=g+1}^{T} \mathbf{Z}_t,$$

where $\mathbf{Z}_t$ is a $g \times 1$ vector defined by

$$\mathbf{Z}_t = \begin{pmatrix} (\epsilon^2{}_t - \bar{\epsilon^2})(\epsilon^2{}_{t-1} - \bar{\epsilon^2}) - \rho_1(\alpha,\beta)(\epsilon^2{}_t - \bar{\epsilon^2})^2 \\ (\epsilon^2{}_t - \bar{\epsilon^2})(\epsilon^2{}_{t-2} - \bar{\epsilon^2}) - \rho_2(\alpha,\beta)(\epsilon^2{}_t - \bar{\epsilon^2})^2 \\ \vdots \\ (\epsilon^2{}_t - \bar{\epsilon^2})(\epsilon^2{}_{t-g} - \bar{\epsilon^2}) - \rho_g(\alpha,\beta)(\epsilon^2{}_t - \bar{\epsilon^2})^2 \end{pmatrix}.$$

By defining

$$\Gamma_j = \mathbb{E}(\mathbf{Z}_t\mathbf{Z}'{}_{t-j}) \tag{3.23}$$

and

$$\mathbf{V}_z = \sum_{j=-\infty}^{\infty} \Gamma_j, \tag{3.24}$$

it is shown in White (1984) that

$$\frac{1}{\sqrt{T}}\sum_{t=1}^{T} \mathbf{Z}_t \longrightarrow \mathbf{N}(0, \mathbf{V}_z). \tag{3.25}$$

Furthermore

$$\sqrt{T}(\tilde{\rho}_g - \rho_g) = \frac{1}{\sqrt{T}}\big(\frac{1}{\tilde{\gamma}_0}\big)\sum_{t=g+1}^{T} \mathbf{Z}_t \quad \text{and} \quad \tilde{\gamma}_0 \longrightarrow \gamma_0, \tag{3.26}$$

allows to conclude that

$$\sqrt{T}(\tilde{\rho}_g - \rho_g) \longrightarrow \mathbf{N}(0, \mathbf{V}_z/\gamma_0^2). \tag{3.27}$$

Newey and West (1987) provide a consistent way to estimate $\mathbf{V}_z$, namely

$$\hat{\mathbf{V}}_z = \hat{\Gamma}_0 - \sum_{j=1}^{q}\big(1 - \frac{j}{1+q}\big)\big(\hat{\Gamma}_j - \hat{\Gamma}'_j\big) \tag{3.28}$$

where $\hat{\Gamma}_j$ is a covariance matrix estimator at lag j, and

$$\hat{\Gamma}_j = \frac{1}{T}\sum_{t=j+1}^{T} \mathbf{Z}^*{}_t \mathbf{Z}^{*\prime}{}_{t-j} \tag{3.29}$$

with

$$\mathbf{Z}^*{}_t = \begin{pmatrix} (\epsilon^2{}_t - \bar{\epsilon^2})(\epsilon^2{}_{t-1} - \bar{\epsilon^2}) - \rho_1(\hat{\alpha},\hat{\beta})(\epsilon^2{}_t - \bar{\epsilon^2})^2 \\ (\epsilon^2{}_t - \bar{\epsilon^2})(\epsilon^2{}_{t-2} - \bar{\epsilon^2}) - \rho_2(\hat{\alpha},\hat{\beta})(\epsilon^2{}_t - \bar{\epsilon^2})^2 \\ \vdots \\ (\epsilon^2{}_t - \bar{\epsilon^2})(\epsilon^2{}_{t-g} - \bar{\epsilon^2}) - \rho_g(\hat{\alpha},\hat{\beta})(\epsilon^2{}_t - \bar{\epsilon^2})^2 \end{pmatrix}.$$

This Newey West technique gives an optimal weighting matrix $\mathbf{W_{NW}}$ consistently estimated by

$$\hat{\mathbf{W}}_{\mathbf{NW}} = (\frac{1}{\hat{\gamma}_0})^2 \hat{\mathbf{V}}_z, \tag{3.30}$$

where

$$\hat{\gamma}_0 = T^{-1}\sum_{t=1}^{T}(\epsilon^2{}_t - \bar{\epsilon^2})^2. \tag{3.31}$$

$\hat{\gamma}_0$ consistently estimates γ_0 and $\hat{\mathbf{V}}_z$ consistently estimates $\mathbf{V}_z$.

The (feasible) Minimum Distance Estimators $\hat{\alpha}$ and $\hat{\beta}$ for α and β in the case of non *i.i.d.* disturbances are obtained as

$$\arg\min_{\alpha,\beta}[\hat{\rho} - \rho(\alpha,\beta)]' \hat{\mathbf{W}}^{-1}_{\mathbf{NW}} [\hat{\rho} - \rho(\alpha,\beta)]. \tag{3.32}$$

The efficiency of this estimator relative to the Maximum Likelihood estimator is evaluated in detail by Baillie and Chung (2001). They found that this estimator can be surprisingly efficient for quite a small number of autocorrelations. It is even more efficient than the quasi maximum likelihood for some regions of the parameter space and for some specific conditional densities. Using Monte Carlo similations, Baillie and Chung (2001) found that for $(\alpha, \beta) \in \{(0.1, 0.55); (0.15, 0.55)\}$ for example, and the conditional density being chi-squared with two degrees of freedom, MDE was better than QMLE both in terms of parameter estimation bias and root mean squared error.

In their empirical example comparing QMLE to MDE, they study 3138 observations of the hourly exchanges rate returns Deutsche Mark versus US dollar. They found that the two estimates had the similar diagnostic properties but exhibit remarkable differences in the value of the autocorrelations of the squares of the fitted returns. It is known see e.g Jacquier, Polson, and Rossi (1994) that MLE estimation of GARCH models are unable to properly replicate stylized facts such as the autocorrelation function of the squared returns. The minimum distance estimator (MDE) of Baillie and Chung (2001) thanks to its criterion function, accurately reproduces the nature of the sample autocorrelations of the squared observations where the MLE produces autocorrelations that are much higher than the corresponding sample equivalents.

3.5 Invariance of the estimated MDE GARCH parameters to the choice of $\mu(b)$ and ω

As done in the case of Maximum Likelihood Estimation, we justify in this section the invariance of the estimated ARCH and GARCH parameters of the GARCH(1,1)-model to the choice of $\mu(b)$, ω and to the scaling of the whole time series. As easily seen in equations (3.18) the theoretical autocorrelations depend solely on α and β, so any change in any of the other parameters or even a scaling of the time series will have no effect on the theoretical autocorrelations.

The empirical autocorrelations as defined in (3.19) depend on the residuals $\tilde{\epsilon} = r_t - \mathbb{E}(r_t) = r_t - \bar{r}$. So if $r'_t = r_t + k$, then

$$\tilde{\epsilon}' = r'_t - \mathbb{E}(r'_t) = r'_t - \bar{r'} = r_t + k - \mathbb{E}(r_t) - k = \tilde{\epsilon}.$$

This shows that the empirical autocorrelations are not affected by a shift of the whole time series. The residuals in this case remain unchanged. A change in ω has no effect on the residuals and therefore no effect in the empirical autocorrelations. This result is different from what will be claimed in chapter 6 and 7 where a structural in the parameters at some point in time within the time series changes the empirical autocorrelations. Here, we stress that changes in μ and ω happen for the whole time series. We see that none of this changes affect the empirical autocorrelations.

We consider now the case where we scale the time series by a constant k. We have a new time series say $r't = kr_t$. Then $\tilde{\epsilon}'_t = kr_t - k\mathbb{E}(r_t) = k\tilde{\epsilon}_t$. The autocorrelations do not change because the scaling factor of k^2 appear in both the numerator and the denominator in (3.19). Again, here we consider the whole time series. We shall see in chapter 6 that scaling some part of the time series changes the empirical autocorrelations.

The remaining possible problem could be the effect of this scaling on the weighting matrix. Going through the derivation of the consistent estimation of the covariance matrix, one easily sees that $\hat{\mathbf{V}}_z$ will be multiplied by k^4. This comes from the product $\mathbf{Z_t Z'_t}$ where $\mathbf{Z_t}$ and $\mathbf{Z'_t}$ each is is multiplied by k^2. But the weighting matrix $\hat{\mathbf{W}}_{\mathbf{NW}} = (\frac{1}{\hat{\gamma}_0})^2\hat{\mathbf{V}}_z$ contains the term $\hat{\gamma}_0$ which as well is scaled by k^4. So k^4 appear in both the numerator and the denominator. The weighting matrix remain unchanged under the scaling of the time series. Even if weighting matrix would be scaled my a non zero constant factor, this might change the value of the criterion function at the extremum but not the extremum itself.

So as the case of maximum likelihood estimators, changes in parameters μ and ω or a scaling of the whole GARCH time series do not affect the minimum distance estimators α and β.

3.6 Conclusion

In this chapter, we have introduced the minimum distance estimation of a GARCH(1,1)-model and discussed the consistent estimation of the weighting matrix given when the disturbances are not *i.i.d.*. This allows us to obtain feasible MDE parameters. We have shown that the theoretical autocorrelations by definition are insensitive to changes of parameters others than α and β. For the empirical autocorrelations, we have found that they remained unchanged under changes of μ, ω and the scaling of the whole time series. The weighting matrix being as well unchanged under these changes, we reached the conclusion that the MDE are invariant under these operations.

As found in Baillie and Chung (2001), MDE favorably competes with QMLE in some regions of the parameter space and for some conditional densities. In the following chapter we will show that the estimated persistence in GARCH(1,1)-model is severely biased in small sample. An alternative will be the MDE just introduced.

Chapter 4

Small Sample Bias in the Estimated Persistence of GARCH(1,1)-Models

4.1 Introduction

The Tsunami of December 26 in 2004, September 11, the German Reunification of October 3rd 1990 and natural catastrophes are among many other economic shocks that shake financial markets. Particularly, returns data (stocks, interest rates, foreign exchange rates) indicate that investments at certain periods are riskier than others. As is well known, these risky times do not occur randomly across time. There is degree of autocorrelation in the riskiness of financial returns.

A question of interest is naturally the accuracy of the prediction of the GARCH model. We are interested in the variance of the error term and what can make it fluctuate. Fitting financial daily data into a GARCH(1,1)-model reveals a sum of the ARCH and GARCH parameter close to one (see chapter 6).

Bollerslev and Engle (1986) have introduced the IGARCH to highlight this fact. Fractionally integrated financial data also exhibits high persistence. This property is becoming a stylized fact in empirical finance (see Granger (2005)).

The finite sample evidence on the performances of the GARCH maximum likelihood estimator is still very limited. Lumsdaine (1995) and Bollerslev and Wooldridge (1992) belong to the very few that investigate these performances and their findings are reported in Bollerslev, Engle, and Nelson (1994) where one reads *" For the GARCH(1,1) with conditional normal errors, the available Monte Carlo evidence suggests that the estimates for* $\hat{\alpha} + \hat{\beta}$ *is downward biased and skewed to the right in small samples. The bias in* $\hat{\alpha} + \hat{\beta}$ *comes from a downward bias in* $\hat{\beta}$ *while* $\hat{\alpha}$ *is upward biased".* In particular, Lumsdaine (1995) in the study of the finite sample properties of the maximum likelihood estimator in the GARCH(1,1)-model finds that this estimator has a normal limiting distribution and a constant covariance matrix. Furthermore, the asymptotic distribution is, for the most part, well approximated by the estimated t statistics. Other statistics such as the Lagrange Multiplier test, Likelihood ratio and Wald test however, do not behave as well in small samples. The ARCH and GARCH estimators are skewed in small samples. The study reports a "pileup" effect at the boundary of parameter true value, an effect which decreases as the sample size increases. The tails of the small sample distribution are heavier than those of the normal distribution. The skewness of the estimated persistence in finite sample is more pronounced in the ARCH parameter. This is caused in part by the estimation design where the parameter α and β are constrained to be in the open unit interval. This restriction is the so called Bollerslev non-negativity condition. Recent research, see (e.g Soosung and Pereira (2006)) addresses the issue of relaxing the Bollerslev non-negativity condition and we will be discussing this as well in the next paragraph. This restriction does create a truncation in the estimation causing the finite sample distribution to resemble a truncated distribution. This effect disappear when the sample size is large enough to allow the estimated distribution to lie entirely within the unit interval. The finite sample distribution differs the most from its limiting distribution when the true value of α and β are close to the boundary of the unit interval. So a question of interest as well is to determine how large a sample size must be before the asymptotic distribution of the maximum likelihood estimator is well approximated by the finite sample distribution.

Recent evidence is found in Soosung and Pereira (2006) where the Bollerslev non-negativity condition is relaxed. They find that the estimated persistence is still biased but the size of the bias is smaller than the size of the bias in Lumsdaine (1995). In fact as already mentioned by Lumsdaine (1995) and further documented in Nelson and Cao (1992), they find as well that the Bollerslev non-negativity condition are a serious restriction in small samples in GARCH(1,1)-model. In small samples, this restriction causes a huge number of convergence problems. When they impose the following weaker non-negativity condition

$$\omega > 0 \;\text{ and }\; \sigma_t^2 > 0 \tag{4.1}$$

instead of the Bollerslev non-negativity condition

$$\omega > 0,\; \alpha \geq 0,\; \text{and } \beta \geq 0, \tag{4.2}$$

the number of convergence errors decreases significantly. Comparing their parameter estimate results with Lumsdaine (1995), they find that the ARCH estimates $\hat{\alpha}$ are negatively biased in their case while positively biased in the case of Bollerslev non-negativity condition as studied in Lumsdaine (1995). Their estimated GARCH parameter $\hat{\beta}$ is severely negatively biased and with small α, some $\hat{\beta}$'s are even negative. When the Bollerslev non-negativity conditions are imposed, the negative bias of $\hat{\beta}$ becomes smaller but the convergence error increases significantly. The bias decreases gradually as the sample increases.

They analyze as well the interdependence of α and β and find for example that $\hat{\beta}$ is affected by the size of α, becoming in fact less biased for large α. They conclude their study by recommending the use of their weak non-negativity conditions rather than the Bollerslev one at least as a pre check because, although they do not guarantee positive volatility, they do reduce convergence errors significantly.

Next we extend their investigation to the case where the estimated parameters of the GARCH(1,1)-model are obtained by minimum distance estimation. Our point is to show that with the errors being normally distributed, for some regions of the parameter space, the minimum distance estimator performs better than the exact maximum likelihood in small samples. Asymptotically, the exact maximum likelihood will of course perform better and we highlight this as well by including sample size of 1000 and 2000 and in some case 3000 and 4000 to show that ultimately, as the sample size grows, minimum distance estimators are outperformed. Our investigation relies on a Monte Carlo study under different parameters. Our conclusion describes as well the source of the bias in different cases.

4.2 Small Sample Bias of the Estimated Persistence

In this section, we compare the performances of the minimum distance estimators with the (exact)maximum likelihood estimators by means of a Monte Carlo study. Table 4.1 shows the parameters combinations investigated. The symbol X in the table points cases where the MDE outperforms the MLE. This analysis is particular because the general consensus is that no other estimator competes with exact maximum likelihood. This is certainly true asymptotically and we can observe it as the sample sizes grows. However, in small samples, we find regions of the parameters space where the distribution free method of minimum distance estimation of GARCH(1,1)-models competes and even outperforms the exact maximum likelihood especially in terms of the parameter bias and in some cases in terms of the mean squared error. We focus on sample size up to five hundred, this case only represents already about 2 years of daily data, 10 years of weekly data and about 40 years of monthly data. The last two are time spans usually called for in macro-econometric analysis. Our conclusion is based on samples of this size and we additionally report the result for the sizes 1000, 2000, and in some case 3000 and 4000 to strengthen the common wisdom that exact maximum likelihood is asymptotically unbeatable. In this particular minimum distance estimation, we used the first 10 lags and the weighting matrix calculated using Newey and West (1987). Other number of lags and other weighting matrices provide roughly the same results and yield the same conclusions. In the alternative of maximum likelihood, procedures programmed in Gauss Fanpac module are called in estimating the exact maximum likelihood GARCH model.

We found and reported 15 cases where the distribution free method outperforms the exact maximum likelihood in terms of the parameter estimation bias. Comparative tables of the MDE and MLE estimated parameters are in the appendix. We designed the experiment in such a way that we can as well address the issue of small sample bias in terms of explaining the particular behavior of the estimated ARCH and GARCH parameters as done in Bollerslev, Engle, and Nelson (1994) and Lumsdaine (1995). Their results are obtained using maximum likelihood methods. Here, we address the same using the minimum distance estimation. To achieve that, we generate GARCH data of sample size 500. We estimated subsamples of sizes 100, 200, 300, 400 and the entire simulated data which has a sample size of 500. For each subsample, the ARCH and

α \ β	0.60	0.70	0.73	0.75	0.77	0.80	0.82	0.83	0.90
0.05									X
0.055						X			X
0.070			X					X	
0.075		X							
0.080						X	X		
0.09					X	X			
0.10		X		X	X	X			
0.16	X								

Table 4.1: Parameters region investigated

GARCH parameter are estimated both in MDE and MLE. Reported results are averages of 10000 replications. The study of Lumsdaine (1995) studies sample sizes of 500 and 200 (results not reported) with 500 replications. The research of Soosung and Pereira (2006) considers sample sizes of 100, 250, 500 and 1000 in which 1000 replications are used.

The different sample sizes considered are representative of the sizes of monthly and quarterly data sets commonly used empirically. We are not concerned with large samples here because the asymptotic properties are are less questionable and Monte Carlo evidence of Baillie and Chung (2001) and Storti (2006) address this larger sample issues . They both find that exact maximum likelihood is the best estimation method asymptotically. In our investigation here, we cover a wider region of the parameter space, and we run simulations based on 10000 replications. The $\omega = 0.1$ and $\mu(b) = 0$ as opposed to Lumsdaine (1995) who used $\omega = \mu(b) = 1$. We justified earlier that the choice of the parameter other than α and β does not changes the values of the estimated parameters $\hat{\alpha}$ and $\hat{\beta}$.

Figures 4.1 and 4.2 present two cases where MDE outperforms MLE in small samples. Entries in these figures are averages of 10000 runs. In figure 4.1, one sees that the estimated MLE persistence increases rapidly and table 10.4 shows that by the sample size of 1000, it has already crossed the estimated MDE persistence. In the second figure (figure 4.2) the estimated MLE persistence increases at a slower rate and as it can be read in table 10.6, it crosses the estimated MDE persistence by the sample size of 3000. The point here is that although MLE is better asymptotically, the rate at which is catches and crosses the estimated MDE persistence depends on the region of the parameter space.

The estimated α and β for both exact maximum likelihood and minimum distance estimation are reported next to each others to ease direct comparison in the appendix (see tables 10.4, 10.5, 10.6, 10.7 and 10.8). For all the parameter settings presented, the minimum distance estimator outperformed the exact maximum likelihood in terms of the parameter bias. In all the other cases, with a few exceptions in the sample sizes of 100 of 200 or sometimes 300, exact maximum likelihood was better than the minimum distance estimation.

We first look at the estimates of the maximum likelihood method. There is downward bias of the estimated persistence as documented by Bollerslev, Engle, and Nelson (1994). This downward bias is explained by an overestimated ARCH parameter and and underestimated GARCH parameter. The estimated ARCH parameter decreases as the sample size increases, but stays above the true parameter in our estimations, in all the parameters settings and for all the different sub sample sizes. The estimated GARCH parameter increases as the sample size increases but stays under the true parameter. Its increase rate is however strong enough to push the sum $\hat{\alpha} + \hat{\beta}$ to increase as the sample size increases. The estimation gets very accurate as the sample size increases since the mean squared errors decrease sharply at the same time.

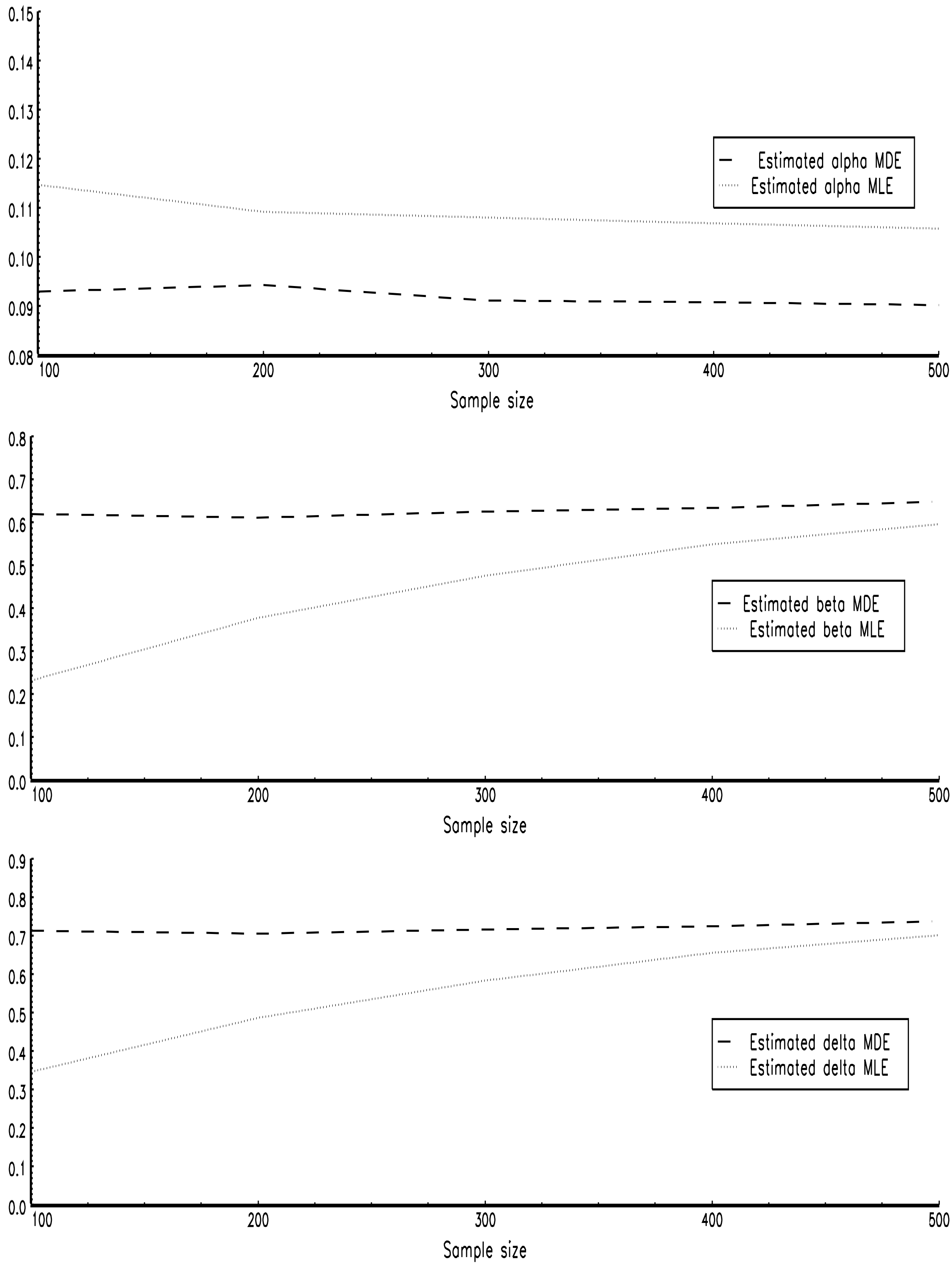


FIGURE 4.1: TRUE PARAMETER $(\alpha, \beta) = (0.10, 0.75)$ FOR $\delta = \alpha + \beta = 0.85$

Next, we consider the estimates of minimum distance estimation method. As opposed to the maximum likelihood estimation case where the estimated persistence $\hat{\delta} = \hat{\alpha} + \hat{\beta}$ increases as the sample increases, the $\hat{\delta}_{MDE}$ seems not to have a clear variational path. However, in most of the cases where $\alpha > 0.10$, the estimated persistence increases as the sample size increases. We observe as well a downward bias. The estimated ARCH parameter $\hat{\alpha}$ and the GARCH parameter $\hat{\beta}$ do not have a clear and determined behavioral path as in the exact maximum likelihood case. The minimum distance method is very good in estimating very small sample size. Already with sample size of about 100 one gets an estimated persistence already close to the true persistence. Again the estimation gets very accurate as the sample size increase and this as well is noticeable in the mean squared error which then becomes very small.

Looking at $\hat{\alpha}_{MDE}, \hat{\alpha}_{MLE}, \hat{\beta}_{MDE}, \hat{\beta}_{MLE}, \hat{\delta}_{MDE}$ and $\hat{\delta}_{MLE}$ all together, we can observe that, while having $\hat{\delta}_{MDE} = \hat{\alpha}_{MDE} + \hat{\beta}_{MDE} > \hat{\delta}_{MLE} = \hat{\alpha}_{MLE} + \hat{\beta}_{MLE}$ in all the 15 cases, $\hat{\alpha}_{MDE} < \hat{\alpha}_{MLE}$ and $\hat{\beta}_{MDE} > \hat{\beta}_{MLE}$. As sample sizes grows as seen for example in the sample sizes of 1000 and above, while $\hat{\alpha}_{MDE}$ stays smaller than $\hat{\alpha}_{MLE}$, $\hat{\beta}_{MLE}$ becomes so greater than $\hat{\beta}_{MDE}$ that $\hat{\delta}_{MLE}$ gets superior to $\hat{\delta}_{MDE}$.

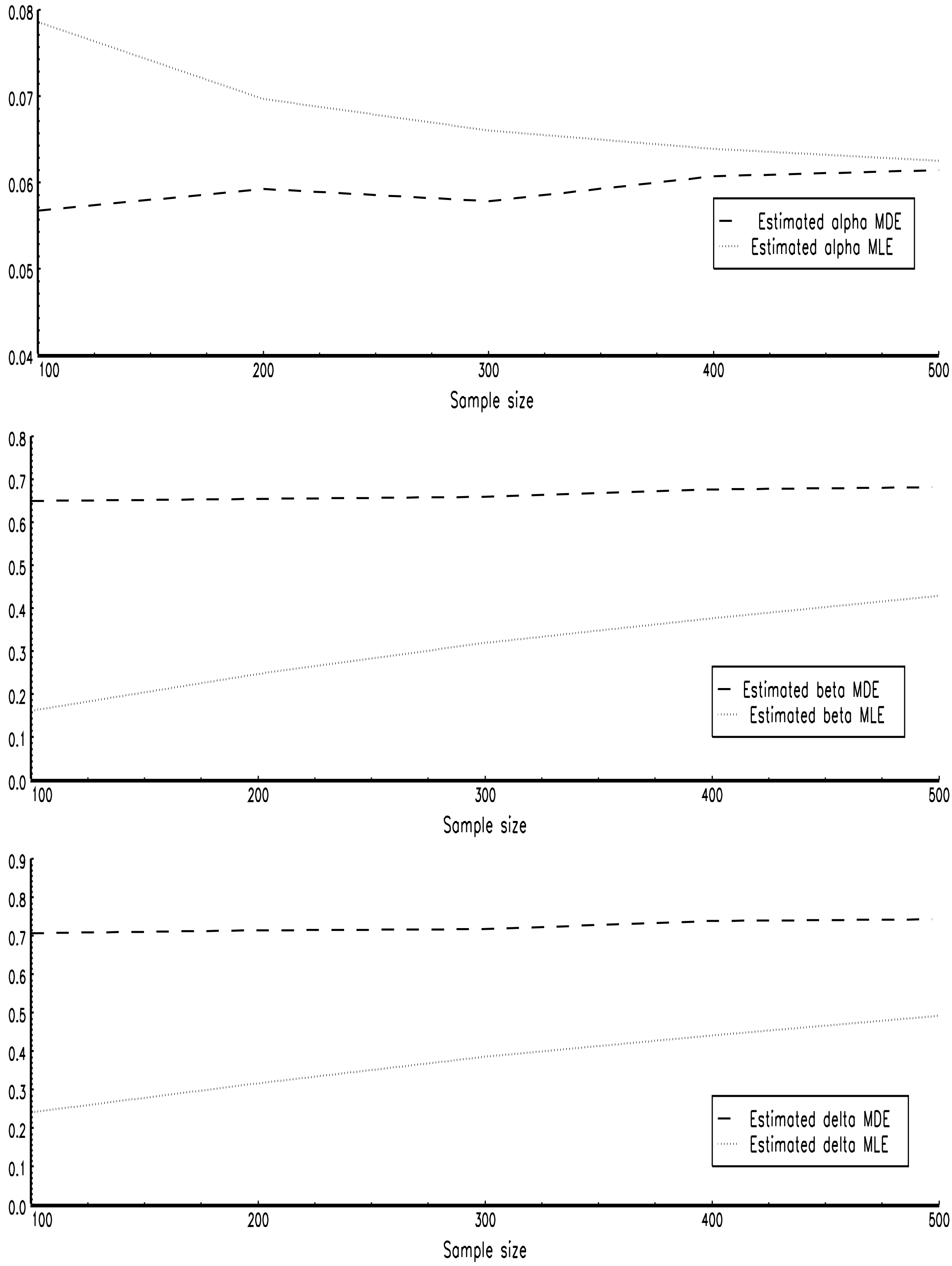


FIGURE 4.2: TRUE PARAMETER $(\alpha, \beta) = (0.055, 0.80)$ FOR $\delta = \alpha + \beta = 0.855$

4.3 Conclusion

In this chapter, we have addressed the issue of small sample bias of the estimated persistence of the GARCH(1,1)-model. In an attempt to describe, compare and identify the driving forces behind the behavior of the estimated persistence in both estimation methods, we simulated GARCH data in increasing sample sizes in 40 different parameters settings. We found 15 cases where the Minimum distance Estimator outperforms the Maximum Likelihood Estimator. Specifically on these cases, we could further say that:

- The estimated persistence increases as the sample size increases in both methods of estimation.

- The estimated persistence in the Maximum Likelihood case is downward biased and this bias is explained by and upward biased in $\hat{\alpha}_{MLE}$ and a downward biased in $\hat{\beta}_{MLE}$. This confirms results in Bollerslev, Engle, and Nelson (1994). In the case of minimum distance estimators, the downward bias in the estimated persistence is explained by the downward bias of $\hat{\alpha}_{MDE}$ and the downward bias in $\hat{\beta}_{MDE}$.

- For some regions of the parameter space, Minimum Distance Estimators perform better than Maximum Likelihood in small samples, in term of bias of the estimated persistence. This fact in small sample is explained by the difficulties of the model to properly estimate the β's. As the sample sizes grow however, the while $\hat{\alpha}_{MDE} < \hat{\alpha}_{MLE}$, we realize that $\hat{\beta}_{MDE}$ gets smaller than $\hat{\beta}_{MLE}$ in such a way that $\hat{\delta}_{MLE} > \hat{\delta}_{MDE}$.

- The estimation becomes more accurate when the sample sizes grows as the mean squared error decreases.

  We have found cases where Minimum Distance Estimators are better than Maximum Likelihood Estimators. We are not saying that Minimum

Distance Estimators are better than Maximum Likelihood Estimators in small samples, rather that Maximum Likelihood is not always the best estimator when sample sizes are smaller than say 500. Our recommendation is to run pre check Monte Carlo simulations to compare the two estimation methods before choosing any of them.

The investigations in the case of minimum distance estimation in this chapter have been done using the first 10 lags. Nothing prevents us from using 5 or 15 or even 70 lags. The issue of optimal lag choice the minimum distance estimation framework is an important one (see e.g Storti (2006)). If the number of lags is too large, we face time constraints in computing the parameters estimates necessary for risk measures such as the hourly or daily value at risk. In the following chapter, we will be addressing this issue.

Chapter 5

Lag Choice in Minimum Distance Estimation of GARCH(1,1)-Models

5.1 Introduction

Empirically, Baillie and Chung (2001) have applied the MDE to hourly exchange rates using the first 10 lags. Storti (2006) uses the first 20 lags to estimate hourly returns on the FTSE100 futures. Storti (2006) acknowledges that *" the development of a formal identification procedure for the value of g is of course an important point worth to be investigated in future work"*. Our purpose in this chapter is to do just that. Our primary optimality criterion is the mean squared error. It is a better optimality criteria than the bias alone because it considers not only the bias (more precisely the squared bias), but the variance of the estimated parameter as well.

In chapter 3, we have introduced both the GARCH(1,1)-model and the minimum distance estimator of its parameters. In the next section, we directly present the results of our investigations.

5.2 Finite Sample Results on the Optimal Lag

This section presents the results based on 1000 replications of the Monte Carlo analysis. We deal successively with the following sample sizes: 250, 500, 1000, 2000, 4000 and 8000. We consider the two following cases :
$\alpha = 0.15, \beta = 0.70$ and $\alpha = 0.20, \beta = 0.50$.
The ω is the same in the two settings $\omega = 0.001$. These two parameters settings are chosen because of their frequent use in this type of analysis, see Baillie and Chung (2001) and Storti (2006).

Figures 5.1 and 5.2 plot the mean squared error as a function of lags for the sample sizes of 1000 and 4000. The figures for the remaining sample sizes are found in the appendix. Entries in the figures are based on means of 1000 runs. Figures 5.3 and 5.4 plot the mean squared error as a function of lags in this second parameter setting for the sample sizes of 250 and 500 respectively.

We could not find a direct functional relationship between the optimal lag for the estimated α, β and δ and the sample size. In itself, this is not a bad result as any type of relationship between the optimal number of lags and the sample size might not keep the number of lags bounded. In particular, a functional relationship of the type of $g = \sqrt{T}$ will still require 100 lags for a sample size 10000, which will be computationally demanding even for a high speed computer.

As a general finding in this investigation, the mean squared error decreases sharply in the first lags for all the three estimated parameters. This finding is consistent with all the six samples size used in our analysis. In general, the optimal lags for the parameter estimation of α, β and δ are different from one another. For the ease of the discussion and because the whole GARCH literature is centered about the persistence parameter, we will limit ourselves to optimal lag choice in the case of the estimated persistence.

The main finding is that the mean squared error decreases sharply within the first lags (see figures 5.1 to 5.4 hereafter). The decreasing rate as seen in the figure is very high in the very first lags, especially from lag 2 to lag 10. Afterwards, it decreases, but at a much slower rate and after lag 30 and in some case 40, the changes are only marginal. The interesting fact here is that this behavior is independent of the sample size. The optimal number of lag is often found to be beyond 50. A mistake of at most 2 percent will be made when choosing a smaller lag instead of these optimal lags across the different sample sizes considered.

A similar investigation under the same parameter settings has been done using the bias as optimality condition. Figures 5.5 till 5.8 show the corresponding graphs. In both graphs, one sees that a mistake of about 3% is made by choosing a small lag instead of the optimal lag with is a large number. The remaining graphs are in the appendix. A similar argument looking at those remaining graphs justifies the choice of a small number of lags.

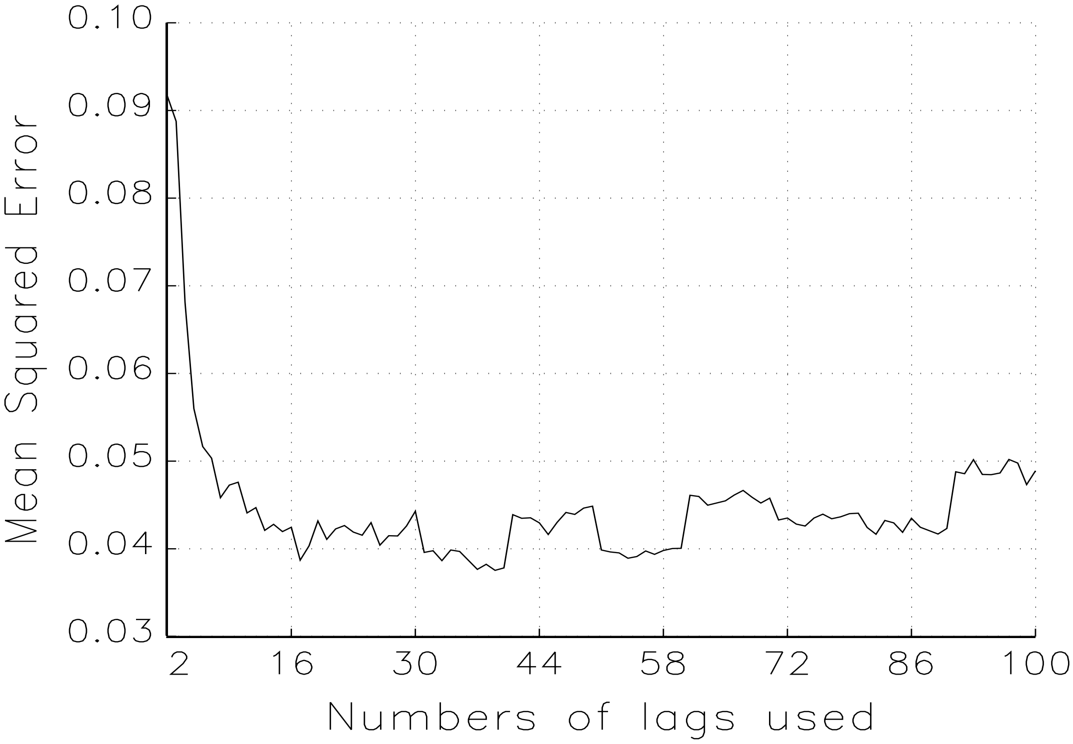


FIGURE 5.1: MSE OF $\hat{\delta}$ AS A FUNCTION ON NUMBER OF LAGS IN CASE OF $\alpha = 0.15$, $\beta = 0.70$ AND $\omega = 0.001$ SAMPLE SIZE OF 1000

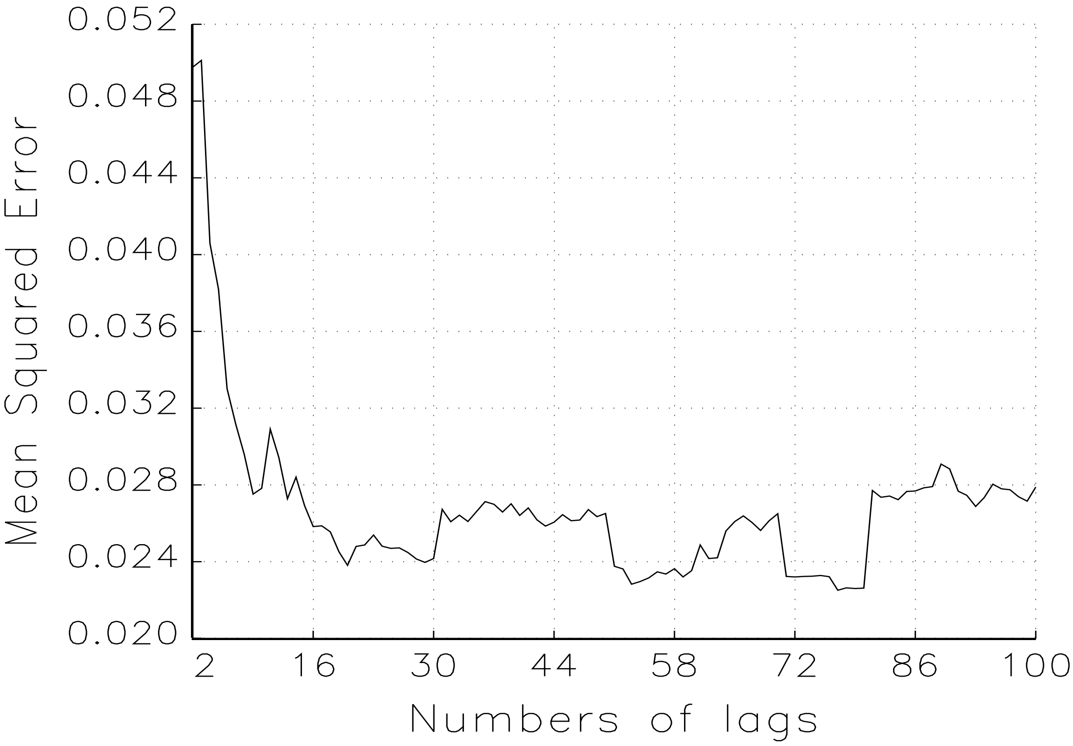


FIGURE 5.2: MSE OF $\hat{\delta}$ AS A FUNCTION ON NUMBER OF LAGS IN CASE OF $\alpha = 0.15$, $\beta = 0.70$ AND $\omega = 0.001$ SAMPLE SIZE OF 4000

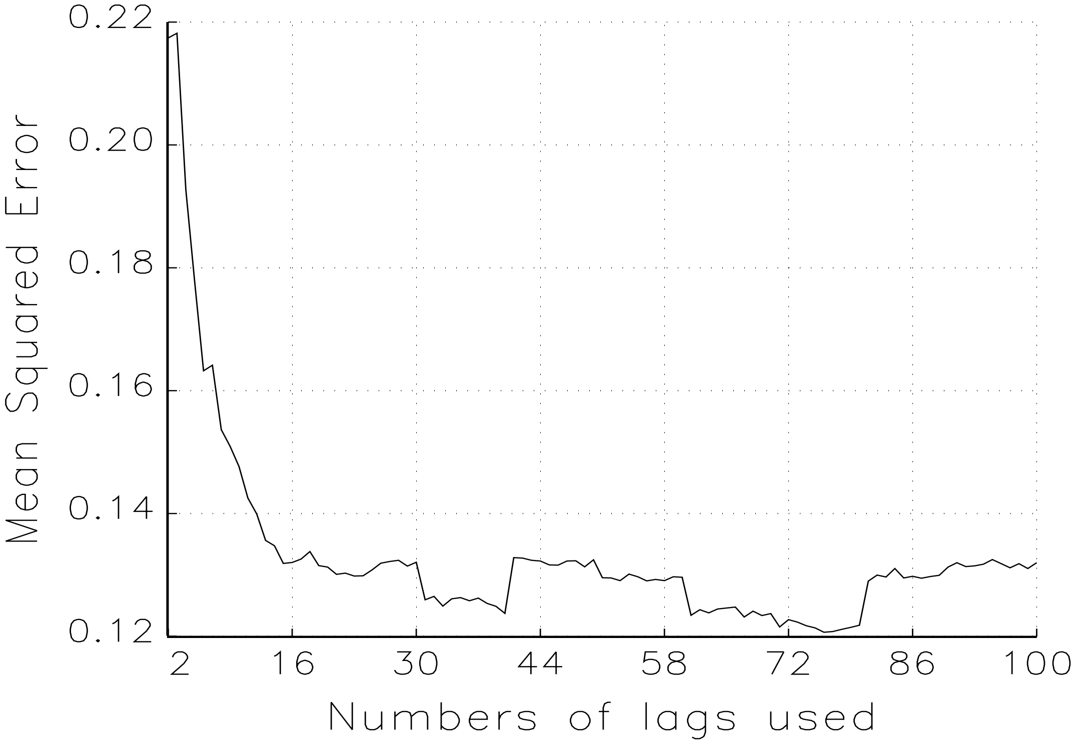


FIGURE 5.3: MSE OF $\hat{\delta}$ AS A FUNCTION OF NUMBER OF LAGS IN CASE OF $\alpha = 0.20$, $\beta = 0.50$ AND $\omega = 0.001$ SAMPLE SIZE OF 250

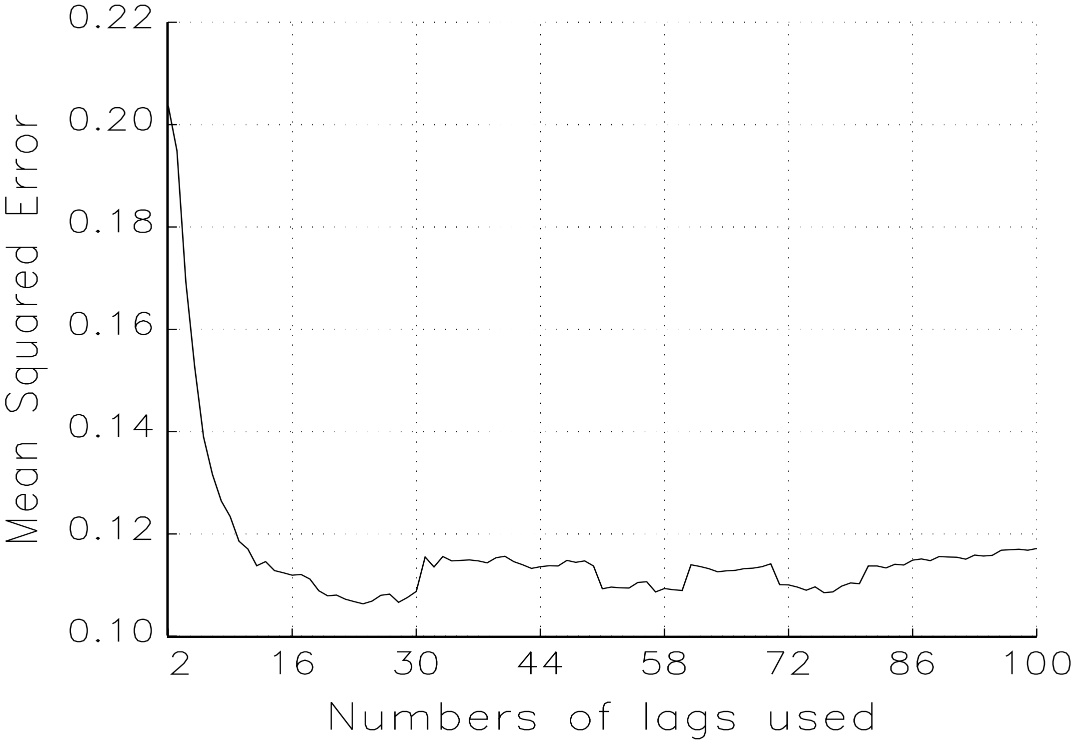


FIGURE 5.4: MSE OF $\hat{\delta}$ AS A FUNCTION OF NUMBER OF LAGS IN CASE OF $\alpha = 0.20$, $\beta = 0.50$ AND $\omega = 0.001$ SAMPLE SIZE OF 500

5.3 Conclusion

The estimation of the parameters of a GARCH(1,1) model using a minimum distance estimator applied to the empirical autocorrelations of the squared process is a viable method. We have properly addressed the issue of optimal lag choice through intensive Monte Carlo investigations under different settings. We have found that:

- There is no functional relationship between the optimal lag and the sample size.
- The optimal lag for the estimated persistence occured in general beyond lag 50. These high lags would be time consuming and computationally demanding if they were to be used.
- Similar results holds for the estimated α and β.
- Our main finding is that the mean squared error decreases sharply within the first lags. This is robust with all the sample sizes simulated and the two estimated parameters.
- By allowing a mistake of less than 2% one can choose a small number of lags in estimating persistence, speeding up therefore the estimation procedure. This holds as well for the two estimated parameters α and β taken separately.
- A similar investigation has been done using the bias as optimality condition. The results were mixed and a clear path was not found. But interestingly, the fluctuations of the graphs in early lags justify the possible use of a small number of lags. The mistake made is around 3%.

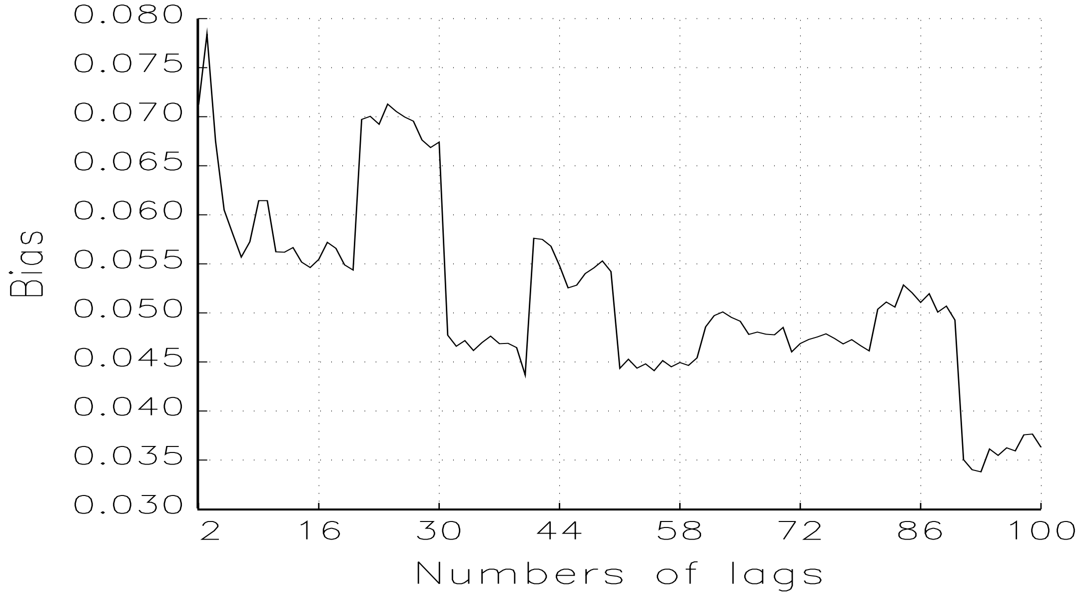


FIGURE 5.5: BIAS OF $\hat{\delta}$ AS A FUNCTION OF NUMBER OF LAGS IN CASE OF $\alpha = 0.15$, $\beta = 0.70$ AND $\omega = 0.001$ SAMPLE SIZE OF 2000

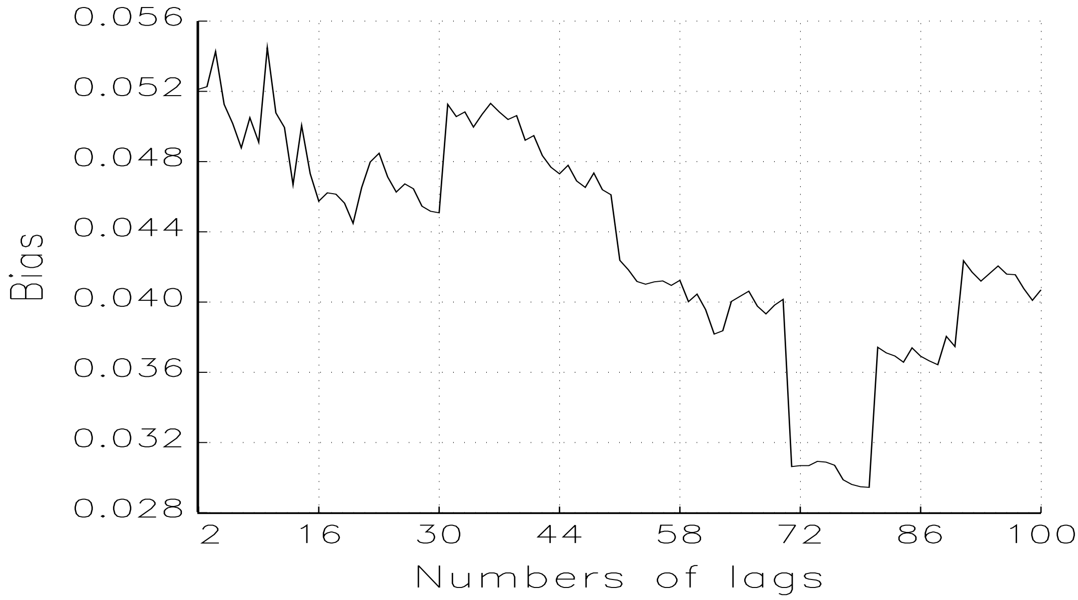


FIGURE 5.6: BIAS OF $\hat{\delta}$ AS A FUNCTION OF NUMBER OF LAGS IN CASE OF $\alpha = 0.15$, $\beta = 0.70$ AND $\omega = 0.001$ SAMPLE SIZE OF 4000

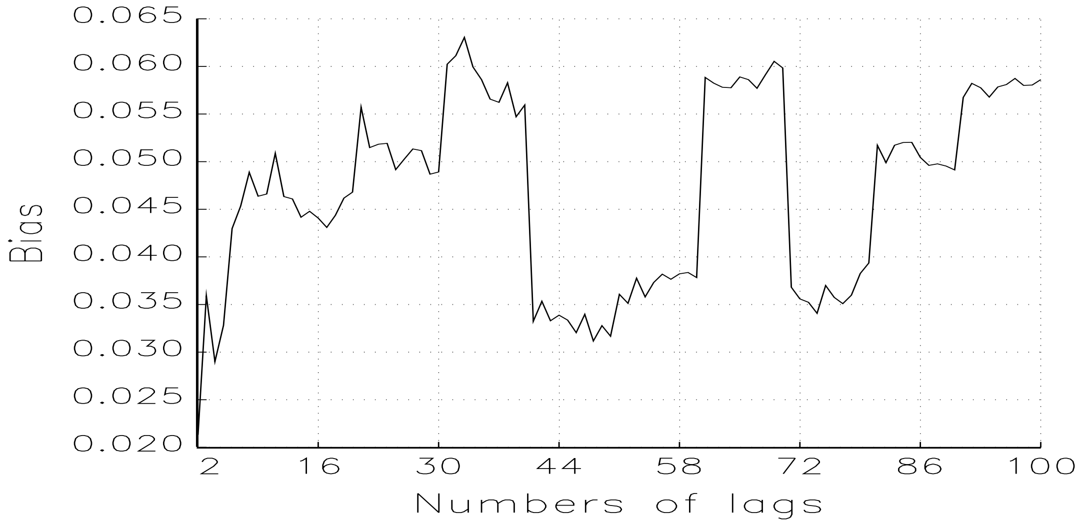


FIGURE 5.7: BIAS OF $\hat{\delta}$ AS A FUNCTION OF NUMBER OF LAGS IN CASE OF $\alpha = 0.20$, $\beta = 0.50$ AND $\omega = 0.001$ SAMPLE SIZE OF 500

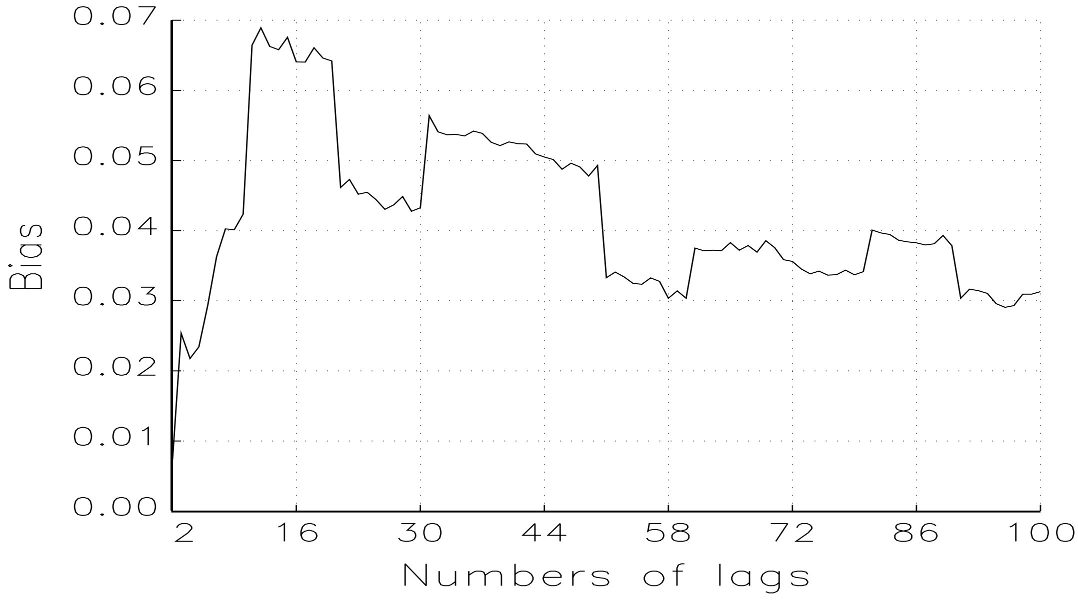


FIGURE 5.8: BIAS OF $\hat{\delta}$ AS A FUNCTION OF NUMBER OF LAGS IN CASE OF $\alpha = 0.20$, $\beta = 0.50$ AND $\omega = 0.001$ SAMPLE SIZE OF 1000

Chapter 6

Structural Change and Estimated Persistence

6.1 Introduction

This chapter is based on Krämer and Tameze (2007)

The GARCH(1,1) - model as already defined in (2.77) and (2.79) is

$$r_t = \epsilon_t + \mu \ , \tag{6.1}$$

$$\epsilon_t = \eta_t \sigma_t \ , \tag{6.2}$$

$$\sigma_t^2 = \omega + \alpha \epsilon_{t-1}^2 + \beta \sigma_{t-1}^2, \tag{6.3}$$

where $\eta_t \sim iid(0,1)$ and r_t - the variable to be "explained" - is typically the rate of change of some economic quantity. This model is still the main workhorse in all areas of applied economics whenever conditional heteroskedasticity is seen to be a problem. Almost from the moment it was born, it was however plagued by the observation that in many applications, the estimate of the "persistence parameter" $\delta := \alpha + \beta$, no matter in which way obtained, was viewed as much too large (in the sense that the superior forecasting performance implied by high persistence did not materialize in empirical applications), and that this upward bias towards the maximum of 1 increases with increasing sample size.

For illustration, figure 6.1 plots various estimates that have been reported in the literature against the sizes of the respective samples. For ease of comparison, we confine ourselves to studies which use daily data. A more detailed description of the studies summarized in figure 6.1 is in table 10.2. The figure clearly demonstrates that estimated persistence increases with sample size and is almost indistinguishable from unity for samples of size 2000 or more.

Diebold (1986) was probably the first to point out that this upward tendency of estimated $\delta's$ might be due to a switch in regime somewhere in the sample, the probability of which increases with increasing calendar time. Lamoureux and Lastrapes (1990), Hamilton and Susmel (1994) or Mikosch and Starica (2004), among many others, show that empirical estimates of δ indeed decrease when the sample is split according to some sensible criterion, and propose generalizations of (6.1) and (6.3) to account for changes in the parameters.

When standard GARCH(1,1)-models are fitted to data generated from such more general models, empirical estimates $\hat{\delta}$ of δ are rather close to, but usually less than one. Haas, Mittnik, and Paollela (2004a) (Figure 1) show by Monte Carlo simulations that $\hat{\delta}$ approaches 1 as persistence in their Markov - switching model increases; Mikosch and Starica (2004) show analytically that the Whittle - estimator of δ becomes arbitrarily close to one if the differences in the variances of their sub-models tend to infinity. This chapter considers the Minimum Distance Estimator (MDE) of α and β and shows that the sum of the estimated α and β can likewise be made arbitrarily close to one if there are

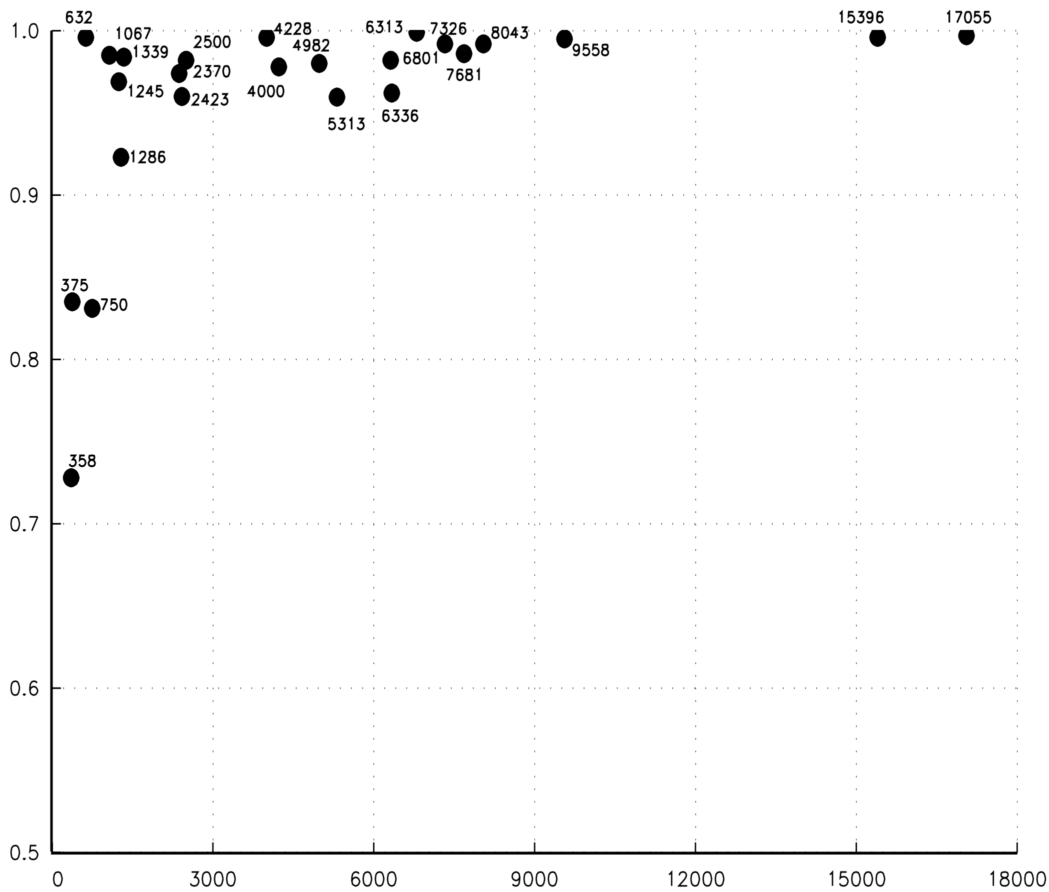


FIGURE 6.1: ESTIMATED PERSISTENCE AS A FUNCTION OF SAMPLE SIZE

structural changes in the unconditional expectation μ of the r_t-process whose number is small relative to sample size, and if the size of the structural changes is large enough.

6.2 Empirical Estimates of the QMLE GARCH(1,1)-model

As said earlier, our interested in GARCH models is around the estimation of persistence parameter. It tells us how long shocks would stay in the GARCH time series. In this section and in the next, we will discuss empirically reported estimates of the persistence parameter of the GARCH(1,1)-model using stocks, interest rates of foreign exchange rates returns at different frequencies.

Tables (10.2, 10.3 and 10.1) in the appendix presents empirical estimates of the persistence parameter of the GARCH(1,1)-model. We present successively daily, weekly and monthly data, all estimated with the quasi maximum likelihood. For the minimum distance estimation, only two results have been reported in the literature. These entries are selected from the studies published in leading economic journals. These tables contain authors names, nature of the data, sample size and the estimated persistence.

In general, as already highlighted by numerous authors (see e.g Baillie and Bollerslev (1990) and Granger (2005)) high persistence does occur in the very special case of daily data, especially for sample sizes covering long time span. The average of the estimated persistence in the case of daily data estimated from sample size of 2000 or more is 0.981. While studies with daily data have frequently found integrated GARCH behavior, studies with higher frequency data (hourly data, tick by tick data) over shorter time spans show weaker persistence. Focussing on daily data however, ensures that sample size is proportional to calendar time, which appears to be the real driving force behind the increase in the estimated persistence.

Considering hourly exchange rate data of the British Pound, the Deutsche Mark, the Swiss Franc, the Japanese Yen each with respect to the US Dollar, Baillie and Bollerslev (1990) found an average persistence of 0.558 with a sample size of 3191. Others frequency studies revealed weaker persistence compared to the persistence shown by daily data. In the case of weekly and monthly data, the averaged estimated persistence for the data in tables 10.3 and 10.1 are 0.903 and 0.916 respectively.

6.3 Empirical Estimates of the MDE GARCH(1,1)-model

Two studies report Minimum Distance estimates of the GARCH(1,1)-model in the literature. The first is done by Baillie and Chung (2001) where the minimum distance estimator is based on the autocorrelations of the GARCH squared process. They used a lag length of 10 and the Newey West weighting matrix to find a persistence of 0.606 for hourly exchange rate between the British Pound and the US Dollars, a total of 3191 observations. The second study is the recent work of Storti (2006) where his minimum distance estimator estimates the GARCH parameters by minimizing the Mahalanobis distance between empirical and theoretical auto-covariances of the GARCH squared process. He uses a lag length of 20 to find a persistence of 0.967 with a sample size of 1673 of hourly futures.

6.4 Structural Change and Sample Correlations

There is a particular relationship between certain types of structural change in the model (6.1) and (6.3) and the estimated autocorrelations of the ϵ_t^2. Most models that allow for changes in the coefficients of (6.1) and (6.3) do so by letting μ, ω, α or β depend on the (unobserved) state of a finite - dimensional Markov chain. Recent examples and variants thereof, with useful surveys of the literature, are Francq, Zakoian, and Roussignol (2001), Klaassen (2002) or Haas, Mittnik, and Paollela (2004a). Alternatively, Hamilton and Susmel (1994), Liu (2000) or Wong and Li (2001) consider

$$\epsilon_t^* := f(\Delta_t)\epsilon_t, \tag{6.4}$$

where ϵ_t is generated by(6.1) (or some variant thereof), and f again depends on the state of some Markov-process $\{\Delta_t\}$ or some other stochastic process. Here, structural changes do not affect the dynamics of the process, just the scale. Other examples are Dueker (1997), who considers changes in the variance of the innovations η_t, or Mikosch and Starica (2004), who simply collect together different sub-samples from different stationary models. All of these models imply that $E(r_t^2)$ is not constant over time.

This chapter considers the Minimum-Distance Estimator of α and β when there are structural changes in μ which are ignored when the model (6.1) and (6.3) is fitted to the data. No matter which way the process changes, it is easily seen that any such change will in general increase empirical autocorrelations of the ϵ_t^2. For illustration, figure 6.2 depicts the first 16 empirical autocorrelations computed from $n = 4000$ observations, for a stationary MA(2) process

$$r_t = m + \epsilon_t + 0.5\epsilon_{t-1} + 0.5\epsilon_{t-2}, \epsilon_t \sim nid(0,1), \tag{6.5}$$

where m switches from -t to t in the middle of the sample. Without such change in m, the theoretical autocorrelations are $\rho_1 = 0.5$, $\rho_2 = 0.17$, $\rho_3 = \rho_4 = ... = 0$. As the figure shows, estimated correlations are much larger and tend to one as t increases.

The same effect is found for the GARCH(1,1) process with parameters $\alpha = 0.2$, $\beta = 0.4$ and $\omega = 0.001$ where μ successively takes the values 0, 0.25 and 0.5 (see figure 6.3 where entries are averages of 1000 replications). For the GARCH squared process with the same parameters, which itself is an ARMA(1,1) process, figure 6.4 shows the empirical autocorrelations when the process undergoes structural shifts in the mean μ. The structural change in the mean of the process inducing an increase of the empirical autocorrelations holds for all types of stochastic processes.

Let in general r_t $(t = 1, ..., n)$ be any short memory sequence of random variables with bounded variance and k shifts in mean at $1 < n_1 < ... < n_k < n$, and consider the empirical g'th order autocorrelation coefficient

$$\hat{\rho}_g = \frac{\sum_{t=1}^{n-g}(r_t - \bar{r})(r_{t+g} - \bar{r})}{\sum_{t=1}^{n}(r_t - \bar{r})^2} \ . \tag{6.6}$$

Rewriting the numerator as

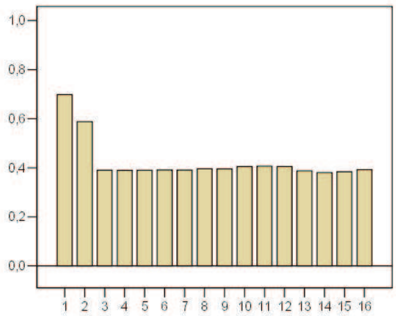
$$\sum_{t=1}^{n-g}(r_t - \bar{r})(r_{t+g} - \bar{r}) = \sum_{t=1}^{n}(r_t - \bar{r})^2 \tag{6.7}$$

$$- \sum_{t=n-g+1}^{n} (r_t - \bar{r})^2 + \sum_{t=1}^{n-g}(r_t - \bar{r})(r_{t+g} - r_t) \ , \tag{6.8}$$

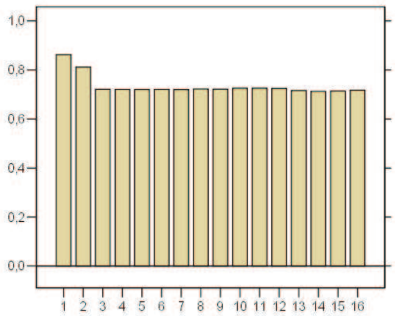
we see that

$$\hat{\rho}_g = 1 - \frac{\sum_{t=n-g+1}^{n}(r_t - \bar{x})^2}{\sum_{t=1}^{n}(r_t - \bar{r})^2} + \frac{\sum_{t=1}^{n-g}(r_t - \bar{r})(r_{t+g} - r_t)}{\sum_{t=1}^{n}(r_t - \bar{r})^2} \ , \tag{6.9}$$

a) $t = 1$



b) t = 2



c) t = 4

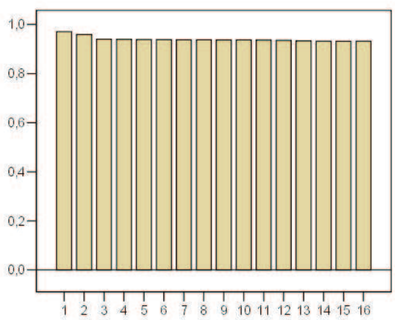


Figure 6.2: Empirical autocorrelations of the MA(2) process with a shift in expectation

where the last two terms can be made as close to 0 as desired if n is "large" relative to k and $\sum(r_t - \bar{r})^2 \xrightarrow{P} \infty$. This is so because the first term tends to zero has $g/n \to 0$, and the second term tends to zero in view of the fact that $(r_t - \bar{r})(r_{t+g} - r_t)$ is "small" relative to $(r_t - \bar{r})^2$ whenever r_{t+g} and r_t belong to the same regime. When the number of shifts is small relative to sample size, this will apply to an increasing number of terms in the sum, so the ratio becomes arbitrarily small.

This reasoning is of course purely heuristic, but suffices for the purposes of the present investigation, which is not concerned with the particular mechanics which lead to $\hat{\rho}_i \xrightarrow{p} 1$. Rather, we take this limiting behavior as given and explore its implications for the estimated persistence of the data.

6.5 Estimating Persistence

The Minimum-Distance-Estimator (MDE) of α and β was explained in chapter 1. Its efficiency relative to the Maximum Likelihood estimator is evaluated in detail in Baillie and Chung (2001); it depends on the particular choice of g and W and shall not concern us here. Rather, we take g and W as given and consider the behavior of $\hat{\delta} = \hat{\alpha} + \hat{\beta}$ as the size of structural changes increases.

This behavior in turn depends crucially on the fact that, in view of section 6.4,

$$\hat{\rho} = (\hat{\rho}_1, \dots, \hat{\rho}_g)' \xrightarrow{p} e := (1, \dots, 1)' \tag{6.10}$$

as the structural changes become larger. This implies that

$$\arg\min_{\alpha,\beta}[plim\ \hat{\rho} - \rho(\alpha,\beta)]'W[plim\ \hat{\rho} - \rho(\alpha,\beta)] \tag{6.11}$$

$$\subseteq \arg\min_{\alpha,\beta}[e - \rho(\alpha,\beta)]'W[e - \rho(\alpha,\beta)], \tag{6.12}$$

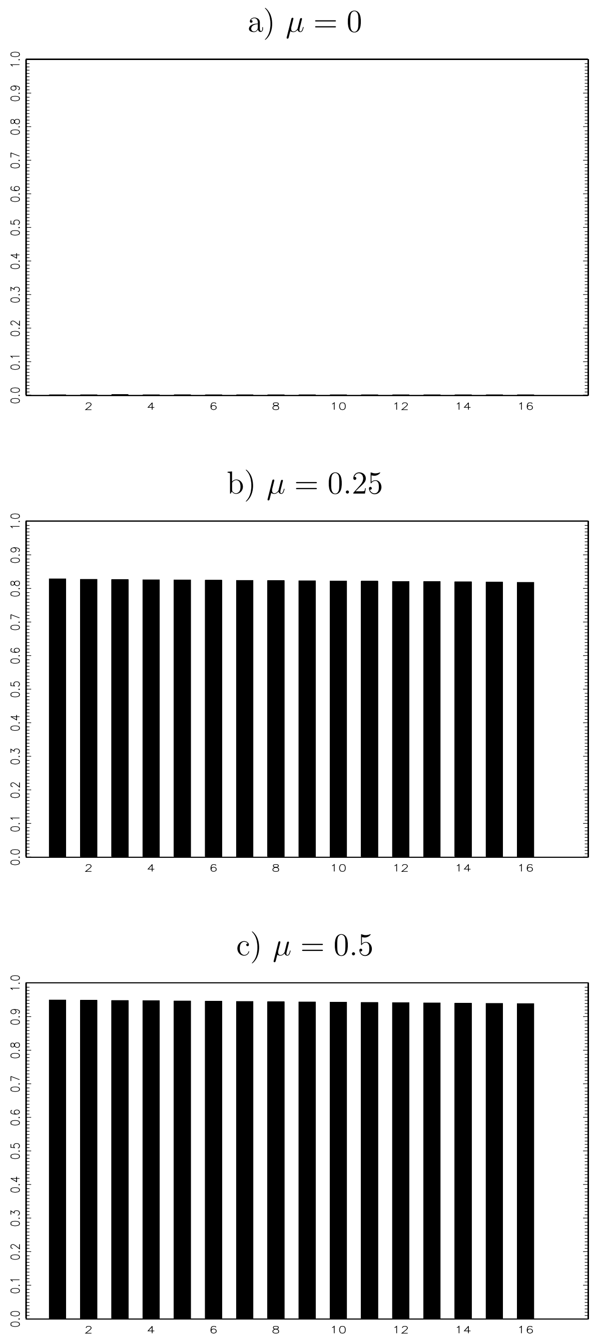


FIGURE 6.3: EMPIRICAL AUTOCORRELATIONS WITH A SHIFT IN EXPECTATION FOR THE GARCH(1,1) PROCESS WITH $\alpha = 0.20$ $\beta = 0.40$, $\omega = 0.001$ SAMPLE SIZE OF 4000.

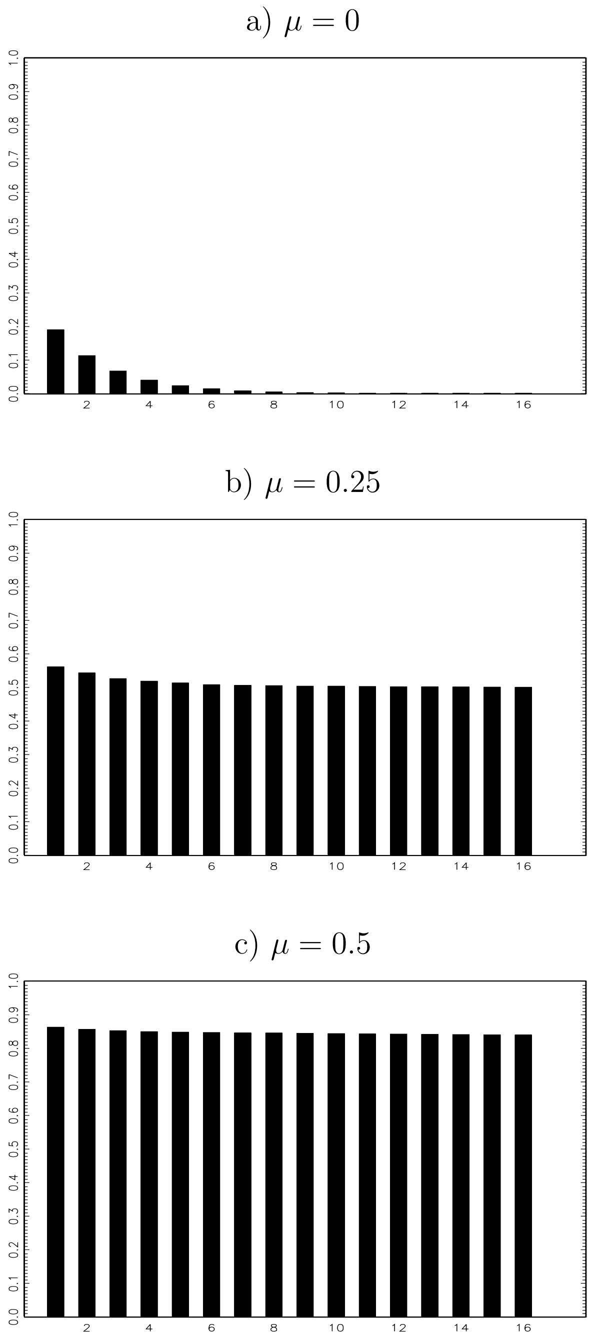


Figure 6.4: Empirical autocorrelations with a shift in expectation for the GARCH(1,1) squared process with $\alpha = 0.20$ $\beta = 0.40$, $\omega = 0.001$ sample size of 4000.

where the latter set of minimizing values of α and β is in view of[1] (3.18) determined by

$$\alpha + \beta = 1 \quad \text{and} \tag{6.13}$$

$$\alpha + \frac{\alpha^2 \beta}{1 - 2\alpha\beta - \beta^2} = 1. \tag{6.14}$$

This is so because (6.13) and (6.14) are equivalent to $\rho(\alpha, \beta) = e$, which is equivalent to

$$[e - \rho(\alpha, \beta)]' W [e - \rho(\alpha, \beta)] = 0, \tag{6.15}$$

which in view of the positive definiteness of W is the minimum value which can be attained.

It is easily checked that (6.13) implies (6.14), so all pairs of α and β with $\alpha > 0$, $\beta \geq 0$ and $\alpha + \beta = 1$ are candidates for $plim_{\hat{\rho} \to e}(\hat{\alpha}, \hat{\beta})$. Which one of these will eventually materialize depends on the particular way in which $\hat{\rho}$ approaches e. In practise, it appears that small values of $\hat{\alpha}$ and large values of $\hat{\beta}$ are preferred (see e.g. Haas, Mittnik, and Paollela (2004a), figure 1). The point of interest here is that, no matter what the particular probability limits of $\hat{\alpha}$ and $\hat{\beta}$ are, they must always sum to one.

6.6 Some Finite Sample Simulations

This section reports on various Monte Carlo simulations to check the finite sample relevance of the above result. Table 6.1 summarizes the experiments. We use various sample sizes and magnitudes of the structural shift in μ as defined in (6.1) and (6.3). The number of lags for the minimum distance

[1](3.18) is $\rho_g = (\alpha + \frac{\alpha^2 \beta}{1 - 2\alpha\beta - \beta^2})(\alpha + \beta)^{g-1}, \quad g \geq 2.$

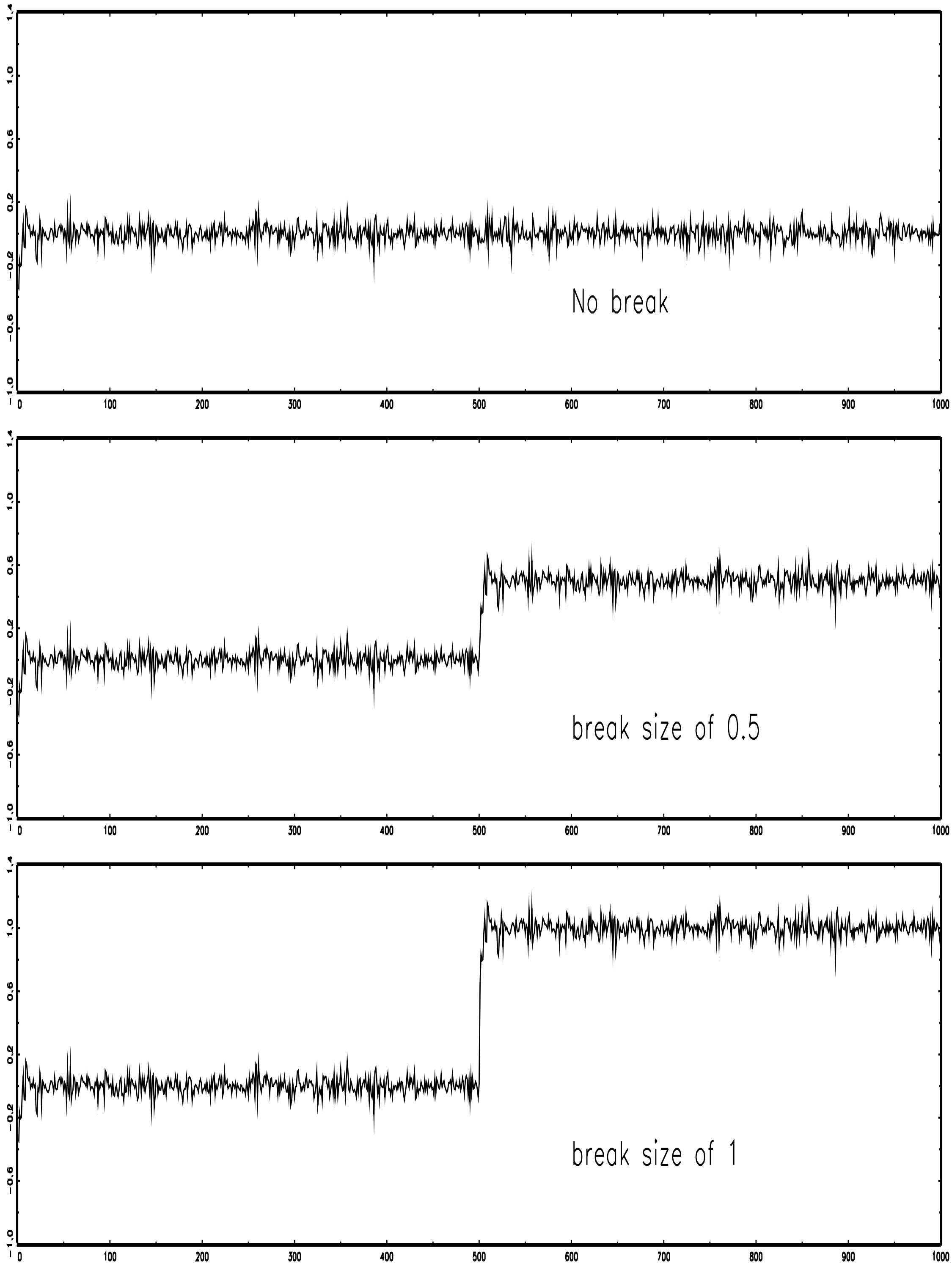


FIGURE 6.5: PATHS OF THE GARCH(1,1) PROCESS WITH A DETERMINISTIC SHIFT IN THE MEAN μ (TYPE I) WITH $\alpha = 0.2,\ \beta = 0.4,\ \mu_1 = 0,\ \mu_2 = 0.5,\ \mu_3 = 1$.

estimator is fixed at $g = 10$, and the weighting matrix W is obtained via the Newey West procedure; results remain virtually unchanged for different lags and weighting matrices. The innovations η_t are standard normal, and the shift always occurs in the middle of the sample. $N = 1000$ experiments are performed for any given parameter and sample size combination. The value for ω is 0.001, uniformly across experiments. In this particular experiment, the size of the shift occurring in μ will be increasing all the time at each step by 0.1 and starting from 0 till 1. We labeled these types of breaks as "type I". Figure 6.5 shows the paths of GARCH the time series with a deterministic shift in the mean, this break always occurring the the middle of the sample. Figure 6.6 plots the estimated persistence as a function of the size of the break in the case where $\alpha = 0.2$ and $\beta = 0.4$. Figures 10.17 and 10.18 in the appendix correspond to the cases $(\alpha, \beta) = (0.3, 0.3)$ and $(\alpha, \beta) = (0.4, 0.2)$.

Another type of breaks, labeled "type II" is pictured in figure 6.7. These are constructed by assuming that the constant term of the conditional mean equation μ is non zero as opposed to the previous types of breaks where one starts with a time series with a zero mean. As soon as the μ of the time series is non zero, we then can multiply the second half of the time series by a sufficiently large constant[2] to shift it up, obtaining a figure similar to figure 6.5. We also find here that the estimated persistence is an increasing function of the size of the break. Interestingly, it can be made arbitrarily close to one. Table[3] 6.2 reports the results of the simulations.

---

[2] We denote this constant by c.

[3] In this table, c is the break coefficient which represents the multiplicative factor of the second half of the time series.

T	$\Delta\mu$										
	0	0.1	0.2	0.3	0.4	0.5	0.6	0.7	0.8	0.9	1
					a) $\alpha = 0.2$, $\beta = 0.4$						
1000	0.556	0.715	0.868	0.940	0.959	0.974	0.989	0.992	0.994	0.997	0.998
2000	0.557	0.774	0.904	0.956	0.977	0.993	0.996	0.996	0.997	0.998	0.998
4000	0.562	0.824	0.934	0.967	0.988	0.995	0.995	0.998	0.999	0.999	0.999
					b) $\alpha = 0.4$, $\beta = 0.2$						
1000	0.545	0.786	0.817	0.903	0.953	0.971	0.984	0.989	0.993	0.995	0.995
2000	0.546	0.739	0.881	0.945	0.971	0.987	0.994	0.996	0.997	0.998	0.998
4000	0.556	0.776	0.926	0.962	0.983	0.992	0.996	0.997	0.998	0.998	0.999
					c) $\alpha = 0.3$, $\beta = 0.3$						
1000	0.538	0.655	0.816	0.914	0.934	0.960	0.980	0.984	0.988	0.989	0.993
2000	0.541	0.710	0.866	0.931	0.952	0.982	0.990	0.991	0.994	0.995	0.998
4000	0.566	0.766	0.898	0.958	0.978	0.986	0.990	0.996	0.997	0.998	0.999

Table 6.1: Estimated δ's (averaged over 1000 runs) for various sizes of structural shifts in μ – Type I

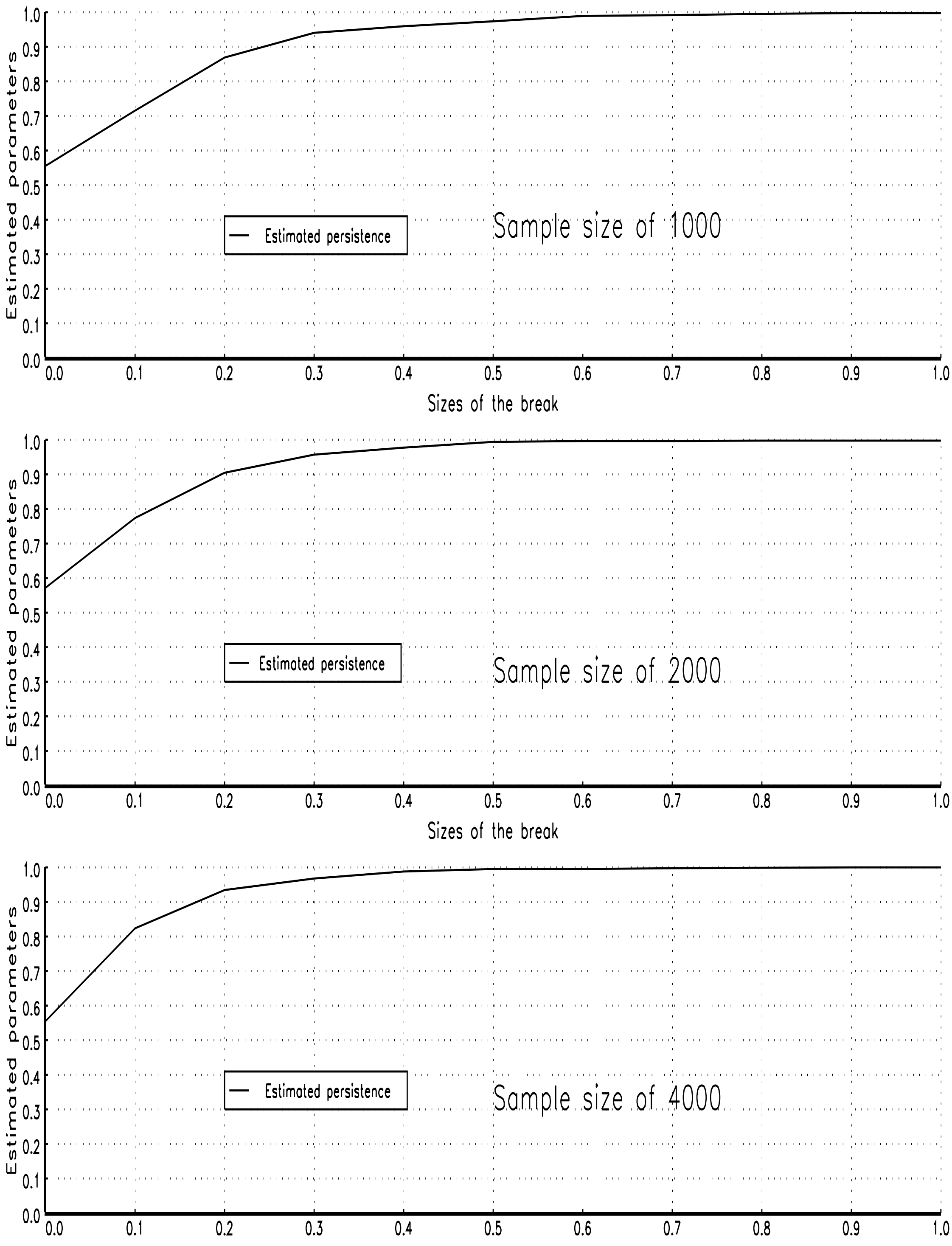


FIGURE 6.6: ESTIMATED PERSISTENCE AS A FUNCTION OF THE SIZE OF THE BREAK TYPE I ($\alpha = 0.20,\ \beta = 0.40$ AND $\omega = 0.001$)

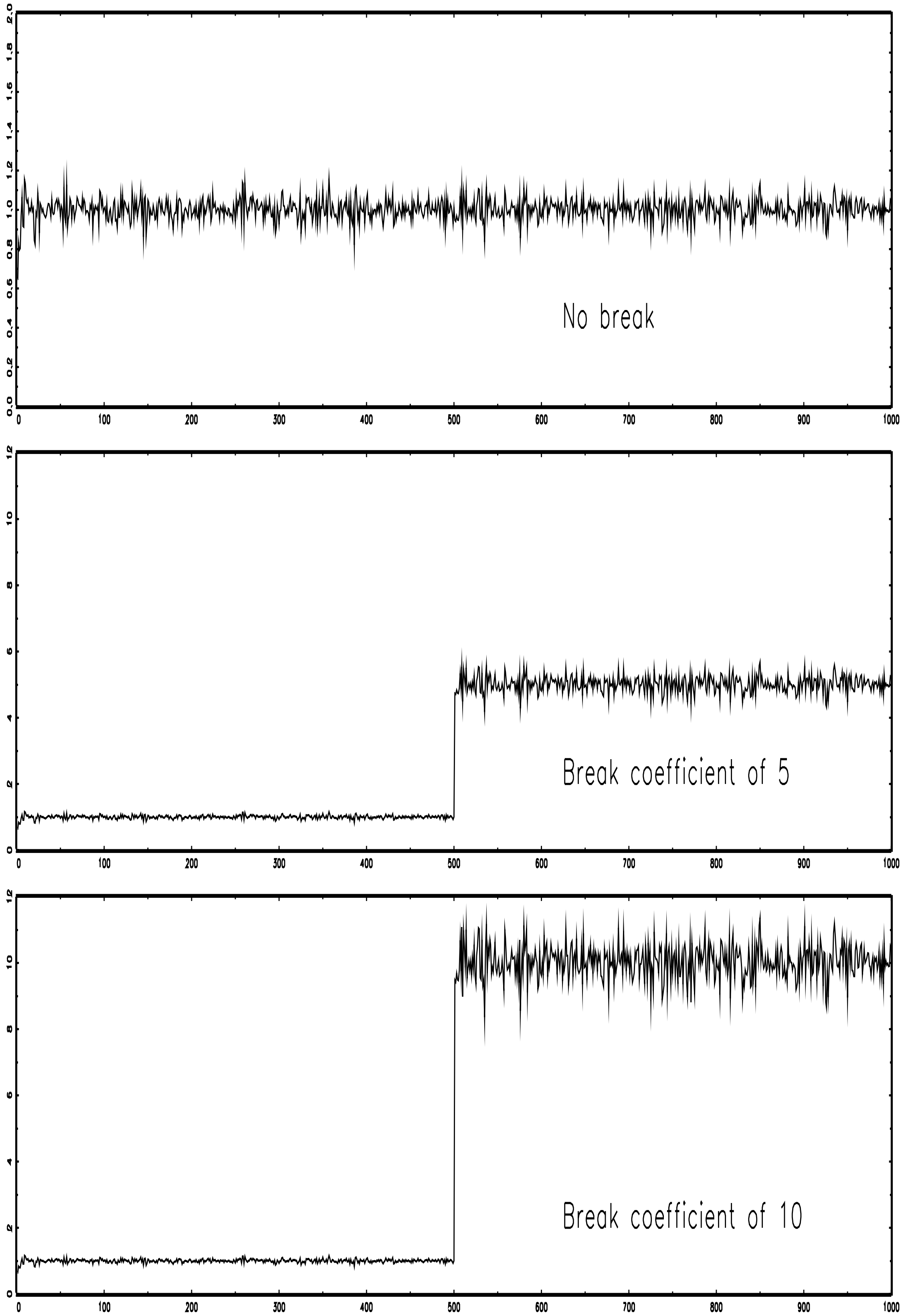


FIGURE 6.7: PATHS OF THE GARCH PROCESS WITH A DETERMINISTIC SHIFT (TYPE II) $\alpha = 0.2,\ \beta = 0.4\ \mu = 1$ WITH $c_1 = 1,\ c_2 = 5$ AND $c_3 = 10$

T	c						
	1	1.1	1.2	1.3	1.4	1.5	2
			a) $\alpha = 0.2$, $\beta = 0.4$				
1000	0.556	0.971	0.982	0.993	0.994	0.997	0.997
2000	0.557	0.984	0.993	0.996	0.997	0.998	0.998
4000	0.559	0.995	0.996	0.998	0.999	0.999	0.999
			a) $\alpha = 0.3$, $\beta = 0.3$				
1000	0.545	0.961	0.984	0.986	0.993	0.996	0.997
2000	0.546	0.983	0.0.991	0.994	0.996	0.997	0.998
4000	0.556	0.989	0.997	0.998	0.998	0.999	0.999
			a) $\alpha = 0.4$, $\beta = 0.2$				
1000	0.538	0.958	0.977	0.983	0.993	0.994	0.996
2000	0.541	0.958	0.978	0.988	0.993	0.994	0.996
4000	0.566	0.988	0.995	0.997	0.998	0.998	0.999

Table 6.2: Estimated δ's (averaged over 1000 runs) for various sizes of structural breaks – Type II

6.7 Conclusion

This chapter has considered the Minimum Distance Estimators of α and β suggested by Baillie and Chung (2001) and has showed that the estimated persistence can arbitrarily be made close to one if there are deterministic structural changes in unconditional expectation μ. In the finite sample assessment of this result, the first type of structural breaks (labeled "Type I") is concerned with a shift of μ in the middle of the time series. Another type of break (labeled "Type II") is concerned with multiplying the second half of the time series with a constant, provided that the μ is a non zero constant. In both cases, we reach the conclusion that the estimated persistence can be made arbitrarily close to one.

Our finding, confirmed by the Monte Carlo study is the special case in which the break always occurred in the middle of the sample. In this framework, we argue that this type of break increase the autocorrelations of the squared process, therefore increasing the estimated persistence. This is a sufficient condition for the persistence parameter to become arbitrarily close to one.

Chapter 7

Additional origins of high persistence in GARCH-models

7.1 Introduction

Modeling the conditional mean of macroeconomic and financial time series has highlighted the role of persistence of shocks. Tests for unit roots in the univariate representation of time series were being performed as opposed to simply considering the time series to be stationary. A simple classification of processes being $I(0)$ or $I(1)$ is far too restrictive. An I(0) process is stationary and an $I(1)$ process contains a unit root.

As opposed to $I(0)$ processes where the propagation of shocks to the mean decays exponentially and to the $I(1)$ processes characterized by infinite persistence, Adenstedt (1974), Granger (1980), Granger (1981) and Granger and Joyeux (1980) proposed already around the 80's, families of discrete time stochastic processes in which the propagation of the shocks to the mean were neither $I(0)$ nor $I(1)$. They called this type of processes long memory fractionally integrated processes (see section 2).

Recent studies by Breit, Crato, and Lima (1998), Dacorogna, Muller, R.J, Olsen, and Pictet (1993), Harvey (1993) and Ding, Granger, and Engle (1993) have reported the presence of apparent long memory (to be define in the next section) in the autocorrelations of squared or absolute returns of various financial asset prices. Baille, Bollerslev, and Mikkelsen (1996) introduced therefore the Fractionally Integrated Generalized Autoregressive Conditional Heteroskedaticity or FIGARCH to explain and represent these types of phenomenon in financial market volatility.

The estimated persistence in various types of GARCH - models is known to be too large when the parameters of the model undergo structural changes somewhere in the sample. The present chapter argues that one avenue through which this could happen is apparent long memory in the squares of ϵ_t and in ϵ_t itself. We also show that a particular estimator of $\alpha + \beta = \delta$ must by necessity tend to one if this artificial long memory resembles an I(d)-process with $d > 1/2$. The particular mechanism of interest here which induces this I(d)-behavior is a structural change in the unconditional mean μ. Previous research on structural changes in GARCH models was mainly concerned with changes in the GARCH parameters α, β and ω (Francq, Zakoian, and Roussignol (2001), Haas, Mittnik, and Paollela (2004b) among many others) or in the distribution of the innovations η_t (Dueker (1997)). Here we follow Diebold and Inoue (2001) by taking the distribution of the ϵ_t's as given, and consider changes in the mean μ. Extending results from the previous chapter (these are as well found in Krämer and Tameze (2007)), where the focus is on increasing structural changes when the sample size is fixed, we here consider structural changes in the context of increasing samples, in the vain of Hillebrand (2005). Hillebrand (2005) shows that the ML-estimates of δ will tend to one when the number of structural changes remains finite as sample size increases. We consider again the Minimum Distance Estimator (MDE) of α and β suggested by Baillie and

Chung (2001), and show that the sum of the estimated α and β can likewise be made arbitrarily close to one if there are certain types of structural change in the r_t-process.

7.2 Definitions of Long Memory

This section considers several definitions of long memory and linkages among them. Let recall that a process r_t is said to be (fractionally)integrated of order $d,\ 0 < d < 1$ or $I(d)$, if

$$(1-L)^d r_t = u_t, \tag{7.1}$$

where L is the lag operator, and u_t is a stationary and ergodic process with a bounded and positively valued spectrum at all frequencies.

Usually, long memory can be defined in the time domain in terms of decay rates of the long lags autocorrelations, or in the frequency domain in terms of rates of explosion of low frequency spectra or alternatively in terms of the rate of growth of the variance of the partial sums.

A long lag definition of long memory for a covariance stationary process r_t is

$$\rho_{r_t}(\tau) \sim k\tau^{2d-1} \ \text{ as } \tau \longrightarrow \ \infty. \tag{7.2}$$

Alternatively, Mcleod and Hipel (1978) say r_t possesses long memory if

$$\lim_{n \longrightarrow \ \infty} \ \Sigma_{j=-n}^{n} \mid \rho_j \mid = \ \infty. \tag{7.3}$$

A low frequency definition spectral definition of long memory is

$$f_{r_t}(\omega) \sim g\omega^{-2d} \ \text{ as } \ \omega \longrightarrow \ 0^+. \tag{7.4}$$

Heyde and Yang (1997) gives an even more general low-frequency spectral definition of long memory as

$$f_{r_t}(\omega) \longrightarrow \infty \ \text{ as } \ \omega \longrightarrow \ 0^+. \tag{7.5}$$

A third definition of long memory which will be used in this chapter involves the rate of growth of variances of partial sums. We define

$$S_T = \sum_{t=1}^{T} r_t. \tag{7.6}$$

The r_t display long memory if

$$\mathrm{Var}(S_T) = \mathrm{O}(T^{2d+1}). \tag{7.7}$$

Beran (1994) chapter 2, proves that the first two definitions are equivalent under specific hypotheses. There is a tight connection between the variance of partial sums definition of long memory and the spectral and autocorrelation definitions of long memory as discussed in Diebold and Inoue (2001). The spectral density at zero is the limit of $\frac{1}{T}S_T$ so A covariance stationary process has long memory in the generalized spectral sense of Heyde and Yang (1997) if and only if it has long memory for some $d > 0$ in the variance of partial sum sense. Barndorff-Nielsen and Cox (1989) present more insights into this result.

Fractionally integrated processes constitute a special family of long memory processes. In this work, we have been focussing on GARCH models which are special type of stochastic processes used in forecasting volatility. In these models, the variance rate follows a mean-reverting process. Our interest in long memory processes won't center on unit root processes but instead, on mean reverting fractionally integrated processes $I(d),\ 0 < d < 1$ and long memory in our sense will be meaning $I(d),\ 0 < d < 1$.

7.3 Stochastic Structural Change in the Mean and Sample Autocorrelations

Once more, our approach here considers the Minimum Distance Estimator of α and β when there are structural changes in the unconditional expectation μ which are ignored when the model is fitted to the data. These changes can be stochastic in the mean, for instance, by letting μ depend on the state of an independent Markov process Δ_t:

$$r_t := \mu(\Delta_t) + \epsilon_t. \tag{7.8}$$

This is similar to the model described in (6.4) extensively considered in the literature. The difference in (7.8) is that it is the conditional mean and not the conditional variance of r_t that is affected. No matter in which way the process changes, however, any such change will in general increase the empirical autocorrelations of the r_t^2.

In this chapter, we show that the sum of the estimated α and β can likewise be made arbitrarily close to one when there are certain types of structural changes in the expectation μ of the r_t-process, or more generally, when the ${r_t}^2$-process behaves as if it had nonstationary long memory.

Krämer and Tameze (2007) show that the estimated persistence will become undistinguishable from 1 for any given sample size as the size of the structural change increases. Next, we investigate yet another avenue through which empirical autocorrelations may be led to tend to one. This happens for increasing sample size, when the r_t can be made to behave as if they were I(d) with $d \geq \frac{1}{2}$, where, following Diebold and Inoue (2001), I(d) behavior is defined by

$$Var\left(\sum_{t=1}^{T} r_t \right) \;=\; O\left(T^{2d+1} \right). \tag{7.9}$$

It has long been known (see e.g. Krämer (1985)) that for $d = 1$, empirical autocorrelations of r_t of all orders must tend to one in probability as $T \longrightarrow \infty$, and Hassler (1997) shows that this holds for fractional integration parameters with $\frac{1}{2} \leq d < 1$ as well. The intuition behind this is that the last two terms in the following expression already derived in the previous chapter

$$\hat{\rho}_g = 1 - \frac{\sum_{t=n-g+1}^{n}(r_t - \bar{r})^2}{\sum_{t=1}^{n}(r_t - \bar{r})^2} + \frac{\sum_{t=1}^{n-g}(r_t - \bar{r})(r_{t+g} - r_t)}{\sum_{t=1}^{n}(r_t - \bar{r})^2}, \tag{7.10}$$

become arbitrarily small as $T \longrightarrow \infty$ as the numerators are of smaller orders in probability than the denominators.

Diebold and Inoue (2001) show that behavior of type (7.9) occurs for instance whenever μ_t is stochastic and independent of ε_t and displays structural breaks of the form

$$\mu_t = \mu_{t-1} + \nu_t \tag{7.11}$$

$$(7.12)$$

$$\nu_t = \begin{cases} 0 & \text{with probability} \quad 1-p \\ \omega_t & \text{with probability} \quad p, \end{cases}$$

where $\omega_t = i.i.d.(0, \sigma^2)$, and where p may depend on sample size. Since

$$\sum_{t=1}^{T} \mu_t = T v_1 + (T-1) v_2 + ... + v_T, \tag{7.13}$$

we have

$$Var\big(\sum_{t=1}^{T} \mu_t\big) = p\ \sigma^2 \sum_{t=1}^{T} t^2 = p\ \sigma^2 \frac{T(T+1)(2T+1)}{6}\ . \tag{7.14}$$

So, we can have (7.9) for any $d,\ 0 < d < 1$, by letting

$$p = c\frac{1}{T^{2-2d}} \quad (0 < c \leq 1). \tag{7.15}$$

Of course, in the limiting case where $d = 1$ and p does not depend on T, μ_t and therefore also r_t will be I(1) and long memory will be extreme.

Spurious long memory in r_t can also be induced by time varying staying probabilities

$$p_{00} \quad = \quad 1 - c_0 T^{-\delta_0} \tag{7.16}$$

$$p_{11} \quad = \quad 1 - c_1 T^{-\delta_1} \tag{7.17}$$

in the Markov-switching model (7.8) with two states and serially independent ϵ's. Diebold and Inoue (2001) show that then (7.9) applies with

$$d = \frac{1}{2} max\{min(\delta_0, \delta_1) - |\delta_0 - \delta_1|, 0\}, \tag{7.18}$$

and to the extent that this carries over to the case where the ϵ_t's follow a GARCH-process, we will for $d_0 = d_1 = 1$ again have empirical autocorrelations of the r_t which tend to one as a consequence of structural change.

We will not enter into a detailed discussion of this phenomenon here. There might well be many other instances where this tendency towards unity of empirical autocorrelations occurs. Diebold and Inoue (2001) for instance show that the Engle and Smith (1999)–STOP-BREAK model, which generates an I(1)-series, can be generalized to an arbitrary I(d)-behavior where in all cases we have autocorrelations increasing with sample size. For the present purpose, it suffices to know that there do exist meaningful models which induce empirical autocorrelations of a time series to become large. The conditions that guarantee this to happen do not concern us here. Rather, we take this behavior as given and explore its implications for the estimated persistence of a GARCH(1,1)-model.

To that purpose, it remains to show that real or spurious long memory in the r_t's induces real or spurious long memory in the r_t^2 (since the estimator which we consider in this chapter is based on the empirical autocorrelations of the squared observations). For a given sample size and increasing breaks, it is easily seen that the arguments that lead to increasing autocorrelations of r_t also lead to increasing autocorrelations of r_t^2. For "genuine" Gaussian I(d)-processes with $d \geq \frac{1}{2}$, Dittmann and Granger (2002) show that the squared process is also I(d) with the same d, and similar results hold for spurious long memory as well (in the sense that convergence to one of the empirical autocorrelations of the r_t's implies convergence to one of the empirical autocorrelations of the

r_t^2's). For instance, it is easily seen that with μ's changing according to (7.11), the empirical autocorrelation of both the r_t's and the r_t^2's must tend to one as sample size increases. Again, we do not want to enter into a detailed discussion here, as the mechanisms that produce large autocorrelations of the ${r_t}^2$ are not our main concern. Rather, our interest is focused on the consequences which this might have for the estimated persistence $\hat{\alpha} + \hat{\beta} = \hat{\delta}$.

7.4 Some finite sample simulations

This section reports on various Monte Carlo simulations to check the finite sample relevance of the above results. In a first series of experiments, we keep the number of changes fixed at times $[Td_1], [Td_2], ..., [Td_k]$ where $0 < d_1 < d_2 < ...d_k < 1$, along the lines of Hillebrand (2005).

Figure 7.1 reports the first 16 empirical autocorrelations in a GARCH(1,1)-squared process with $\alpha = 0.2,\ \beta = 0.4$ where $k = 1,\ d_1 = 1/2$, where a shift in μ of size 0.8 occurs. The figures are averages over 1000 replications and show that empirical autocorrelations tend to one quite rapidly as sample size increases.

Figure 7.2 shows the resulting estimates of $\hat{\delta} = \hat{\alpha} + \hat{\beta}$, also for a wider range of sample sizes and structural breaks. It is seen that the estimated persistence likewise tends to one quite rapidly as the sample size increases, at least if the structural change is large enough. Similar results were also obtained for other values of $k, d_1, d_2, ..., d_k$ and α and β and can be obtained from the authors upon request.

In a second series of experiments, we let μ change according to the Diebold and Inoue (2001)–scheme from equation (7.11). Figures 7.3, 7.4 and 7.5 show

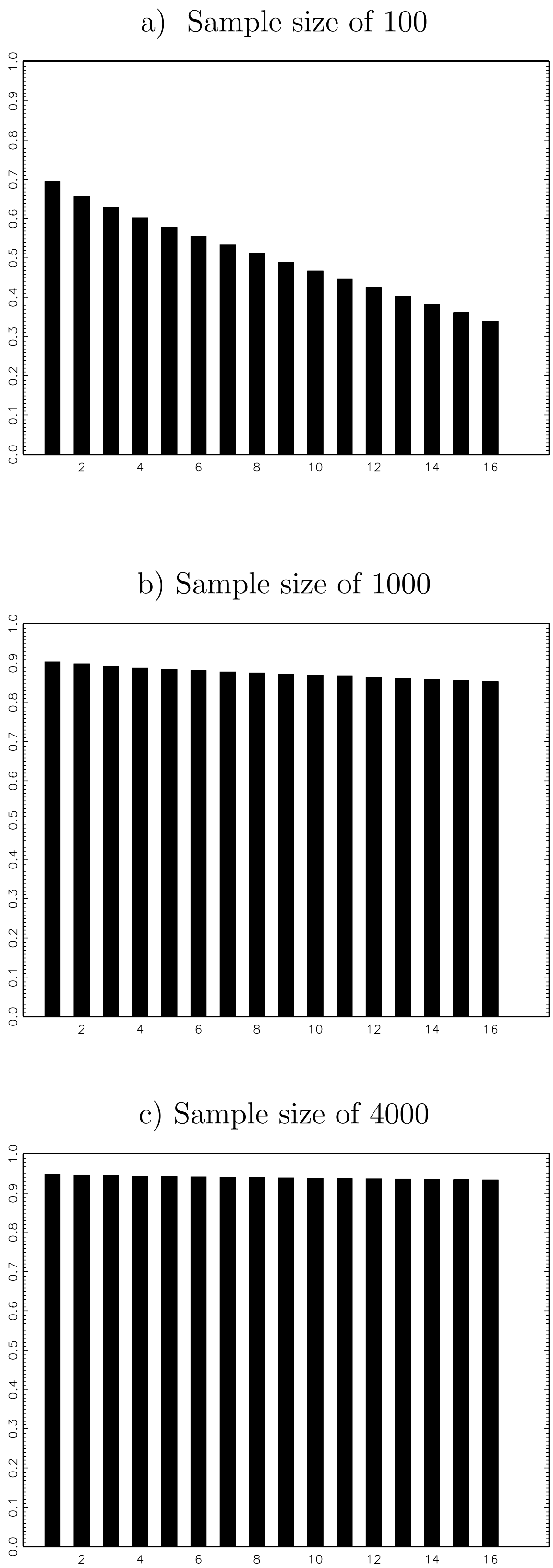


FIGURE 7.1: EMPIRICAL AUTOCORRELATIONS WITH A SHIFT OF 0.8 IN μ FOR THE GARCH (1,1) SQUARED PROCESS $\alpha = 0.2, \quad \beta = 0.4$

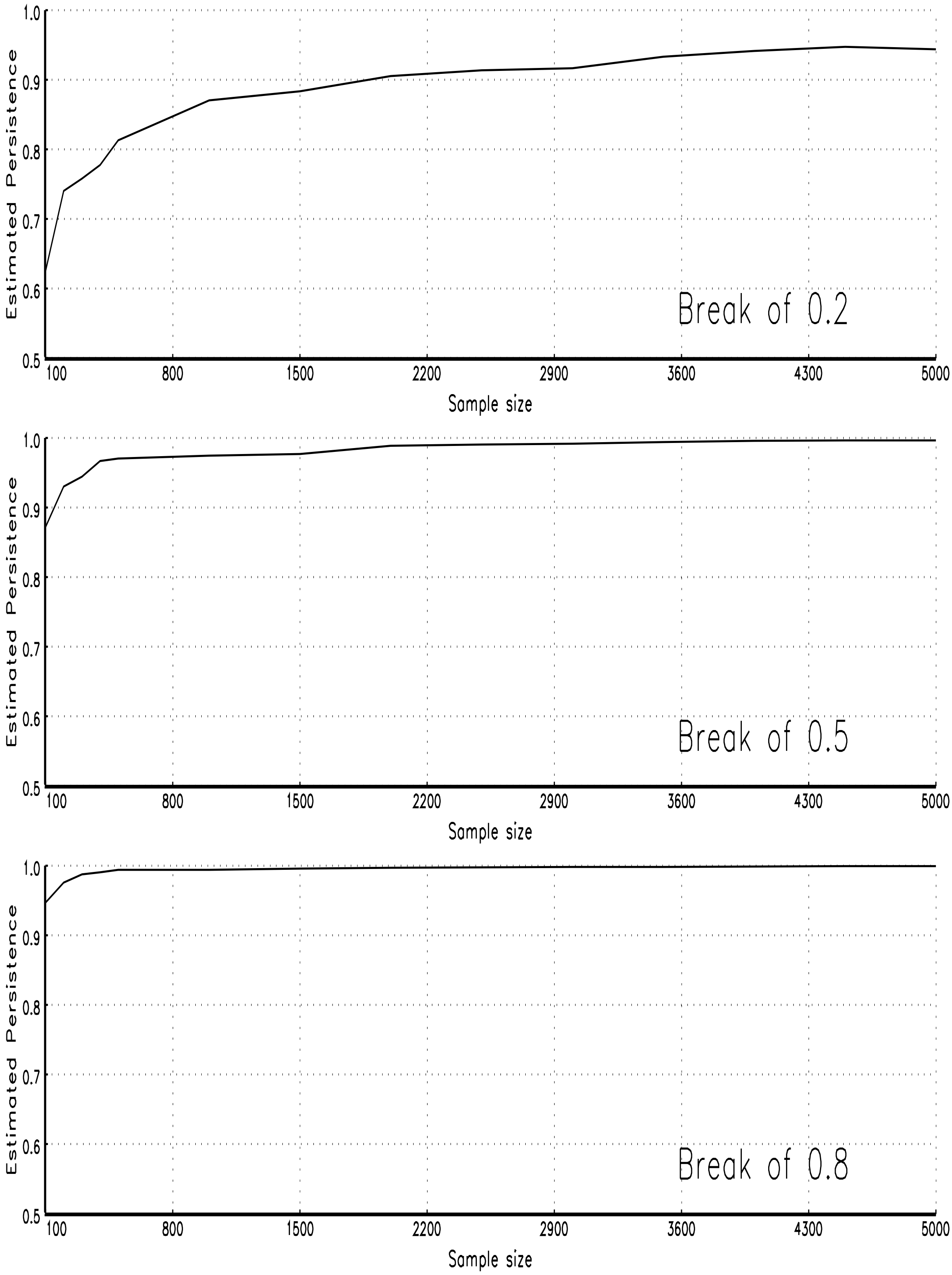


FIGURE 7.2: ESTIMATED PERSISTENCE AS A FUNCTION OF THE SAMPLE SIZE $(\alpha, \beta) = (0.2, 0.4)$

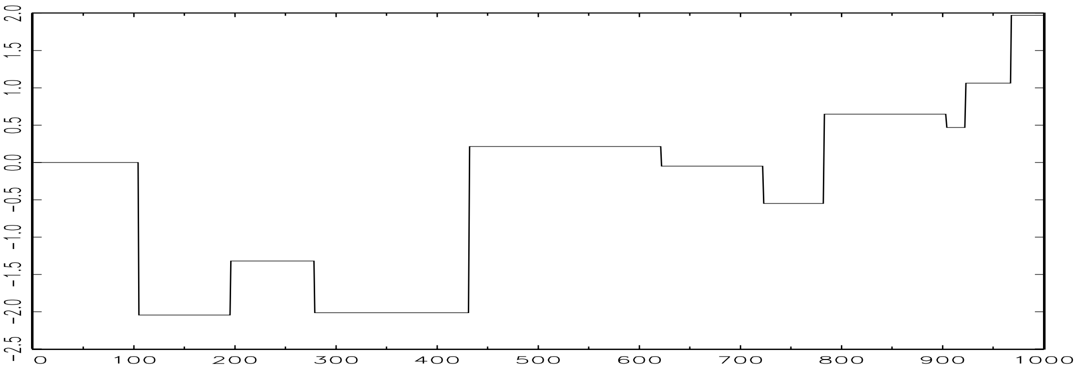


FIGURE 7.3: STOCHASTIC MEAN ACCORDING TO (7.11), P = 0.01 AND T = 1000

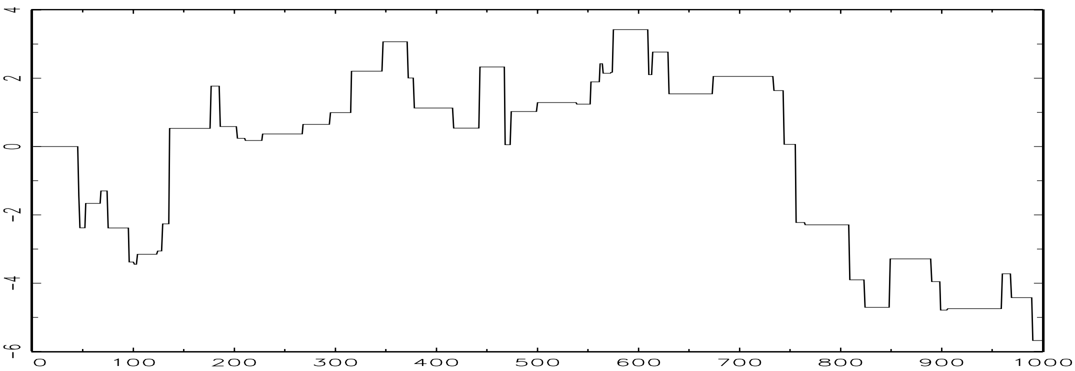


FIGURE 7.4: STOCHASTIC MEAN ACCORDING TO (7.11), P = 0.05 AND T = 1000

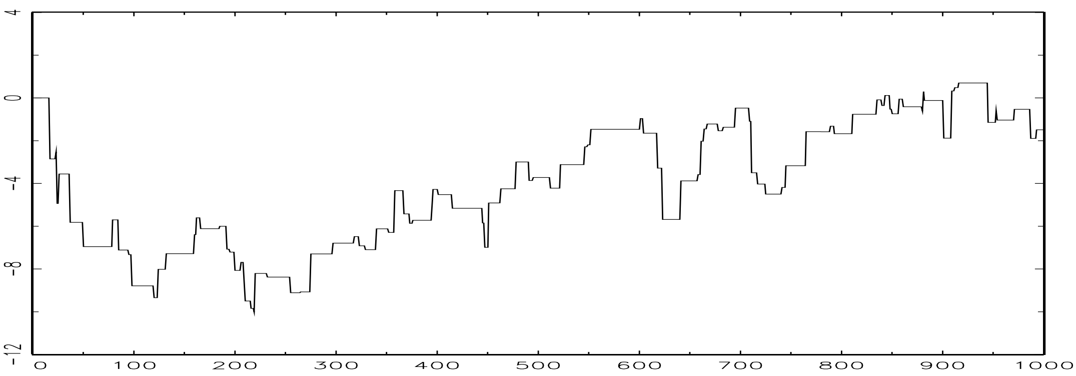


FIGURE 7.5: STOCHASTIC MEAN ACCORDING TO (7.11), P = 0.10 AND T = 1000

some sample time series of the μ's for some fixed sample sizes. In each case, we fix the switching probability and increase the sample size.

Figure 7.6 shows the resulting first 16 empirical autocorrelations of the ${r_t}^2$ for the case where $\nu_t \sim n.i.d(0, \sigma^2)$ and the switching probability is $p = 0.05$. It is seen that sample autocorrelations likewise tend to one as sample size increases, although not as fast as for the case where the structural change is non stochastic.

Figure 7.7 gives the persistence derived from these empirical autocorrelations for a wider range of switching probabilities and sample sizes. Again, it is seen that $\hat{\delta}$ approaches 1 quite rapidly, and similar results were obtained for different parameters of the GARCH-model as well.

7.5 Conclusion

The previous chapter dealt with structural changes in the mean μ in which the sample sizes were fixed and the size of the break was increasing. We find that this setting was sufficient to generate the IGARCH effect in the sense that the estimated persistence could arbitrarily made close to one by sufficiently increasing the size of the break. In this chapter, we show the same effect by keeping the size of the break fixed this time and increasing the sample size. This produces as well in an increase of the empirical autocorrelations of the GARCH process and squared process, which results in an increase of the estimated persistence. As the sample size grows, the persistence becomes undistinguishable from 1. Furthermore, we show that the estimated persistence can as well be made arbitrarily close to one if there are stochastic structural changes in the unconditional expectation μ, in particular in processes of the type (7.8). Our finding, confirmed by the Monte Carlo study is the special case in which the

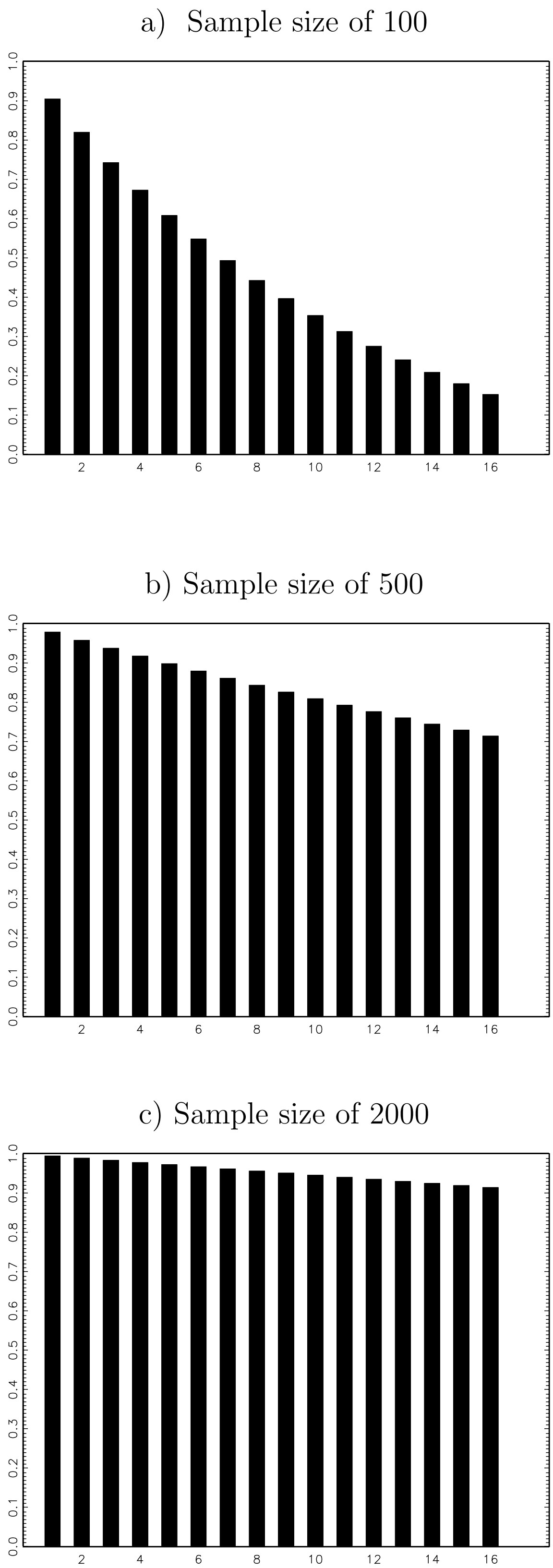


FIGURE 7.6: EMPIRICAL AUTOCORRELATIONS OF THE STOCHASTIC MEAN (AS DEFINED IN (7.11))WITH A SWITCHING PROBABILITY OF $p = 0.05$

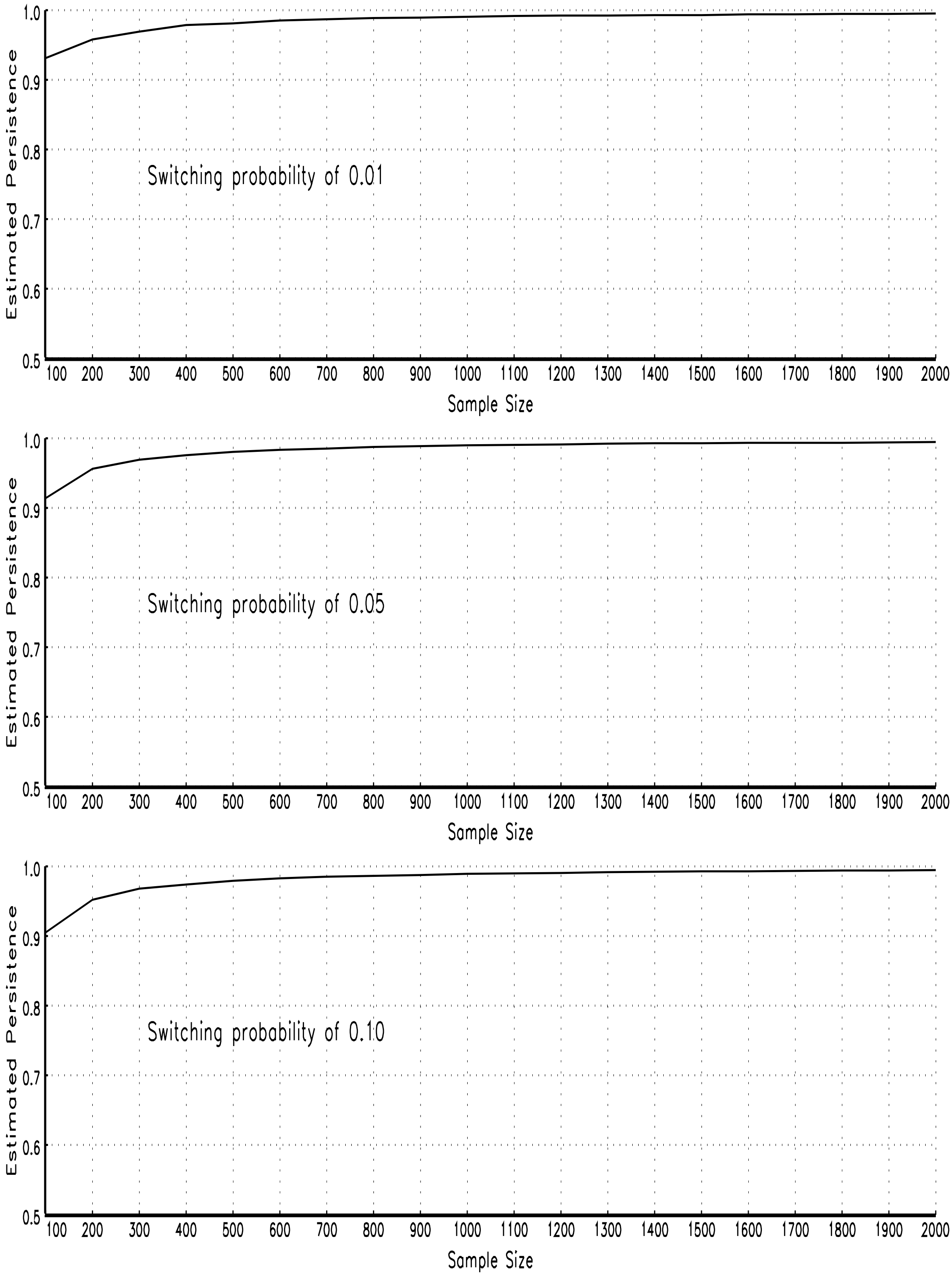


Figure 7.7: Estimated persistence as a function of the size with switching probabilities $p = 0.01,\ 0.05$ and 0.10 respectively

stochastic mean was generated according to (7.11). We argue that this type of break increases the autocorrelations of the GARCH process and squared process, therefore increasing the estimated persistence. In the previous chapter, we showed that a particular deterministic structural changes in μ was sufficient to produce the almost IGARCH behavior. This other particular deterministic structural change in μ and the stochastic structural change are two others sufficient conditions for the persistence parameter to become arbitrarily close to one.

Chapter 8

Value at Risk and Expected Tail Loss from Minimum Distance Estimation of GARCH(1,1)-models

8.1 Introduction

Hull (2006) defines a GARCH model as an econometric model to forecast risk where the variance follows a mean reverting process. In this chapter, we consider risks in financial markets. Unexpected market movements, often due to exogenous shocks, can cause stocks, interest rates, exchange rates and commodities to fluctuate enormously. Market risk is the risk an institution suffers from these changes. One of the most important tasks of financial institutions such as banks is to evaluate and manage market risk exposure. Stocks, interest rates, exchange rates and commodities are termed "derivatives". These derivatives are best managed by specific sensitivity risk measures called "Greek" letters. These Greeks describe different aspects of risk in a portfolio of derivatives. A financial institution usually calculates each of these measures each day for every market variable to which it is exposed. But even small banks sell lots of different products in many markets. So often, there are hundreds or even thousands of these markets variables and these lead to a large amount of different risk measures being produced everyday. These risk measures provide valuable information for a trader who is responsible for managing the part of the portfolio that is dependent on the particular market variable. They do not provide a way to measure the total risk to which a financial institution is exposed. Value at Risk (VaR) calculation is a process that provides a single number summarizing the total risk in a portfolio of financial assets. It has become widely used by corporate treasurers and fund managers as well as by financial institutions. Central bank regulators also use VaR in determining the capital required to reflect the market risk it is bearing. BASEL II is the latest international regulations for calculating bank capital. It is expected to come into effect in 2007, see e.g Hull (2007) for a discussion. In the context BASEL II for example, require an internal risk measurement model based on the 10 day VaR at a 99% confidence level.

In the Risk Management framework, VaR has emerged as the standard tool for measuring market risk. It provides a single number that quantifies the worst possible financial loss to a portfolio over a fixed time horizon and a given confidence level, see Jorion (2002) and Dowd (2005). This single estimate is of importance to managers as they need a number to quantify the possible loss. However as a risk measure VaR suffers from two drawbacks. First VaR is not a coherent risk measure(see section 8.2) as it is not sub-additive (to be defined in section 8.2). Next, it does not incorporate the happening of an extreme event as well as the size of the possible corresponding losses. An alternative simple measure extending the value at risk into a coherent risk measure and accounting for extremal events is the expected tail loss, see Embrechts, McNeil, and Frey (2005).

Forecasting the volatility using the history of the returns is intensively done in the ARCH and GARCH framework, mainly with maximum and quasi maximum likelihood methods. In this chapter, we take another approach, namely we forecast the needed volatility using the GARCH model in which the parameters are obtained by the minimum distance estimation as discussed in chapter 2. These minimum distance estimators are different at each lag. So the risk measure we will be calculating will depend on the number of lags used in the minimum distance estimator. We propose to compare two measures of market risk, namely Value at Risk (VaR) and Expected Tail Loss (ETL), both calculated as a function of the number of lags used in the minimum distance estimation framework with the same risk measures computed when the risk factors are forecasted under standard maximum likelihood approach. Value at Risk and expected tail loss are considered as risk measures so we start by discussing the properties of risk measures.

8.2 Properties of Risk Measures

Banks have to comply with some capital requirements under what is known as BASEL II. In this case, a risk measure is defined as the amount of cash that must be added to a position to make its risk acceptable to regulators.

Let ϖ be a risk measure, X_1 and X_2 be two risks, ϱ a positive real number τ and be a real number. Following Hull (2007), a risk measure ϖ is called coherent if it has the following properties:

- *Monotonicity*: $X_1 \leq X_2 \Rightarrow \varpi(X_2) \leq \varpi(X_1)$.

  If a portfolio has lower returns than another portfolio for every state of the world, its risk measure should be greater.

- *Translation invariance*: $\varpi(X_1 + \tau) = \varpi(X_1) - \tau$.

  If we add an amount of cash τ to a portfolio, it risk measure should go down by τ.

- *Homogeneity* : $\varpi(\varrho X_1) = \varrho(\varpi(X_1))$.

  Changing the size of the portfolio by a factor r while keeping the relative amounts of different items in the portfolio the same should result in the risk measure being multiplied by $\varrho\times$.

- *Subadditivity* : $\varpi(X_1 + X_2) \leq \varpi(X_1) + \varpi(X_2)$

  The risk measure for two portfolios after they have been merged should be no greater than the sum of their risk measures before they were merged. In the literature, one simply reads "a merger does not create extra risk".

The first three conditions are straightforward, given the intuitive definition of a risk measure. The fourth condition states that diversification helps reduce risk. When we aggregate two risks, the total risk should either decrease or stay the same. In the next section, we will show that VaR is not a coherent risk measure by showing that it is not sub-additive. The ETL on the other side does satisfies all these properties and is therefore a coherent risk measure.

8.3 Value at Risk

Value at Risk (VaR) is still the standard tool in measuring market risk. It is an attempt to provide a single number that summarizes the total risk in a portfolio of financial assets. Indeed, VaR tells the worst loss over a target horizon with a given level of confidence. Beyond funds managers, corporate treasurers and financial institutions, central banks also use VaR in determining the bank capital requirement.

Considering a portfolio, a confidence level $L$and a time horizon N, saying that V is the VaR of the portfolio means that we are L percent certain that *we will not lose more than V dollars in the next N days.* This seems simple and pretty intuitive as it answers the question "how bad can things get". At this time of writing, under the Basel II prescriptions for example, the capital market risk for financial institution is the 10 days VaR at 99% confidence level multiplied by a factor between 3 and 4 to provide the minimum capital requirements for regulatory purposes. VaR is just a simple quantile of the distribution of the losses.

In the special case of the loss distribution function being normal, i.e $F_X \sim \mathcal{N}(\mu, \sigma^2)$ for a given $\alpha \in (0, 1)$ one has

$$VaR_\alpha = \mu + \sigma\Phi^{-1}(\alpha), \tag{8.1}$$

where Φ denotes the standard normal distribution function and $\Phi^{-1}(\alpha)$ is the α-quantile of Φ.

The whole attraction of the VaR is its ability to summarizes all sort of risk in a single number. This charm however goes as at the cost of its weaknesses. As pointed out by numerous authors, VaR is not a good risk measure In particular, Artzner, Delbaen, Eber, and Heat (1999) Dowd (2005), Rachev, C, and Fabozzi (2005), Jorion (2002) and numerous references therein provide detailed discussions on weaknesses of VaR as a risk measure.

VaR as opposed to the expected tail loss (see next section), only tells us the most we can lose if a tail event does not occur. In case of an appearance of an extreme event, we can expect to lose more than the VaR but the VaR itself does not give any information about the magnitude of this possible huge loss. This implies that two positions can have the same VaR number meaning that they have the same risk when we use the VaR to measure it, but effectively have different risk exposures.

From the seminal paper Artzner, Delbaen, Eber, and Heat (1999), VaR is not a coherent risk measure because it is not sub-additive. A classical example adapted from Dowd (2005), page 34, is the following: Consider two identical securities A and B, each defaulting with probabilities 9% and we get a loss of says 50 if default occurs and a loss of 0 if no default occurs. The 90% VaR in each security is therefore 0. By assuming that the default are independent, one gets a lost of 0 with probability $(1 - 9\%)^2$ and a loss of 100 with probability $9\%^2$ and a loss of 50 with probability $1 - (1 - 9\%)^2 - 9\%^2$. VaR(X+Y) = 50 $> 0 =$ VaR(A)+ VaR(B). So VaR is not sub-additive.

The consequences of VaR being not sub-additive are painful, both in risk management perspective and in the regulatory perspective. If firms were to meet a risk control criterion that does not satisfy this property, the firm might have incentives to break up into several sub units to avoid satisfying the criterion. This will be an issue for the regulator as a huge amount of entities are always more difficult to control. This might lead to financial problems including possible tax evasions.

We learned in basic portfolio theory that diversification decreases the total risk exposure. Non subadditivity suggests that diversification is a bad thing and it recommends that putting all the assets in one basket is a good risk management practise. If risk measures are not sub-additive, adding them together gives us an underestimate of combined risk making the sum of individual risk not as a back-up- envelope-measure any longer. In case of subadditivity, the combined risk is bounded by the sum of the individual risk and this is seen as a conservative estimated of the combined risk.

Facing these drawbacks, particularly the non sub-additivity, researchers introduced recently the Expected Tail Loss as a natural coherent alternative to the VaR. In addition to all the properties already satisfied by the value at risk, the ETL is sub additive making it therefore a coherent measure of risk.

8.4 Expected Tail Loss

Although VaR is seen as a standard in measuring market risk related issues, we have highlighted its drawback. There is a need to transform the VaR into a proper risk measure which would be at least sub additive. An alternative to VaR is the so called expected tail loss.

Formally, following the notations as in the case of VaR in section 8.3, the ETL at confidence level $\alpha \in (0,1)$ is defined as:

$$ETL_\alpha(X) = E\big[X|X \geq VaR_\alpha(X)\big]. \tag{8.2}$$

Intuitively, it is the average loss for losses larger than the VaR. Expected Tail Loss takes into account losses beyond the VaR level. It is proven (see for example Dowd (2005) page 35) that it is sub additive making it therefore a coherent risk measure. In the literature, Expected Tail Loss, Conditional Value at Risk or Expected Shortfall are interchangeably used, when the underlying loss distribution is continuous.

When denoting F_L as a continuous distribution of the loss function, one rewrite it as

$$\begin{aligned} ETL_\alpha(X) &= E\big[X|X \geq VaR_\alpha(X)\big] \\ &= \frac{E(X\mathbb{I}_{[q_\alpha(X),\infty)}(X))}{P(X \geq q_\alpha(X))} \\ &= \frac{1}{1-\alpha}E(X\mathbb{I}_{[q_\alpha(X),\infty)}(X)) \\ &= \frac{1}{1-\alpha}\int_{q_\alpha(X)}^{\infty} x dF_X(x), \end{aligned}$$

where $\mathbb{I}_A$ is the indicator function with the value 1 if $x \in A$ and 0 otherwise. Remembering that $VaR_\alpha(X) = q_\alpha(X)$ and assuming that X is continuous, it follows

$$ETL_\alpha(X) = \frac{1}{1-\alpha}\int_\alpha^1 VaR_\alpha(X)dp.$$

When facing discrete distributions, we use the so called Generalized Expected Tail Loss (GETL) defined as

$$GETL_\alpha(X) = \frac{1}{1-\alpha}\left[\, E(XI_{[q_\alpha(X),\infty)}(X)) + q_\alpha(X)(1-\alpha-P(X\geq q_\alpha(X)))\,\right].$$

The second term in the expression on the right hand side above disappears when the distribution of X is continuous. In this case GETL and ETL are equal.

In the specific case where the returns are normally distributed, say $X \sim \mathcal{N}(\mu,\sigma^2)$ and assuming ϕ and Φ the density and the distribution function of the standard normal distribution , we easily derive

$$\begin{aligned}
ETL_\alpha(X) &= E\big[X|X\geq VaR_\alpha(X)\big] \\
&= E\big(\mu+\sigma Y|\mu+\sigma Y\geq VaR_\alpha(\mu+\sigma Y)\big) \text{ where } Y\sim\mathcal{N}(0,1) \\
&= E\big(\mu+\sigma Y|Y\geq VaR_\alpha(Y)\big) \\
&= \mu+\sigma ETL_\alpha(Y) \\
&= \mu+\sigma\frac{\phi(\Phi^{-1}(\alpha))}{1-\alpha}\ .
\end{aligned}$$

A simple proof that $ETL_\alpha(Y) = \frac{\phi(\Phi^{-1}(\alpha))}{1-\alpha}$ where $Y\sim\mathcal{N}(0,1)$ is found in Embrechts, McNeil, and Frey (2005) page 45.

Unlike VaR, ETL has the following interesting features: ETL gives information about the losses beyond the VaR, using information in the tail. It tells us what to expect in bad situations. It even gives an idea about how bad things might be while VaR tell nothing other than to expect a loss higher than the VaR itself.

ETL possesses the complete set of coherent risk measure as defined and explained in section 8.2. It is by definition and construction a form of conditional expected loss and it happen to be convenient to use in portfolio optimization. In particular, ETL is sub-additive while VaR is not and therefore has all the various attractions of sub-additivity and the VaR does not (see Dowd (2005) and Jorion (2002)). Finally, ETL gives an indication of extreme losses, in case they occur. Although it is yet to become a standard in the financial industry, Expected Tail Loss will surely play a major role, as it currently does in the insurance industry.

8.5 Results

The main assumption in our investigation is that the conditional profit/loss distribution is normal. The normal distribution is elliptic and in this case the VaR is sub additive. This allows us as well to use the closed forms solutions presented in the previous sections writing the programming routines for calculations.

We used a sample size of 4000 artificially generated data in two parameter settings $\omega = 0.001$ and $(\alpha, \beta) \in \{(0.15, 0.70); (0.20, 0.50)\}$. Entries are averaged over 1000 runs. In our case, VaR and ETL is calculated for a time horizon of 1 day at the confidence levels of 95% and 99%. Practitioners use the $\sqrt{T}$ rule to get an approximation of the T-day VaR. The one day ahead volatility is forecasted using a GARCH(1,1)-model. In the case of MDE GARCH, estimation of the parameters is performed from lags 2 to 40. Indeed, chapter 4 shows that a higher number of lags does not reveal any additional information. The weighting matrix is computed using the Newey and West (1987) procedure. For comparison, the VaR and ETL are calculated under the normal distribution assumption, as implied by the data, for each simulations.

Figures 8.1 and 8.2 show the magnitude of the estimates of two risk measures as a function of the lags. They show little variation and this implies that a lag does not have a significant influence in estimating these market risk measures

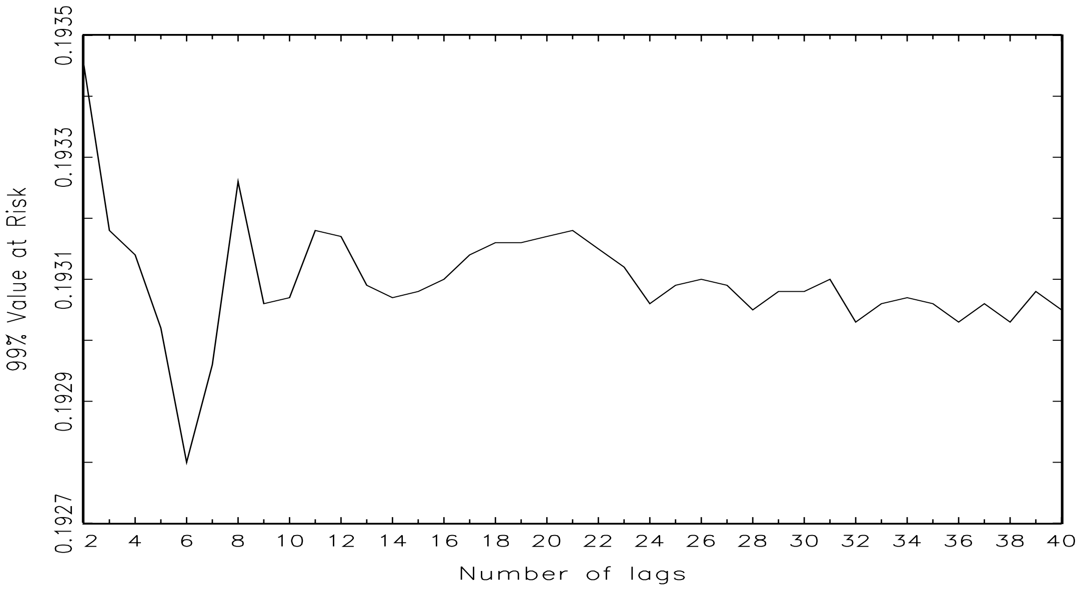


Figure 8.1: 99% VaR as a function of lags $\omega = 0.001,\ \alpha = 0.15,\ \beta = 0.70$

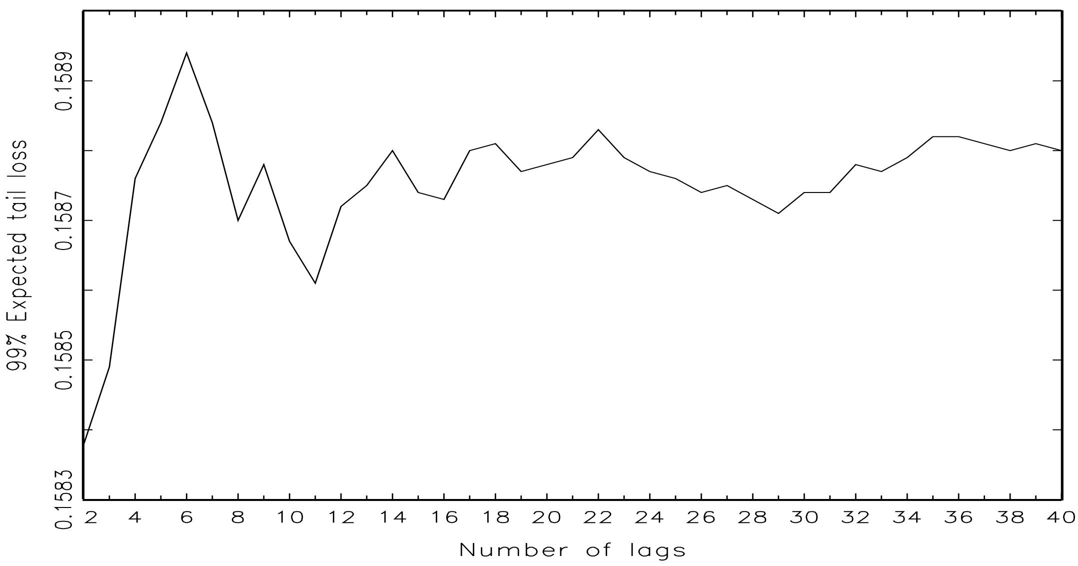


Figure 8.2: 99% ETL as a function of lags $\omega = 0.001,\ \alpha = 0.20,\ \beta = 0.50$

	$\hat{\omega} = 0.0010$	$\hat{\alpha} = 0.1493$	$\hat{\beta} = 0.6987$	
	(6.17e-005)	(0.0217)	(0.0353)	
95% MLE-VaR =	0.0658		95%MLE-ETL =	0.0825
	(0.0004)			(0.0013)
99%MLE-VaR =	0.0933		99%MLE-ETL=	0.1058
	(0.0014)			(0.0016)
95%VaR =	0.1307		95%ETL =	0.1638
	(0.0009)			(0.0012)
99%VaR =	0.1848		99%ETL=	0.2116
	(0.0014)			(0.0016)
	$\hat{\omega} = 0.0010$	$\hat{\alpha} = 0.2006$	$\hat{\beta} = 0.4978$	
	(7.10e-005)	(0.0308)	(0.0305)	
95%MLE-VaR =	0.0647		95%MLE-ETL =	0.0811
	(0.0003)			(0.0004)
99%MLE-VaR =	0.0916		99%MLE-ETL=	0.1039
	(0.0004)			(0.0005)
95% VaR =	0.0947		95%ETL =	0.1186
	(0.0002)			(0.0002)
99%VaR =	0.1339		99%ETL=	0.1533
	(0.0002)			(0.0002)

Table 8.1: 95% and 99% MLE-VaR, MLE-ETL, VaR and ETL.

and therefore a choice of a small number of lag within the minimum distance framework is reasonable and recommended.

Alternatively, calculations in which the GARCH parameters are estimated from the traditional maximum likelihood methods were performed. VaR and ETL as implied by the data, are as well calculated. Table 8.1 presents all these risk measures calculated at 95% and 99% confidence level with a holding period of one day. The numbers in brackets are the standard errors. Concretely, in the case $\omega = 0.001$, $\alpha = 0.15$ and $\beta = 0.70$ for example, the 99% MDE VaR (see figure 8.1) varies between 0.1927 and 0.1935 which is an interval of length 0.0008 (8bp!). The corresponding MLE VaR is 0.0933 only (see table 8.1). The implied 99% corresponding VaR is 0.1848, very close to the corresponding MDE VaR. This holds for all the MDE calculated risk measures, they are greater than their MLE equivalents in all the investigated cases and are very close to the risk measures implied by the data. This holds for all the parameter settings considered. This finding suggest that MDE value at risk and expected tail loss are reliable risk measures. They are more conservative than their MLE counterparts in the sense that they will have fewer exceptions (an exception is a value that is higher than the risk measure). This will then lead to higher reserves cash money which will happen to be helpful in case of claims but also reduces the investment capital of the financial institution.

8.6 Conclusion

The results in this paper show that Value at Risk and Expected Tail Loss calculations can be made with the daily volatility being forecast by a GARCH(1,1)-model where the parameters are minimum distance estimators as developed by Baillie and Chung (2001). Assuming the specific case where the loss distribution is normal, closed forms solutions allow one to calculate the Value at Risk and Expected Tail Loss as a function of the number of lags and evaluate their magnitude at each lag. The use of a small number of lag is recommended as the total error made using any lag among the first 40 investigated is less than 0.08%. The Value at Risk and Expected Tail Loss obtained by standard MLE GARCH models were anyhow clearly smaller in magnitude. The MDE calculated risk measures are very close to the risk measures naturally implied by the data. They outperform clearly the MLE counterparts.

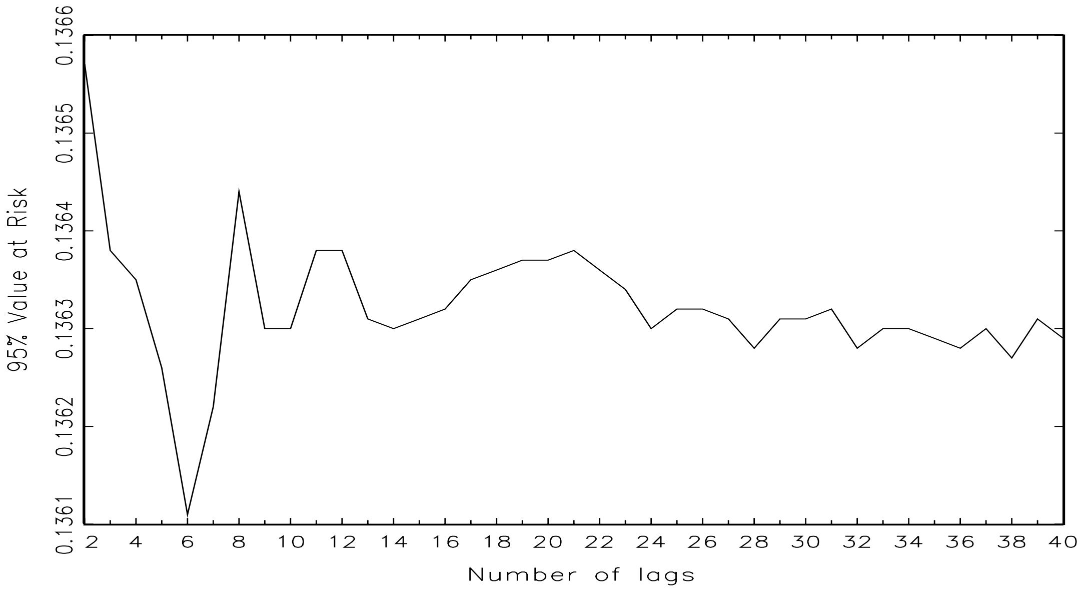


FIGURE 8.3: 95% VaR AS A FUNCTION OF LAGS $\omega = 0.001,\ \alpha = 0.15,\ \beta = 0.70$

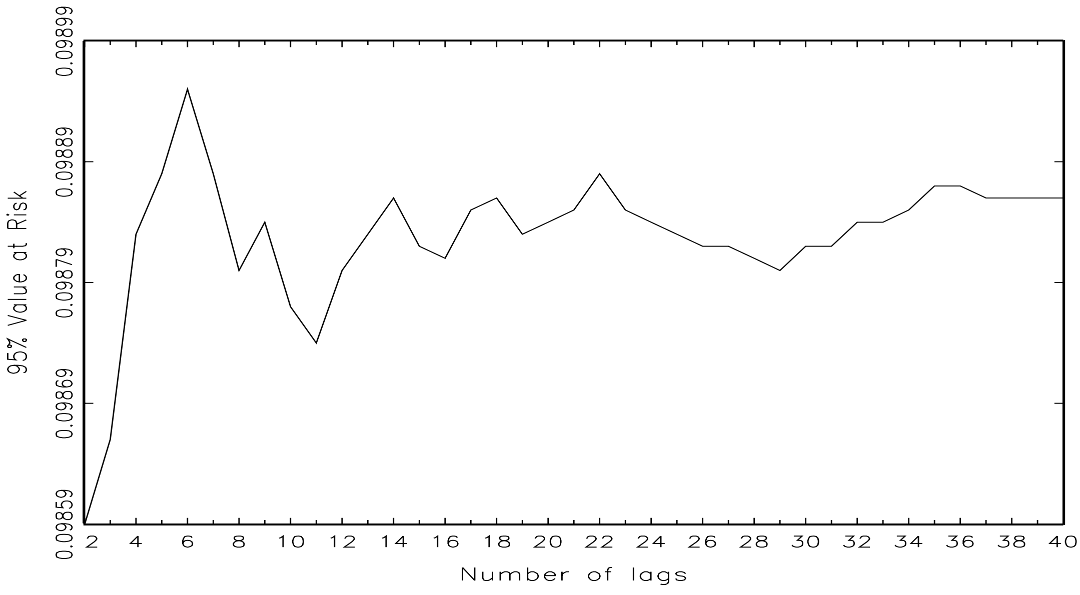


FIGURE 8.4: 95% VaR AS A FUNCTION OF LAGS $\omega = 0.001,\ \alpha = 0.20,\ \beta = 0.50$

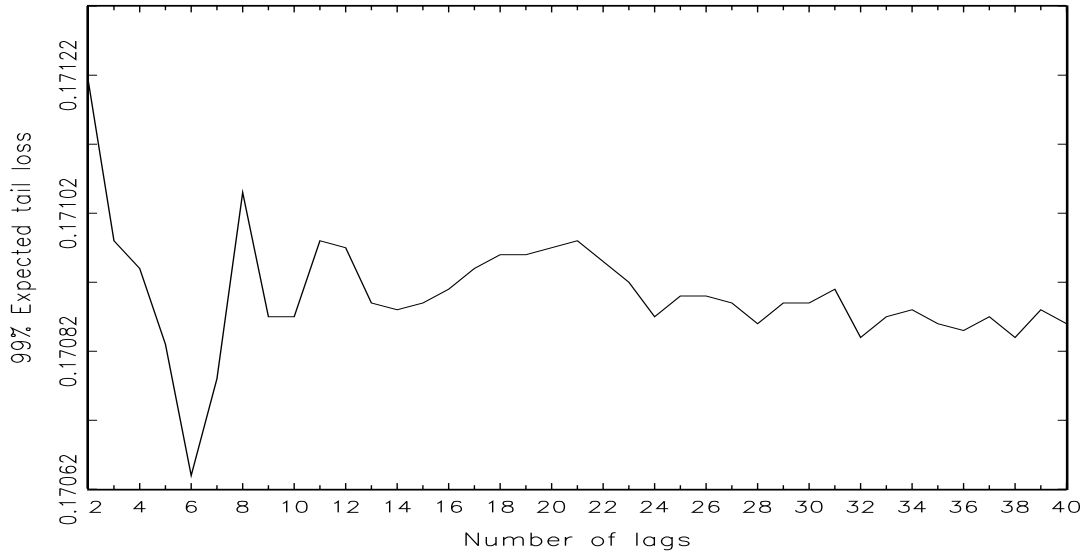


FIGURE 8.5: 95% ETL AS A FUNCTION OF LAGS $\omega = 0.001,\ \alpha = 0.15,\ \beta = 0.70$

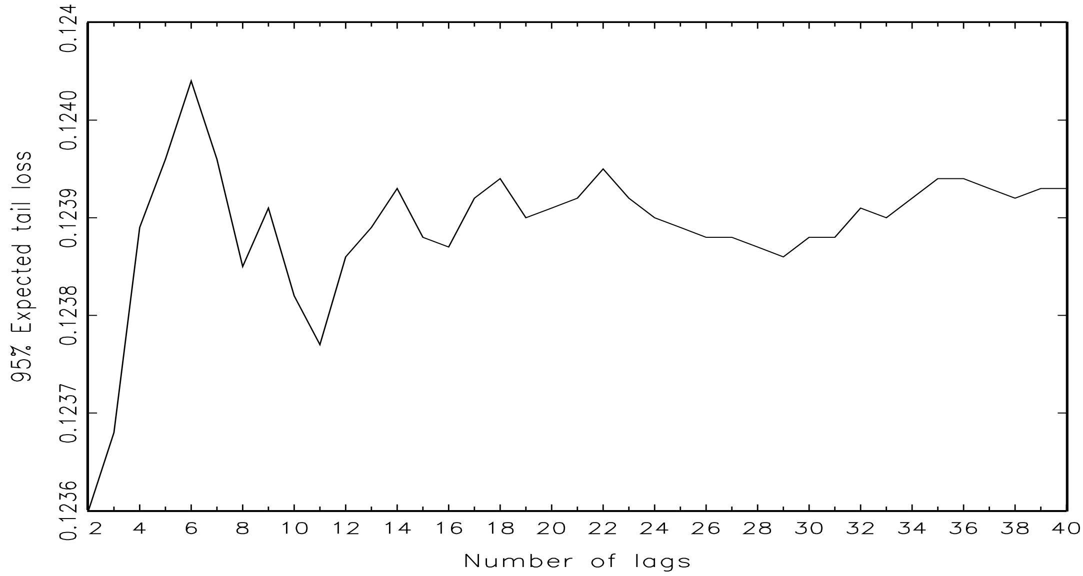


FIGURE 8.6: 95% ETL AS A FUNCTION OF LAGS $\omega = 0.001,\ \alpha = 0.20,\ \beta = 0.50$

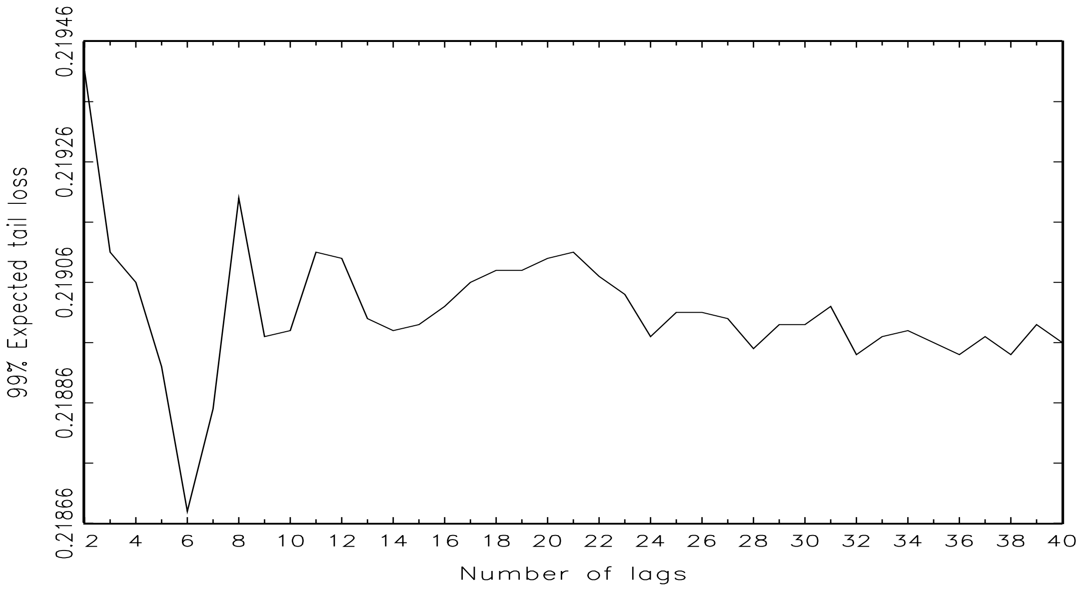


FIGURE 8.7: 99% ETL AS A FUNCTION OF LAGS $\omega = 0.001,\ \alpha = 0.15,\ \beta = 0.70$

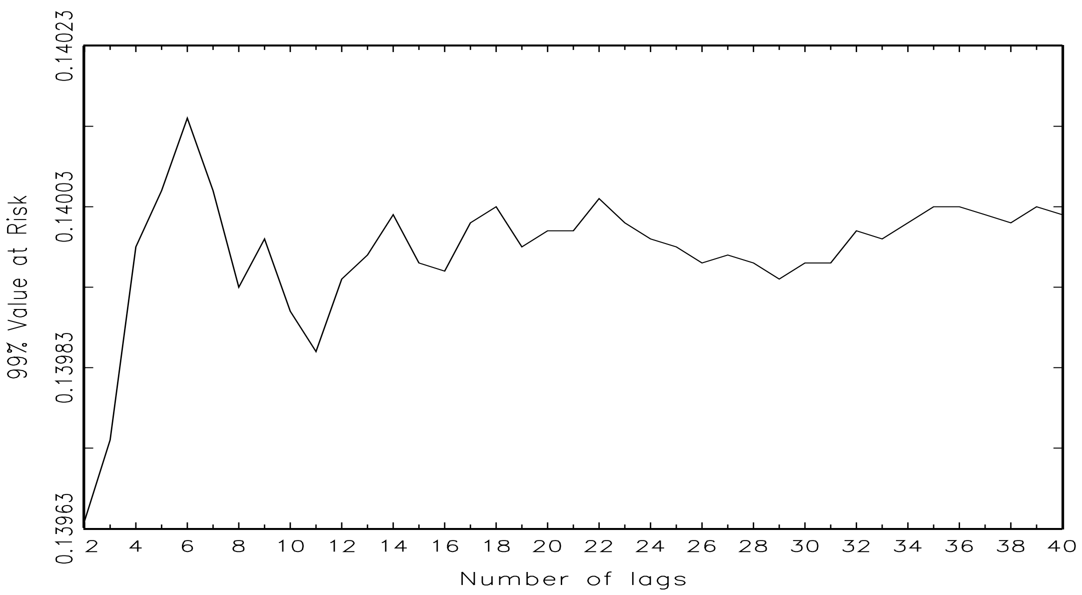


FIGURE 8.8: 99% VAR AS A FUNCTION OF LAGS $\omega = 0.001,\ \alpha = 0.20,\ \beta = 0.50$

Chapter 9

Concluding Remarks

Conditional heteroskedasticity is an important stylized fact of financial returns series. ARCH and GARCH models have been the workhorse in modern risk management when facing such issues. To account for some specific properties of financial data and to be as well in accordance with the economic theory, different reformulations of GARCH models have been studied in the literature, including nonlinear ones. Quasi maximum likelihood methods are usually used whenever estimating these models. In the presence of extreme non normality however, this estimation method can fail to deliver asymptotically efficient parameter estimates. Moreover, maximum likelihood methods have been shown to be inadequate in replicating the behavior of the autocorrelations of the squared observed returns. Distribution free approaches using minimum distance estimators have been proposed in the literature. We used the one presented by Baillie and Chung (2001) to address several issues investigated with the quasi maximum likelihood methods as well. This method presents the advantage that it does not require any distributional assumption of the underlying data.

After presenting GARCH models in general, we deal with the existence of its solution. We follow Bollerslev (1986) in discussing the existence of weakly stationary solutions for linear GARCH(p,q)- processes. The theory of products of random matrices is the appropriate technique to handle the case of strict stationarity as suggested by Nelson (1990). Bougerol and Picard (1992) followed his advice to solve the problem for general GARCH models. They characterized the existence of the solution by the strict negativity of the Lyapounov exponent associated to the GARCH corresponding random matrices.

The minimum distance estimation of the GARCH(1,1)-model is presented in chapter 3. Because this estimation is based on the autocorrelations of the squared process, we start by reviewing the calculations of the theoretical autocorrelations of the GARCH squared process as discussed in Bollerslev (1988) and Ding and Granger (1996). The empirical counterparts are easily and consistently estimated by (3.19). Having both the theoretical and the empirical autocorrelations, the only ingredient left in the MDE is the weighting matrix. We discussed the estimation of a consistent covariance matrix when, as it is the case in (2.44), the disturbances are not *i.i.d.*. Newey and West (1987) present a way of estimating it a practical situations and in all our applications, the weighting matrix is obtained in this way.

We addressed the issue of small sample bias of the estimated persistence of the GARCH(1,1)-model in chapter 4. Previous research in this direction has been done in Lumsdaine (1995) and Soosung and Pereira (2006) in the context of quasi and exact maximum likelihood. Here we have extend this work to the case where α and β are minimum distance estimators of the GARCH(1,1)-process. We covered a wider region of the parameter space and used much more replications in our simulations. Consistency and asymptotic normality of the minimum distance estimators of GARCH(1,1)-models have already been proven by Baillie and Chung (2001) and Storti (2006). They found that in certain regions of the parameter space, and for certain conditional densities, the minimum distance estimator can compete with the quasi maximum likelihood estimator. We find that the estimated persistence increases as the sample size increases in both methods of estimation. Furthermore, minimum distance estimators perform better than exact maximum likelihood in small samples (up to 500) in certain regions of the parameter space, in term of bias of the estimated persistence. As the sample size grows, maximum likelihood estimators becomes better than the minimum distance estimators, recovering the common finding that exact maximum likelihood is the best estimator, as least asymptotically. We recommend one to perform a monte carlo precheck whenever estimating small samples GARCH models. The smaller the sample, the greater the likelihood that minimum distance estimation outperforms maximum likelihood estimation.

Minimum Distance Estimators of a GARCH(1,1)-model depend directly on the autocorrelations of the squared process. The autocorrelations are computed at given lags. The MDE will directly depend on a chosen lag. A question we therefore address in chapter 5 is the lag choice in this context. We find that a number of lags between 10 and 30 is recommendable. In our Monte Carlo investigation, taking the mean squared error as the optimality condition, we find that the mean squared error decreases sharply in the first lags and changes only very marginally after the 30 or 40 first lags. Our result is consistent with different sample sizes and different parameter settings. A functional relationship between the optimal number of lags and the sample size was not found. As such, this is not a bad result because a certain relationship (depending e.g. on the sample size) could force one to use a large number of lags. This could turn out to be computationally demanding and time consuming making it difficult to update daily risk measures such as the value at risk in case of large lags. The optimal lag was often found at a very high lag. A similar investigation has been done using the bias as optimality condition. The results were mixed and a clear path was not found. But interestingly, the fluctuations of the graphs at early lags justify the possible use of a small number of lags.

Time series covering long time span often suffer from structural breaks. In chapter 6, we have studied the effect of deterministic structural breaks in the mean μ on the sum of the minimum distance estimators α and β of the GARCH(1,1)-model. The break in the constant term of the conditional mean equation occurs in the middle of the time series. This results in an increase of the estimated persistence. We support this finding with monte carlo evidence in which sample sizes are fixed and the size of the structural break is increasing. The estimated persistence increases as well as a function of the size of the break and becomes arbitrarily close to one if the structural change is large enough.

We extend the previous analysis in chapter 7 by considering this time again deterministic structural breaks in the mean, but for fixed sizes of the break and increasing sample sizes. This produces an increase in the empirical autocorrelations. We find that the estimated persistence likewise tends to one as the sample size increases. Structural changes in the time series can also occur stochastically. We study the case where the structural changes in the constant term of the conditional mean equation are stochastic. We impose a path to be followed by μ to be a function of a particular process with some switching probabilities. We find out that the estimated persistence increases independently of a particular switching probability and that it can be made arbitrarily close to one by increasing the sample size.

The constant term in the conditional mean equation μ and the constant term in the conditional variance equation ω are just scaling parameters in the GARCH process. The ARCH parameter α and GARCH parameter β are the two most important factors in modeling volatility. Value at Risk and Expected Tail Loss estimated from a GARCH updating volatility scheme are functions of α and β. There is a certain mistake we make by choosing a lag other than the optimal one when estimating these important market risk measures in GARCH

Minimum Distance Estimation framework. Chapter 8 studies these two market risk measures as a function of the lag used in Minimum Distance Estimation. We realize that these market risk measures are almost constant for the first 40 lags studied. Higher lags do not provide additional information as seen in chapter 5. The MDE-GARCH estimated risk measures are found to be much closer to the risk measures implied by the data than the MLE-GARCH risk measures.

List of Figures

List of Tables

Bibliography

ADENSTEDT, R. (1974): "On Large-sample Estimation for the Mean of a Stationary Random Sequence," The Annals of Statistics, 2(1), 1095–1107.

ARTZNER, P., F. DELBAEN, J. EBER, AND D. HEAT (1999): "Coherent measures of Risk," Mathematical Finance, 9(1), 203–228.

BAILLE, R., T. BOLLERSLEV, AND H. MIKKELSEN (1996): "Fractionally Integrated Generalized Autoregressive Conditional Heteroskedasticity," Journal of Econometrics, (74), 3–30,.

BAILLIE, R. T., AND T. BOLLERSLEV (1989): "The Message in Daily Exchange Rate: A Conditional-Variance Tale," Journal of Business & Economic Statistics, 7(3), 60–68.

——— (1990): "Intra-Day and Inter-Market Volatility in Foreign Exchange Rates," Review of Economic Studies, 58, 565–585.

BAILLIE, R. T., AND R. D. GENNARO (1990): "Stock Returns and Volatility," Journal of Financial and Quantitative Analysis, 25(2), 203–214.

BAILLIE, R, T., AND H. CHUNG (2001): "Estimation of GARCH models from the autocorrelations of the squared of a process," Journal of Time Series Analysis, 22(6), 631–650.

BARNDORFF-NIELSEN, O., AND D. COX (1989): "Asymptotic Techniques for

Use in Statistics. Chapman and Hall, New York.

Beran, J. (1994): ”Statistics for Long-Memory Processes. Chapman and Hall, New York.

Black, F. (1976): “The pricing of Commoditiy Contracts,” Journal of Financial Economics, 3(1), 167–179.

Bollerslev, T. (1986): “Generalized autoregressive conditional heteroskedasticity,” Journal of Econometrics, (31), 307–327.

Bollerslev, T. (1987): “A conditionally Heteroskedastic Times Series Model for Speculatives Prices and Rates of Returns,” Review of Economic Studies, 1(1), 542–547.

Bollerslev, T. (1988): “On the correlation structure for the generalized autoregressive conditional heteroskedasticity process,” Journal of Time Series Analysis, (9), 121–131.

Bollerslev, T., R. Chou, and K. Kroner (1992): “ARCH Modelling in Finance: A Review of Theory and Empirical Evidence,” Journal of Econometrics, 52, 5–59.

Bollerslev, T., R. Engle, and D. Nelson, B. (1994): ARCH MODELS Handbook of Econometrics volume 4. Elsevier Science B.V, Amsterdam.

Bollerslev, T., and R. F. Engle (1986): “Modelling the persistence of conditional variances,” Econometric Reviews, 5(1), 1–50.

Bollerslev, T., and H. Mikkelsen (1996): “Modelling and Pricing long Memory in Stock Market Volatility,” Journal of Econometrics, (73), 151–184.

Bollerslev, T., and J. Wooldridge (1992): “Quasi maximum likelihood

Estimations and Inference in Dynamic Models with Time varying Covariances," Econometric Reviews, (11), 143–172.

Bougerol, P. (1987): "Tightness of Products of Random Matrices and Stability of Linear Stochastic Systems," Annals of Probability, (15), 40–74.

Bougerol, P., and N. Picard (1990): "Strict Stationarity of Generalized Autoregressive Processes," Annals of Probability, (20), 1714–1730.

——— (1992): "Stationarity of GARCH processes and of some nonnegative time series," Journal of Econometrics, (52), 115–127.

Breit, T., N. Crato, and Lima (1998): "The Detection and Estimation of Long Memory in Stochastic Volatility," Journal of Econometrics, (83), 325–348.

Cai, J. (1989): "A Markov Model of Switching-Regime ARCH," Journal of Business & Economic Statistics, 12(3), 309–316.

Campbell, J., A. Lo, and A. MacKinlay (1997): The Econometrics of Financial Markets. Princeton University Press, New Jersey.

Cao, C., and R. Tsay (1992): "Nonlinear Time Series Analysis of Stock Return Volatility," Journal of Applied Econometrics, 52, 165–185.

Dacorogna, M., U. Muller, N. R.J, R. Olsen, and O. Pictet (1993): "A Geographical Model for the Daily and Weekly Seasonal Volatility in the FX Market," Journal of International Money and Finance, 12(1), 413–438.

Day, T., and L. C.M (1992): "Stock Market Volatility and the Information Content of Stock Index Options," Journal of Econometrics, 52, 267–287.

Diebold, F. (1986): "Modeling the persistence of the conditional variances:

a comment.," Econometric Reviews, (5), 51–56.

Diebold, F., and A. Inoue (2001): "Long Memory and Regime Switching," Journal of Econometrics, (105), 139–159.

Ding, Z., and C. Granger (1996): "Modeling Volatility Persistence of Speculative Returns: A New Approach," Journal of Econometrics, (73), 185–215.

Ding, Z., C. Granger, and R. Engle (1993): "A Long Memory Property of Stock Market Return.," Journal of Empirical Finance, (1), 83–106.

Dittmann, I., and C. Granger (2002): "Properties of nonlinear tranformations of fractionally integrated processes," Journal of Econometrics, (110), 113–133.

Domowitz, I., and H. White (1982): "Misspecified models with dependent observations," Journal of Econometrics, (20), 35–58.

Dowd, K. (2005): Measuring Market Risk. Wiley, West Sussex.

Drost, F., and T. Nijman (1993): "Temporal Aggregation of GARCH Processes," Econometrica, 61(4), 909–927.

Dueker, M. (1997): "Markov Switching in GARCH Processes and Mean-Reverting Stock-Market Volatility," Journal of Business & Economic Statistics, 15(15), 26–34.

Embrechts, P., A. McNeil, and R. Frey (2005): Quantitative Risk Management Concepts Techniques Tools. Princeton Series in Finance, New Jersey.

Engle, R. (1982): "Autoregressive Conditional Heterokedasticity with estimates of the variance of the united kingdom Inflation.," Econometrica, (50), 987–1007.

——— (2001): "GARCH(1,1): The use of ARCH and GARCH models in applied econometrics," Journal of Economic Perspectives, 15(4), 157–168.

Engle, R., D. Lilien, and R. Robins (1987): "Estimating time varying risk premia in the term structure: The ARCH-M model," Econometrica, 55(1), 391–407.

Engle, R., and A. Smith (1999): """Stochastic Permanent Breaks," The Review of Economics and Statistics, 81(1), 553–574.

Francq, C., J. Zakoian, and M. Roussignol (2001): "Conditional Heteroskedasticity Driven by Hidden Markov Chains," Journal of Time Series Analysis, 22(2), 197–220.

Franses, P., and C. Dijk (2000): Non Linear Time Series Models in Empirical Finance. Cambridge.

French, K., and G. S. Schwert (1987): "Expected Stock Returns and Volatility," Journal of Financial Economics, 22(19), 3–29.

Glosten, L., R. Jagannathan, and D. Runkle (1993): "On the Relation Between the Expected Value and the Volatility of the Nominal Excess Return on Stocks," The Journal of Finance, 48(1), 1779–1801.

Granger, C. (1980): "Long Memory Relationships and the Aggregation of Dynamic Models," Journal of Econometrics, (14), 227–238.

——— (1981): "Some Properties of Time Series Data and their Use in Econometric Model Specification," Journal of Econometrics, (16), 121–130.

——— (2005): "The past and future of empirical finance: some personal comments," Journal of Econometrics, (129), 35–40.

Granger, C., and R. Joyeux (1980): "An Introduction to Long Memory Time Series Models and Fractional Differencing," Journal of Time Series Analysis, (1), 15–39.

Greene, W. (2003): Econometric Analysis. Prentice-Hall, Upper Saddle River, fourth edn.

Haas, M., S. Mittnik, and . M. S. Paollela (2004a): "Mixed Normal conditional Heteroskedasticity," Journal of Financial Econometrics, 2(2), 211–250.

Haas, M., S. Mittnik, and M. S. Paollela (2004b): "A New Approach to Markov-Switching GARCH Models," Journal of Financial Econometrics, 2(4), 493–530.

Hamilton, J. D., and R. Susmel (1994): "Autoregressive Conditional Heteroskedasticity and Change in Regime," Journal of Econometrics, (64), 307–333.

Harvey, A. (1993): "Long Memory in Stochastic Volatility, Manuscript, London School of Economics," .

Hassler, U. (1997): "Sample autocorrelations of nonstationary fractionaly integrated series," Statistical Papers, (38), 43–62.

Hentschel, L. (1995): "All in the Family: Nesting Linear and Nonlinear GARCH models," Journal of Financial Economics, 39(1), 139–164.

Heyde, C., and Y. Yang (1997): "On Defining Long Range Dependence," Journal of Applied Probability, 34(1), 939–944.

HILLEBRAND, E. (2005): "Neglecting parameter changes in GARCH models," Journal of Econometrics, (129), 121–138.

HONG, P. (1991): "The Autocorrelation Structure for the GARCH-M process," Economics Letters, 37(1), 129–132.

HULL, J. (2006): Options Futures and Others Derivatives. Prentice Hall.

——— (2007): Risk Management and Financial Institutions. Prentice Hall.

JACQUIER, E., N. POLSON, AND P. ROSSI (1994): "Bayesian Analysis of Stochastic Volotility Models," 12(1), 371–417.

JORION, P. (2002): Value At Risk. Mc Graw-Hill, Singapore.

KINGMAN, J. (1973): "Sub-additive Ergodic Theory," Annals of Probability, (1), 833–909.

KLAASSEN, F. (2002): "Improving GARCH volatility forecasts with regime-switching GARCH," Empirical Economics, 2(27), 363–394.

KRÄMER, W. (1985): Trend in Ökonometrischen Modellen. Königstein, (Athenauem/Hain).

KRÄMER, W., AND B. TAMEZE (2007): "Structural Change and Estimated Persistence in the GARCH(1,1)-model to appear in Economics Letters," .

LAMOUREUX, C., AND W. LASTRAPES (1990): "Persistence in Variance, Structural Change, and the GARCH Model," Journal of Business & Economic Statistics, (8), 225–234.

LEE, S., AND B. HANSEN (1991): "Asymptotic Properties of the Maximum Likelihood Estimator and Test of the Stability of Parameters of the GARCH and IGARCH Models," Manuscript University of Rochester.

LIU, M. (2000): "Modelling long memory in stock market volatility," Journal of Econometrics, (89), 139–171.

LUMSDAINE, R. (1995): "Finite Sample Properties of the Maximum Likelihood Estimator in GARCH(1,1) Models: A Monte Carlo Investigation," Journal of Business & Economic Statistics, 13(1), 1–10.

——— (1996): "Consistency and Asymptotiy Normality of the Quasi Maximum Likelihood Estimator in IGARCH(1,1) and Covariance stationary GARCH(1,1)-Models," 64(3), 575 – 596.

MANDELBROT, B. (1963): "The Variation of Certain Speculative Prices," Journal of Business, (36), 394–419.

MCCURDY, T., AND I. MORGAN (1988): "Testing the martingale hypothesis in Deutschmark futures with models specifying the form of the heteroskedasticity," Journal of Applied Econometrics, (3), 187–202.

MCLEOD, A., AND K. HIPEL (1978): "Preservation of the Rescaled Adjusted Range,1 :A reassessment of the Hurst Phenomenon," Water Ressources Research, 14(1), 491–508.

MIKOSCH, T., AND C. STARICA (2000): "Limit Theory for the Sample Autocorrelations and Extremes of a GARCH(1,1) Process," The Annals of Statistics, 28(5), 1427–1451.

——— (2004): "Nonstationarities in Financial Times Series, the Long Range Depence, and the IGARCH Effects," The Review of Economics and Statistics, (86), 378–390.

MULLER, U., M. DACOROGNA, N. R.J, R. OLSEN, AND O. PICTET (1997): "Volatility of Different Time Resolutions - Analysing the Dynamics of Market

Components," Journal of Empirical Finance, 4(1), 213–239.

Nelson, D. (1990): "Stationarity and Persistence in the GARCH(1,1)-model," Econometric Theory, 6(1), 318–334.

Nelson, D. (1991): "Conditional Heteroskedasticity in Asset Returns: A new Approach," Econometrica, 59(1), 347–370.

Nelson, D., and C. Cao (1992): "Inequality Constraints in the Univariate GARCH Model," Journal of Business & Economic Statistics, 10(1), 229–235.

Newey, W. K., and K. D. West (1987): "A Simple, Positive Semi-Definite, Heteroskedasticity and Autocorrelation Consistent Covariance Matrix," Econometrica, 55(3), 703–708.

Noh, J., and A. Kane (1994): "Forecasting Volatility and Option Prices of the S and P 500 Index," The Journal of Derivatives, (1), 17–30.

Rachev, S., M. C, and Fabozzi (2005): Fat-tailed and skewed asset retunrs distribtions. Wiley, New Jersey.

Ray Chou, Y. (1988): "Volatility Persistence and Stock Valuations: Some empirical Evidence using Garch," Journal of Applied Econometrics, 3, 279–294.

Sentana, E. (1995): "Quadratic ARCH models," Review of Economic Studies, 62(1), 639–661.

Soosung, H., and P. Pereira (2006): "Small Sample Properties of GARCH estimates and Persistence," The European Journal of Finance, 12(6), 473–494.

Storti, G. (2006): "Minimum Distance Estimation of GARCH(1,1) models," Computational Statistics and Data Analysis, 51(3), 1803–1821.

White, H. (1984): Asymptotic Theory for Econometricians. Academic Press, Orlando.

Wong, C., and W. Li (2001): "On a Mixture Autoregressive Heteroskedastic Model," Journal of the American Statistical Association, (96), 982–995.

Chapter 10

Appendix

Authors	Data	Size	$\hat{\delta}$
Cai (1989)	Returns 3m-T-Bill 08/64-11/91	328	0.980
Bollerslev (1987)	500C 01/47-09/84	453	0.842
Bollerslev (1987)	Indus 01/47-09/84	453	0.834
Bollerslev (1987)	Cap. goods 01/47-09/84	453	0.816
Bollerslev (1987)	Cons goods 01/47-09/84	453	0.885
Bollerslev (1987)	Pub. Util. 01/47-09/84	453	0.943
Baillie and Gennaro (1990)	Retuns VW 02/28-12/84	683	0.924
Cao and Tsay (1992)	S+P 01/28-12/89	744	0.980
Cao and Tsay (1992)	Ret VW 01/28-12/89	744	0.975
Cao and Tsay (1992)	Ret EW 01/28-12/89	744	0.982

Table 10.1: Empirical estimates of the persistence parameter in GARCH(1,1)–model case of monthly data

Authors	Data	Size	$\hat{\delta}$
Lamoureux and Lastrapes (1990) ***	20 stocks (80–84)	358	0.728
Mikosch and Starica (2004)	S+P 86-87	375	0.835
Mikosch and Starica (2004)	S+P 53-56	750	0.831
Bollerslev and Engle (1986)	FX US,SwF 07/73-08/85	632	0.996
McCurdy and Morgan (1988)	Returns Futures	1067	0.985
Baillie and Bollerslev (1989)	FX FF,US 01/03/80-28/01/85	1245	0.943
Baillie and Bollerslev (1989)	FX IT,US 01/03/80-28/01/85	1245	0.961
Baillie and Bollerslev (1989)	FX JPY,US 01/03/80-28/01/85	1245	0.990
Baillie and Bollerslev (1989)	FX CHF,US 01/03/80-28/01/85	1245	0.980
Baillie and Bollerslev (1989)	FX BP,US 01/03/80-28/01/85	1245	0.971
Baillie and Bollerslev (1989)	FX DM,US 01/03/80-28/01/85	1245	0.966
Francq, Zakoian, and Roussignol (2001)	CAC40 1/6/88-31/12/93	1286	0.923
Noh and Kane (1994)	S+P 500 21/04/86-31/12/91	1339	0.984
Dueker (1997)	S+P 12/82-12/91	2370	0.974
Hull (2006)	FX Yen 06/01/88-15/08/97	2423	0.960
Engle (2001)	Portf NDJLB 23/03/90-23/03/00	2500	0.982
Hillebrand (2005)	D J 07/12/87-31/10/03	4000	0.996
Lamoureux and Lastrapes (1990) **	30 stocks 01/01/63-13/11/79	4228	0.978
Klaassen (2002)*	FX 03/01/78-23/07/97	4982	0.980
Haas, Mittnik, and Paollela (2004b)	FX SingD,USD 01/81-06/03	5313	0.933
Haas, Mittnik, and Paollela (2004b)	FX SingD,USD 01/81-06/03	5313	0.986
Haas, Mittnik, and Paollela (2004b)	FX BP,USD 01/81-06/03	6313	0.974
Haas, Mittnik, and Paollela (2004b)	FX BP,USD 01/81-06/03	6313	0.990
Haas, Mittnik, and Paollela (2004b)	FX JPY/USD 01/81-06/03	6336	0.958
Haas, Mittnik, and Paollela (2004b)	FX JPY/USD 01/81-06/03	6336	0.965
Breit, Crato, and Lima (1998)	VW Ret. 07/62-07/69	6801	0.999
French and Schwert (1987)	S+P 01/28-12/52	7326	0.992
Haas, Mittnik, and Paollela (2004b)	Nasdaq Ret 02/71-06/01	7681	0.986
French and Schwert (1987)	S+P 01/53-12/84	8043	0.992
Bollerslev and Mikkelsen (1996)	S+P 500 02/01/53-31/12/90	9558	0.995
French and Schwert (1987)	S+P 01/28-12/84	15369	0.996
Ding, Granger, and Engle (1993)	S+P 500 03/01/28-30/08/91	17055	0.997

Table 10.2: Empirical estimates of the persistence parameter in GARCH(1,1)-model case of daily data

Authors	Data	Size	$\hat{\delta}$
McCurdy and Morgan (1988)	Returns Futures	219	0.888
Drost and Nijman (1993)	FX FF,USD 01/03/80-28/01/85	249	0.799
Drost and Nijman (1993)	FX IT,USD 01/03/80-28/01/85	249	0.845
Drost and Nijman (1993)	FX CHF,USD 01/03/80-28/01/85	249	0.905
Drost and Nijman (1993)	FX BP,USD 01/03/80-28/01/85	249	0.891
Drost and Nijman (1993)	FX DM,USD 01/03/80-28/01/85	249	0.885
Day and C.M (1992)	S+P 11/11/83-28/12/89	319	0.907
Franses and Dijk (2000)	Returns Tokyo 01/86-12/95	520	0.981
Franses and Dijk (2000)	Returns Frankfurt 01/86-12/95	520	0.779
Franses and Dijk (2000)	Returns NY 01/86-12/95	520	0.987
Franses and Dijk (2000)	Returns Paris 01/86-12/95	520	0.926
Franses and Dijk (2000)	Returns FX BP 01/86-12/95	520	0.927
Franses and Dijk (2000)	Returns FX FF 01/86-12/95	520	0.945
Franses and Dijk (2000)	Returns FX DM 01/86-12/95	520	0.842
Ray Chou (1988)	Ret NYSE 07/62-12/89	1225	0.986
Hamilton and Susmel (1994)	Returns NYSE 07/62-12/87	1327	0.960

Table 10.3: Empirical estimates of the persistence parameter in QMLE GARCH(1,1)-model case of weekly data

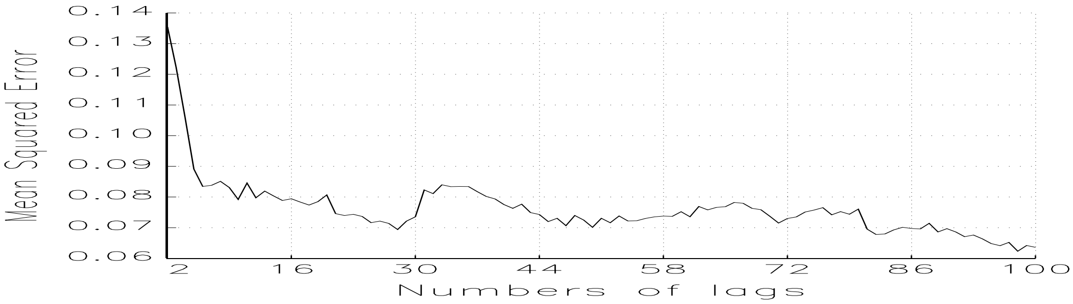


FIGURE 10.1: MSE OF $\hat{\delta}$ AS A FUNCTION ON NUMBER OF LAGS IN CASE OF $\alpha = 0.15$, $\beta = 0.70$ AND $\omega = 0.001$ SAMPLE SIZE OF 250

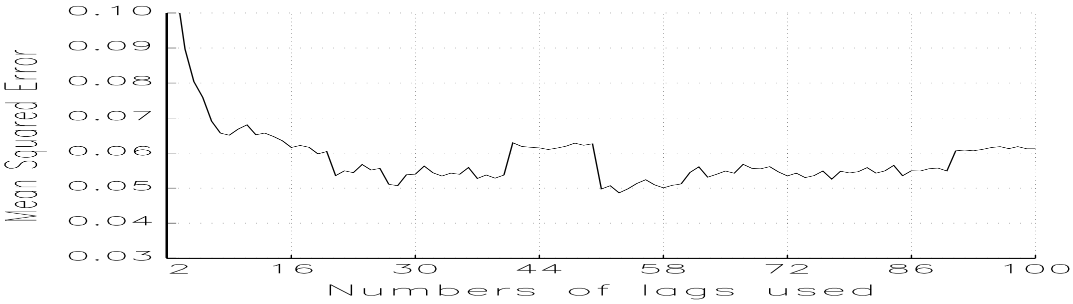


FIGURE 10.2: MSE OF $\hat{\delta}$ AS A FUNCTION ON NUMBER OF LAGS IN CASE OF $\alpha = 0.15$, $\beta = 0.70$ AND $\omega = 0.001$ SAMPLE SIZE OF 500

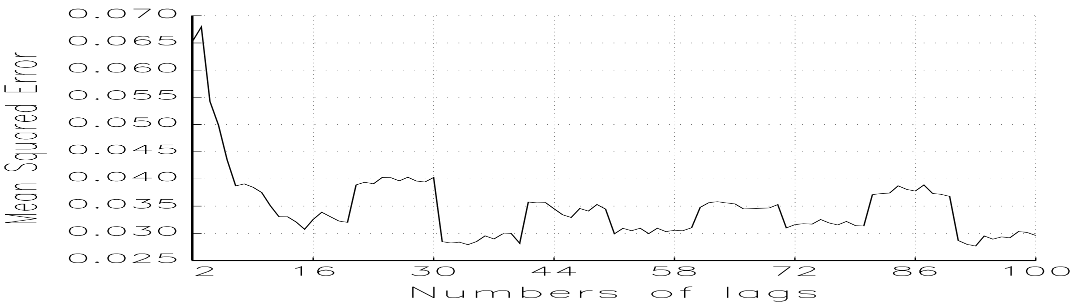


FIGURE 10.3: MSE OF $\hat{\delta}$ AS A FUNCTION ON NUMBER OF LAGS IN CASE OF $\alpha = 0.15$, $\beta = 0.70$ AND $\omega = 0.001$ SAMPLE SIZE OF 2000

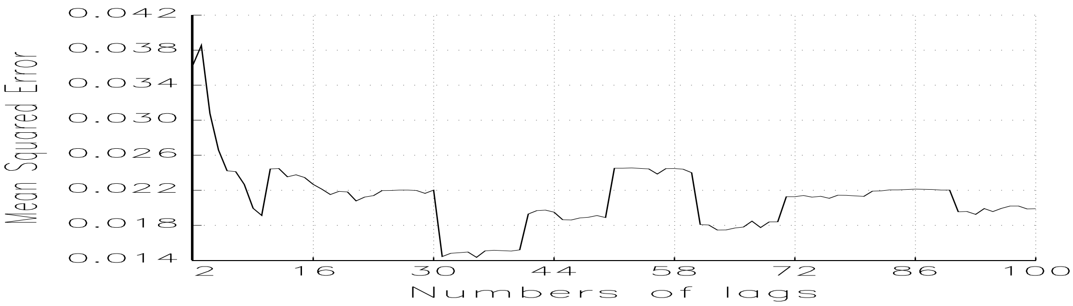


FIGURE 10.4: MSE OF $\hat{\delta}$ AS A FUNCTION ON NUMBER OF LAGS IN CASE OF $\alpha = 0.15$, $\beta = 0.70$ AND $\omega = 0.001$ SAMPLE SIZE OF 8000

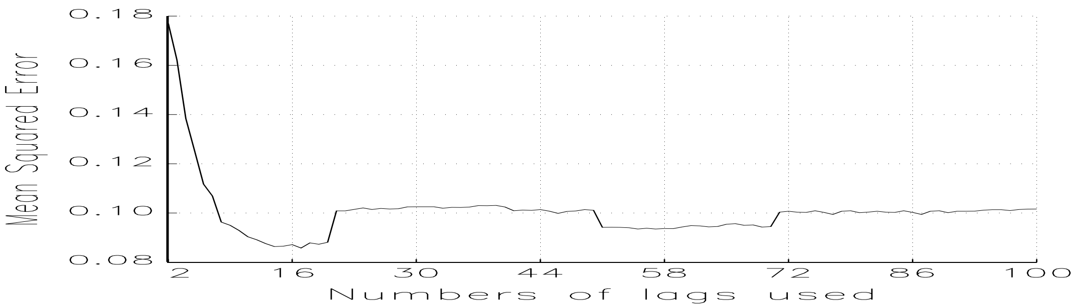


Figure 10.5: MSE of $\hat{\delta}$ as a function on number of lags in case of $\alpha = 0.20,\ \beta = 0.50$ and $\omega = 0.001$ sample size of 1000

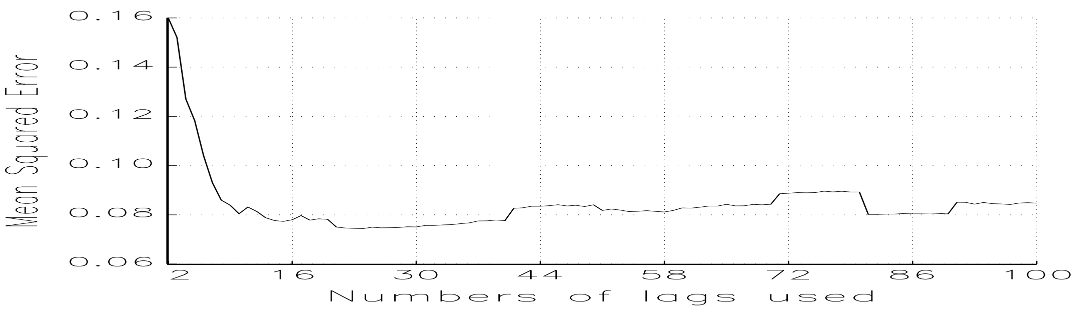


Figure 10.6: MSE of $\hat{\delta}$ as a function on number of lags in case of $\alpha = 0.20,\ \beta = 0.50$ and $\omega = 0.001$ sample size of 2000

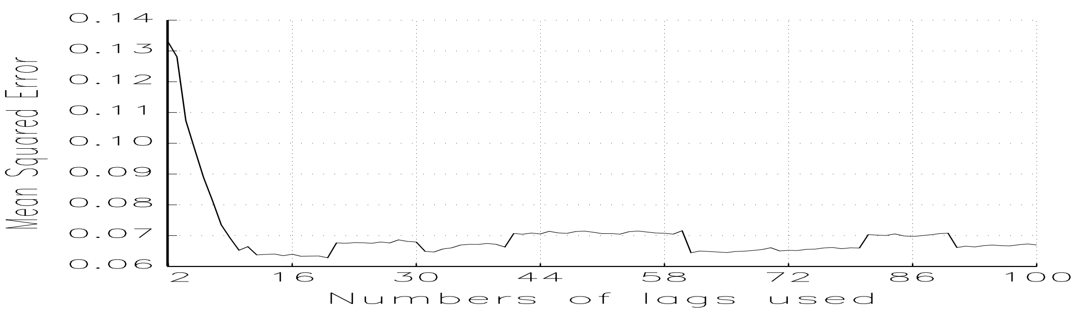


Figure 10.7: MSE of $\hat{\delta}$ as a function on number of lags in case of $\alpha = 0.20,\ \beta = 0.50$ and $\omega = 0.001$ sample size of 4000

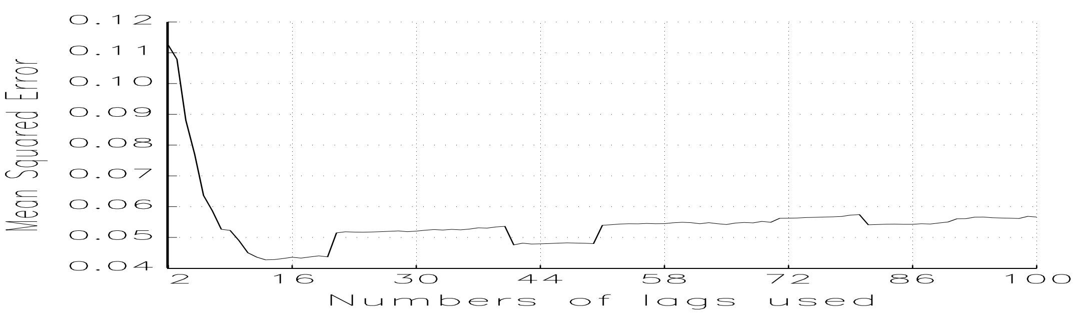


Figure 10.8: MSE of $\hat{\delta}$ as a function on number of lags in case of $\alpha = 0.20,\ \beta = 0.50$ and $\omega = 0.001$ sample size of 8000

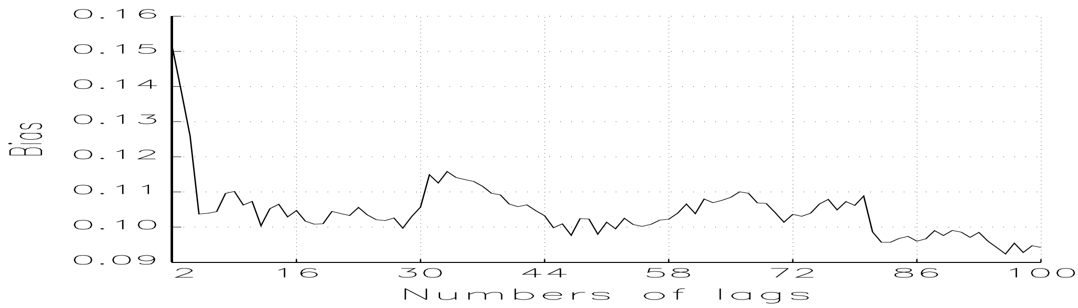


Figure 10.9: Bias of $\hat{\delta}$ as a function on number of lags in case of $\alpha = 0.15$, $\beta = 0.70$ and $\omega = 0.001$ sample size of 250

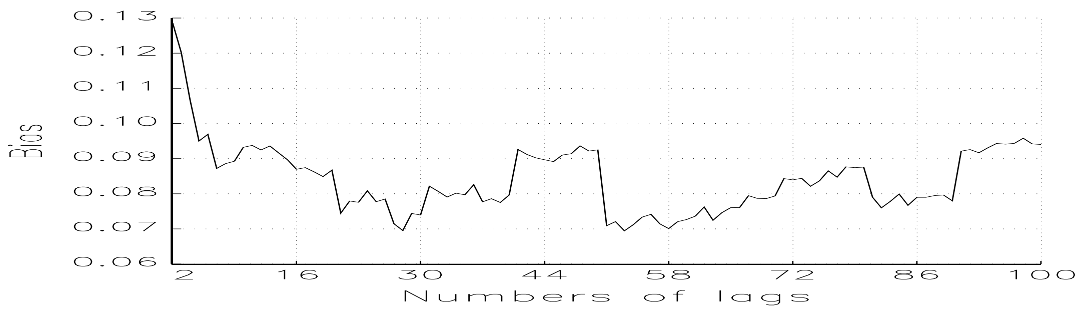


Figure 10.10: Bias of $\hat{\delta}$ as a function on number of lags in case of $\alpha = 0.15$, $\beta = 0.70$ and $\omega = 0.001$ sample size of 500

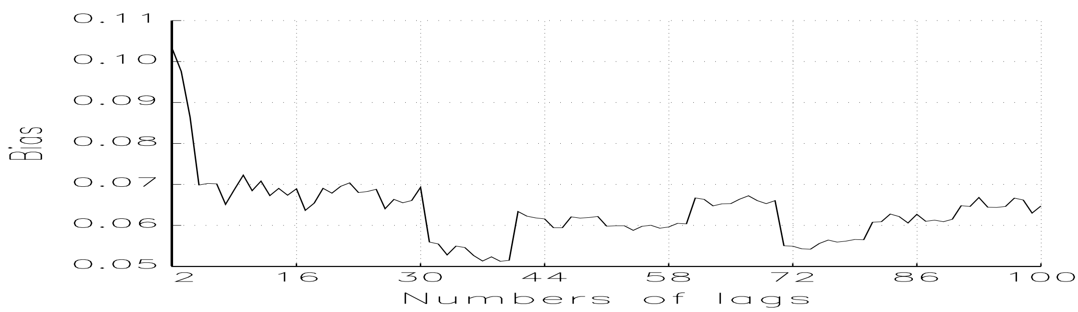


Figure 10.11: Bias of $\hat{\delta}$ as a function on number of lags in case of $\alpha = 0.15$, $\beta = 0.70$ and $\omega = 0.001$ sample size of 1000

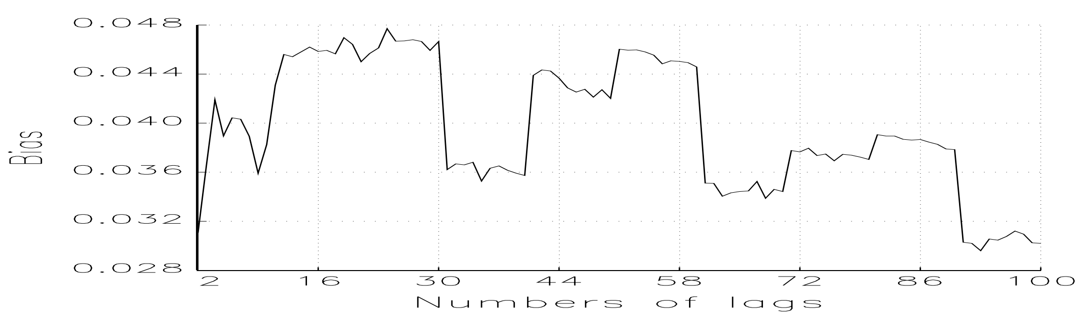


Figure 10.12: Bias of $\hat{\delta}$ as a function on number of lags in case of $\alpha = 0.15$, $\beta = 0.70$ and $\omega = 0.001$ sample size of 8000

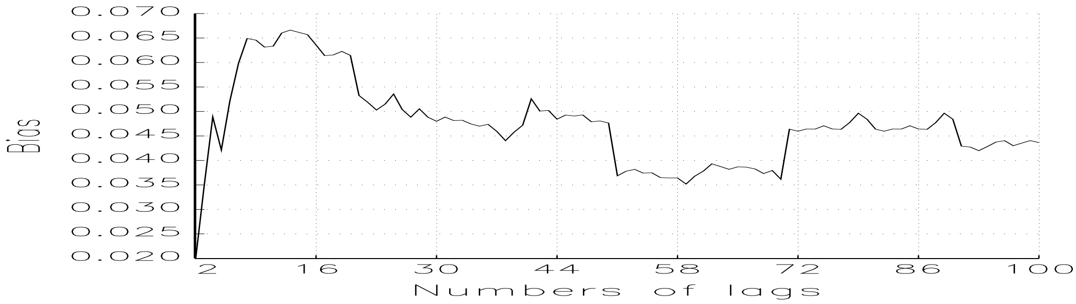


FIGURE 10.13: BIAS OF $\hat{\delta}$ AS A FUNCTION ON NUMBER OF LAGS IN CASE OF $\alpha = 0.20$, $\beta = 0.50$ AND $\omega = 0.001$ SAMPLE SIZE OF 250

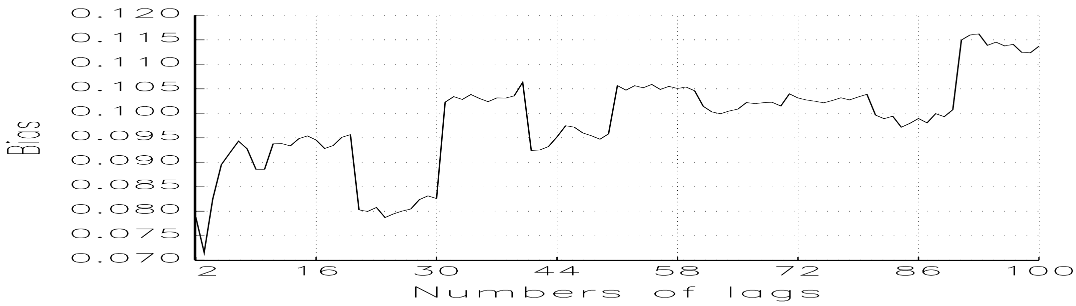


FIGURE 10.14: BIAS OF $\hat{\delta}$ AS A FUNCTION ON NUMBER OF LAGS IN CASE OF $\alpha = 0.20$, $\beta = 0.50$ AND $\omega = 0.001$ SAMPLE SIZE OF 2000

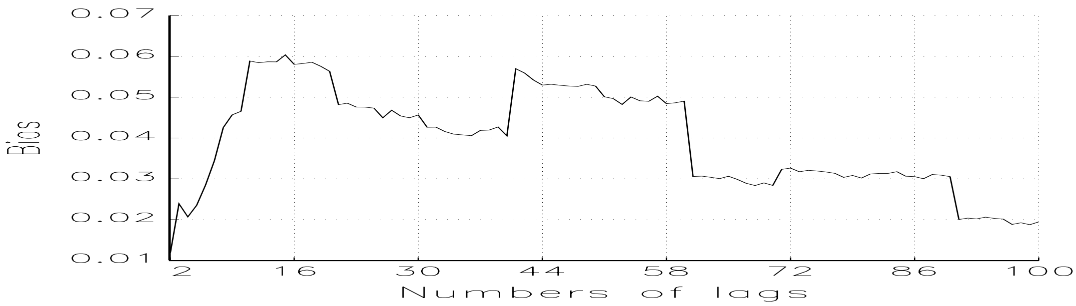


FIGURE 10.15: BIAS OF $\hat{\delta}$ AS A FUNCTION ON NUMBER OF LAGS IN CASE OF $\alpha = 0.20$, $\beta = 0.50$ AND $\omega = 0.001$ SAMPLE SIZE OF 4000

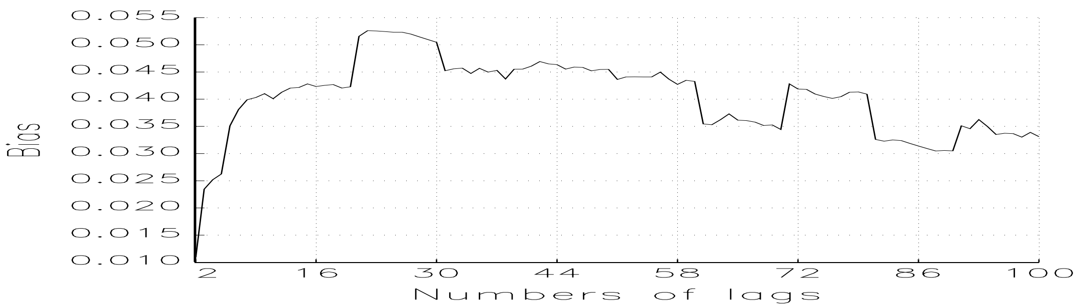


FIGURE 10.16: BIAS OF $\hat{\delta}$ AS A FUNCTION ON NUMBER OF LAGS IN CASE OF $\alpha = 0.20$, $\beta = 0.50$ AND $\omega = 0.001$ SAMPLE SIZE OF 8000

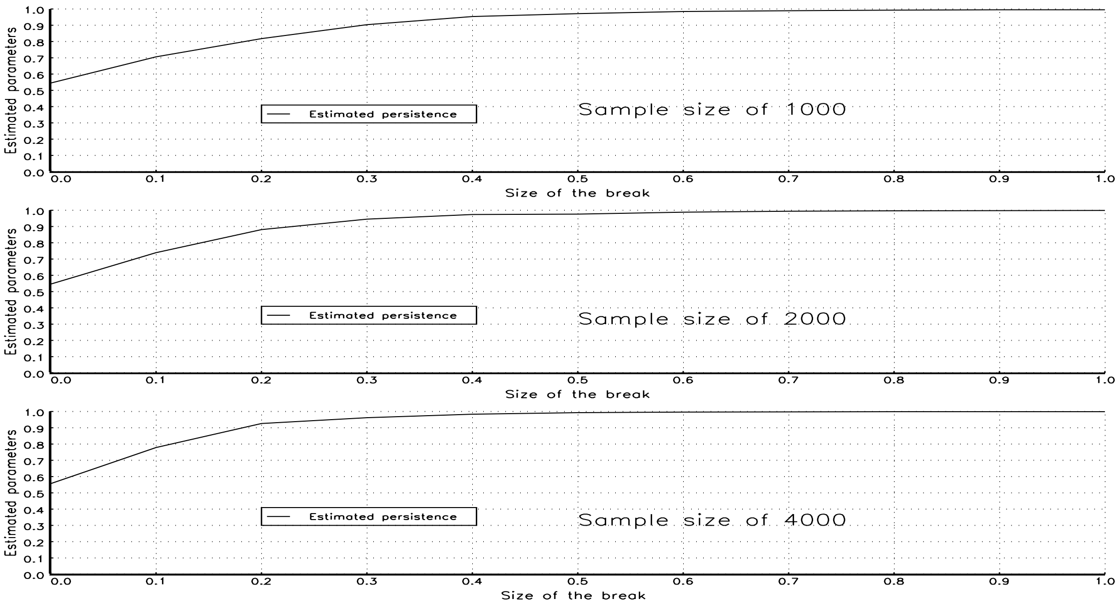


FIGURE 10.17: ESTIMATED PERSISTENCE AS A FUNCTION OF THE SIZE OF THE BREAK ($\alpha = 0.30$, $\beta = 0.30$ AND $\omega = 0.001$)

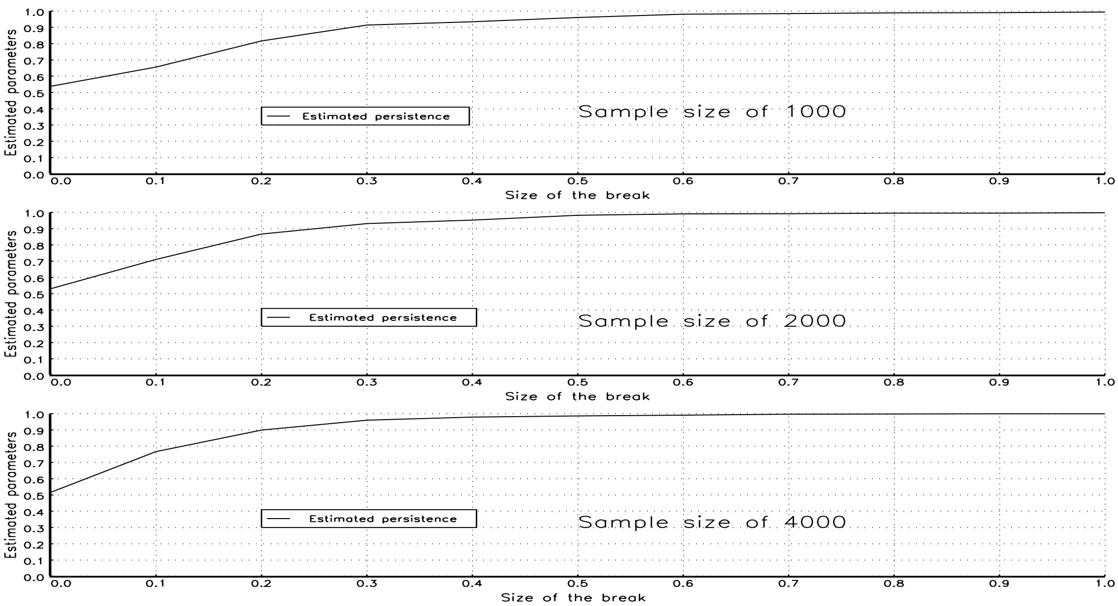


FIGURE 10.18: ESTIMATED PERSISTENCE AS A FUNCTION OF THE SIZE OF THE BREAK ($\alpha = 0.40$, $\beta = 0.20$ AND $\omega = 0.001$)

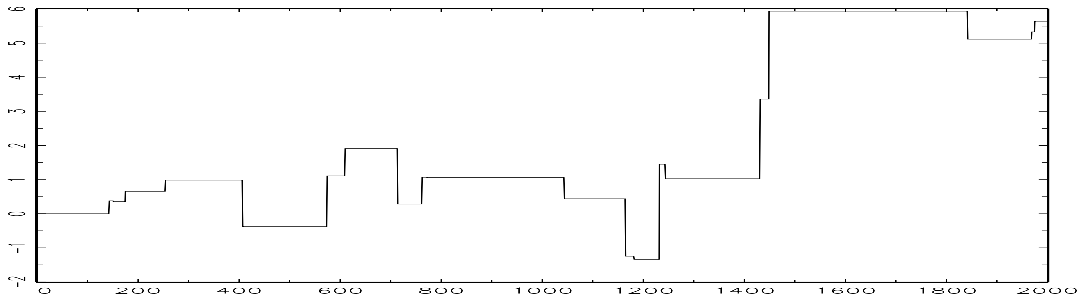


FIGURE 10.19: STOCHASTIC MEAN ACCORDING TO (7.16) AND (7.17), P = 0.01 AND T = 2000

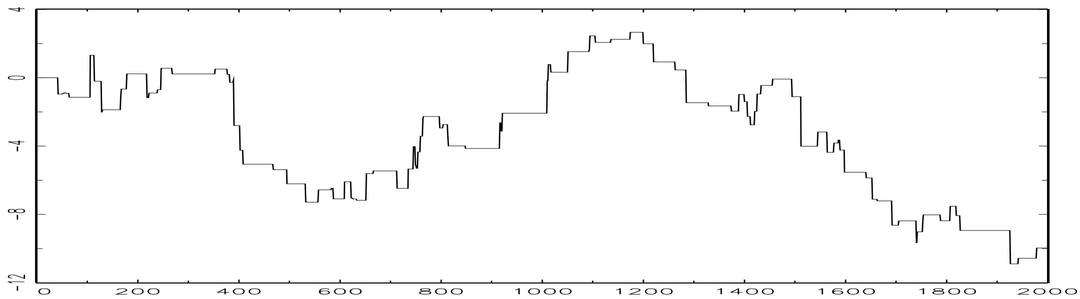


FIGURE 10.20: STOCHASTIC MEAN ACCORDING TO (7.16) AND (7.17), P = 0.05 AND T = 2000

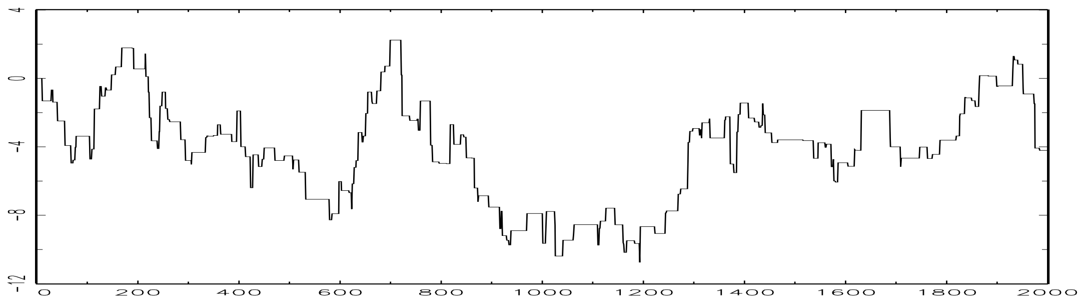


FIGURE 10.21: STOCHASTIC MEAN ACCORDING TO (7.16) AND (7.17), P = 0.10 AND T = 2000

Sample size	$\hat{\alpha}_{MDE}$	$\hat{\alpha}_{MLE}$	$\hat{\beta}_{MDE}$	$\hat{\beta}_{MLE}$	$\hat{\delta}_{MDE}$	$\hat{\delta}_{MLE}$
		a) $\alpha = 0.10,$	$\beta = 0.75$			
100	0.092	0.114	0.619	0.231	0.712	0.345
200	0.094	0.109	0.610	0.376	0.705	0.486
300	0.091	0.108	0.624	0.475	0.715	0.583
400	0.090	0.106	0.633	0.548	0.724	0.655
500	0.090	0.105	0.647	0.595	0.737	0.700
1000	0.089	0.102	0.687	0.702	0.777	0.804
2000	0.092	0.100	0.713	0.736	0.805	0.836
		b) $\alpha = 0.10,$	$\beta = 0.80$			
100	0.090	0.116	0.674	0.302	0.764	0.419
200	0.091	0.111	0.671	0.459	0.762	0.571
300	0.088	0.109	0.688	0.562	0.776	0.671
400	0.087	0.106	0.704	0.631	0.791	0.738
500	0.086	0.105	0.723	0.682	0.809	0.787
1000	0.089	0.102	0.687	0.702	0.777	0.804
2000	0.092	0.100	0.713	0.736	0.805	0.836
		c) $\alpha = 0.05,$	$\beta = 0.90$			
100	0.048	0.073	0.770	0.509	0.818	0.582
200	0.047	0.066	0.773	0.600	0.820	0.667
300	0.044	0.062	0.776	0.668	0.821	0.731
400	0.044	0.060	0.786	0.719	0.830	0.779
500	0.043	0.058	0.795	0.753	0.838	0.812
1000	0.042	0.053	0.838	0.850	0.880	0.904
2000	0.041	0.051	0.877	0.886	0.918	0.937

Table 10.4: MDE and MLE simulated mean of the estimated parameters of the standard GARCH(1,1)-model with different α's and β's

Sample size	$\hat{\alpha}_{MDE}$	$\hat{\alpha}_{MLE}$	$\hat{\beta}_{MDE}$	$\hat{\beta}_{MLE}$	$\hat{\delta}_{MDE}$	$\hat{\delta}_{MLE}$
		d) $\alpha = 0.08,$	$\beta = 0.82$			
100	0.075	0.098	0.6864	0.509	0.762	0.369
200	0.074	0.093	0.688	0.600	0.762	0.503
300	0.074	0.090	0.689	0.668	0.763	0.600
400	0.071	0.089	0.711	0.719	0.782	0.677
500	0.069	0.087	0.726	0.753	0.796	0.728
1000	0.070	0.082	0.766	0.759	0.837	0.842
2000	0.072	0.080	0.794	0.806	0.866	0.886
		e) $\alpha = 0.07,$	$\beta = 0.73$			
100	0.069	0.085	0.582	0.168	0.652	0.254
200	0.067	0.080	0.585	0.269	0.653	0.349
300	0.072	0.079	0.584	0.344	0.657	0.424
400	0.066	0.078	0.597	0.406	0.664	0.484
500	0.076	0.077	0.600	0.445	0.676	0.523
1000	0.066	0.074	0.622	0.588	0.689	0.663
2000	0.067	0.072	0.665	0.683	0.733	0.755
		f) $\alpha = 0.07,$	$\beta = 0.83$			
100	0.067	0.091	0.711	0.254	0.778	0.346
200	0.065	0.083	0.697	0.373	0.762	0.457
300	0.064	0.080	0.699	0.471	0.764	0.552
400	0.063	0.079	0.709	0.538	0.773	0.618
500	0.063	0.077	0.721	0.597	0.785	0.674
1000	0.062	0.073	0.769	0.739	0.831	0.813
2000	0.062	0.071	0.799	0.808	0.862	0.879

Table 10.5: MDE and MLE simulated mean of the estimated parameters of the standard GARCH(1,1)-model with different α's and β's

Sample size	$\hat{\alpha}_{MDE}$	$\hat{\alpha}_{MLE}$	$\hat{\beta}_{MDE}$	$\hat{\beta}_{MLE}$	$\hat{\delta}_{MDE}$	$\hat{\delta}_{MLE}$
		g) $\alpha = 0.075$,	$\beta = 0.70$			
100	0.081	0.083	0.570	0.154	0.651	0.238
200	0.077	0.081	0.547	0.249	0.624	0.330
300	0.074	0.080	0.545	0.324	0.623	0.404
400	0.073	0.080	0.551	0.381	0.624	0.461
500	0.071	0.079	0.563	0.428	0.634	0.508
1000	0.071	0.078	0.594	0.571	0.665	0.649
2000	0.072	0.077	0.630	0.657	0.702	0.733
		h) $\alpha = 0.055$,	$\beta = 0.90$			
100	0.052	0.076	0.788	0.541	0.840	0.618
200	0.050	0.069	0.777	0.637	0.828	0.707
300	0.049	0.066	0.782	0.705	0.831	0.771
400	0.046	0.064	0.797	0.757	0.843	0.821
500	0.047	0.062	0.806	0.789	0.853	0.851
1000	0.041	0.053	0.841	0.849	0.882	0.902
2000	0.040	0.077	0.888	0.885	0.920	0.936
		i) $\alpha = 0.055$,	$\beta = 0.80$			
100	0.056	0.078	0.649	0.161	0.706	0.240
200	0.059	0.069	0.654	0.246	0.713	0.316
300	0.057	0.066	0.659	0.319	0.717	0.385
400	0.062	0.0634	0.675	0.376	0.737	0.440
500	0.061	0.062	0.681	0.428	0.743	0.491
1000	0.052	0.060	0.692	0.591	0.744	0.651
2000	0.051	0.057	0.737	0.718	0.789	0.776
3000	0.048	0.056	0.749	0.767	0.797	0.823
4000	0.048	0.056	0.764	0.782	0.812	0.838

Table 10.6: MDE and MLE simulated mean of the estimated parameters of the standard GARCH(1,1)-model with different α's and β's

Sample size	$\hat{\alpha}_{MDE}$	$\hat{\alpha}_{MLE}$	$\hat{\beta}_{MDE}$	$\hat{\beta}_{MLE}$	$\hat{\delta}_{MDE}$	$\hat{\delta}_{MLE}$
		j) $\alpha = 0.10,$	$\beta = 0.77$			
100	0.093	0.117	0.630	0.254	0.723	0.371
200	0.093	0.110	0.633	0.403	0.726	0.513
300	0.091	0.108	0.645	0.501	0.737	0.609
400	0.089	0.107	0.662	0.577	0.752	0.684
500	0.089	0.106	0.677	0.624	0.766	0.730
1000	0.089	0.102	0.714	0.729	0.803	0.832
2000	0.091	0.100	0.741	0.759	0.832	0.859
		k) $\alpha = 0.09,$	$\beta = 0.77$			
100	0.086	0.108	0.642	0.231	0.728	0.339
200	0.084	0.103	0.626	0.366	0.710	0.468
300	0.082	0.100	0.641	0.462	0.723	0.563
400	0.081	0.099	0.646	0.534	0.727	0.633
500	0.081	0.097	0.665	0.586	0.746	0.683
1000	0.080	0.092	0.705	0.749	0.787	0.841
2000	0.082	0.091	0.734	0.786	0.817	0.877
		l) $\alpha = 0.09,$	$\beta = 0.80$			
100	0.082	0.108	0.679	0.264	0.761	0.372
200	0.081	0.101	0.671	0.410	0.753	0.511
300	0.081	0.099	0.676	0.509	0.758	0.608
400	0.080	0.097	0.692	0.579	0.772	0.677
500	0.080	0.095	0.705	0.631	0.785	0.727
1000	0.080	0.092	0.746	0.749	0.826	0.841
2000	0.073	0.091	0.766	0.786	0.839	0.877

TABLE 10.7: MDE AND MLE SIMULATED MEAN OF THE ESTIMATED PARAMETERS OF THE STANDARD GARCH(1,1)-MODEL WITH DIFFERENT α'S AND β'S

Sample size	$\hat{\alpha}_{MDE}$	$\hat{\alpha}_{MLE}$	$\hat{\beta}_{MDE}$	$\hat{\beta}_{MLE}$	$\hat{\delta}_{MDE}$	$\hat{\delta}_{MLE}$
		m) $\alpha = 0.16$,	$\beta = 0.60$			
100	0.138	0.112	0.473	0.167	0.611	0.279
200	0.141	0.114	0.479	0.254	0.621	0.369
300	0.140	0.114	0.489	0.309	0.629	0.424
400	0.140	0.115	0.500	0.346	0.640	0.462
500	0.140	0.117	0.512	0.375	0.652	0.492
1000	0.142	0.157	0.535	0.553	0.677	0.710
2000	0.145	0.160	0.557	0.584	0.702	0.744
		n) $\alpha = 0.08$,	$\beta = 0.80$			
100	0.076	0.099	0.678	0.230	0.755	0.330
200	0.076	0.091	0.662	0.356	0.738	0.447
300	0.073	0.088	0.671	0.453	0.744	0.542
400	0.073	0.087	0.675	0.526	0.749	0.614
500	0.073	0.086	0.687	0.582	0.760	0.668
1000	0.071	0.084	0.739	0.718	0.811	0.802
2000	0.073	0.081	0.767	0.779	0.840	0.861
		o) $\alpha = 0.10$,	$\beta = 0.70$			
100	0.081	0.109	0.578	0.217	0.659	0.327
200	0.082	0.107	0.568	0.344	0.650	0.452
300	0.083	0.107	0.575	0.430	0.658	0.537
400	0.084	0.106	0.589	0.497	0.673	0.603
500	0.086	0.105	0.597	0.537	0.684	0.642
1000	0.088	0.103	0.638	0.639	0.726	0.742
2000	0.092	0.101	0.667	0.697	0.760	0.780

Table 10.8: MDE and MLE simulated mean of the estimated parameters of the standard GARCH(1,1)-model with different α's and β's

Erklärung

Ich versichere, dass ich diese Dissertation selbständig verfasst habe. Bei der Erstellung der Arbeit habe ich mich ausschließlich der angegebenen Hilfsmittel bedient. Die Dissertation ist nicht bereits Gegenstand eines erfolgreich abgeschlossenen Promotion-oder sonstigen Prüfungsverfahrens gewesen.

Dortmund, den 21.07.2007

Baudouin Tameze Azamo