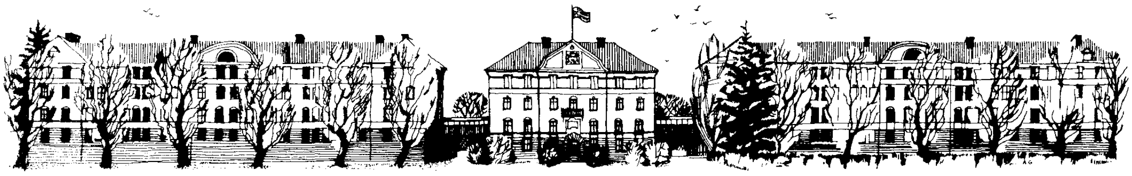
On Estimation of GARCH Models with an Application to Nordea Stock Prices

Chao Li

U.U.D.M. Project Report 2007:6

Examensarbete i matematik, 20 poäng
Handledare och examinator: Silvelyn Zwanzig

Januari 2007



Department of Mathematics

Uppsala University

On Estimation of GARCH Models with an Application to Nordea Stock Prices

Chao Li

Abstract

We are interested in estimation of stationary GARCH models. In simulation studies, we assess the performance of the maximum likelihood estimator and Yule-Walker estimator of the $GARCH\,(1,1)$ model. Finally we attempt to fit the dynamics of daily stock returns on Nordea by a GARCH model.

1 Introduction

The generalized autoregressive conditionally heteroscedastic (hereafter denoted as GARCH) model has gained a lot of attention in the literature since the introduction by Bollerslev(1986). Various works on testing for or modeling of time-varying volatility of stock market returns have been dominated by this model such that includes past variances in the explanation of future variances, which allows to capture the patterns usually exhibited by many financial time series such as volatility clustering, large kurtosis.

In this paper we are interested in parameter estimation of non-negative GARCH models under the weak stationary condition discussed by Bollerslev(1986). Although there exist more than one methods for estimating parameters in GARCH models, we primarily focus on the Maximum Likelihood (abbreviated ML) method facilitated by hypothetically assuming the innovation distribution to be normal, which is arguably the most frequently used technique in practice. We also consider Yule-Walker estimation based on an autoregressive moving average (abbreviated ARMA) representation of GARCH processes established in Theorem 12.9 of Franke, Hardle and Hafner without proof, and will provide a complete proof in Section 2.3. A simulation study of our paper aimed at checking the performance of the above-mentioned estimators is performed on a $GARCH\,(1,1)$ model in R environment. Because there is no available function handling Yule-Walker estimation for $ARMA(p,q)$ models with $q>0$ in R, we write a small program that only deals with the $p=q=1$ case by solving the Yule-Walker derived quadratic equation.

The paper is organized as follows. Section 2 reviews preliminary background information necessary for our work. Parameter estimation is considered in Section 3, while Section 4 contains simulation results comparing the performance of estimators discussed. Section 5 is devoted to an application to Nordea daily stock returns, and finally some remarks are given at the end of the paper.

2 Preliminaries

2.1 Definition

Let ε_t denote a real-valued discrete-time stochastic process, and F_t the information set at time t. In [1] ε_t follows a $GARCH\,(p,q)$ process if:

Definition 1 *The process ε_t, $t \in Z$, is $GARCH\,(p,q)$, if*

$$\varepsilon_t | F_{t-1} \sim N\left(0, \sigma_t^2\right), \tag{1}$$

$$\sigma_t^2 = \omega + \sum_{i=1}^{q} \alpha_i \varepsilon_{t-i}^2 + \sum_{j=1}^{p} \beta_j \sigma_{t-j}^2, \tag{2}$$

where

$$p \geq 0, \qquad q > 0,$$

$$\omega > 0, \qquad \alpha_i \geq 0,\ i = 1, \ldots, q, \qquad \beta_j \geq 0,\ j = 1, \ldots, p. \tag{3}$$

Also another discussion of the $GARCH\,(p,q)$ process definition has appeared in [2] that distinguishes three forms of $GARCH\,(p,q)$ processes, namely strong, semi-strong and weak form as follows:

Definition 2 *The process ε_t, $t \in Z$, is $GARCH\,(p,q)$, if*

$$E\left(\varepsilon_t | F_{t-1}\right) = 0,$$

$$\sigma_t^2 = \omega + \sum_{i=1}^{q} \alpha_i \varepsilon_{t-i}^2 + \sum_{j=1}^{p} \beta_j \sigma_{t-j}^2,$$

and

- $\operatorname{var}\left(\varepsilon_t | F_{t-1}\right) = \sigma_t^2$ *and* $Z_t = \varepsilon_t / \sigma_t$ *is i.i.d. (strong GARCH)*
- $\operatorname{var}\left(\varepsilon_t | F_{t-1}\right) = \sigma_t^2$ *(semi-strong GARCH), or*
- $P\left(\varepsilon_t^2 | 1, \varepsilon_{t-1}, \varepsilon_{t-2}, \ldots, \varepsilon_{t-1}^2, \varepsilon_{t-2}^2, \ldots\right) = \sigma_t^2$ *(weak GARCH),*
  where P is the best linear projection.

Remark 1 *Setting $Z_t = \varepsilon_t / \sigma_t$, it holds for both strong and semi-strong $GARCH$ models that*

$$\begin{aligned} E(Z_t) &= E(\varepsilon_t/\sigma_t) = E(E(\varepsilon_t/\sigma_t | F_{t-1})) \\ &= E(E(\varepsilon_t | F_{t-1})/\sigma_t) = 0, \end{aligned} \tag{4}$$

and

$$\begin{aligned} \operatorname{var}(Z_t) &= E\left(Z_t^2\right) - E^2(Z_t) = E\left(Z_t^2\right) = E\left(\varepsilon_t^2/\sigma_t^2\right) \\ &= E\left(E\left(\varepsilon_t^2/\sigma_t^2 | F_{t-1}\right)\right) = E\left(E\left(\varepsilon_t^2 | F_{t-1}\right)/\sigma_t^2\right) = 1. \end{aligned} \tag{5}$$

Under the assumption that Z_t is normally distributed, (1) allows the difference between strong and semi-strong $GARCH$ models to disappear.

Since this paper is not intended to discuss weak models, Definition 1 of $GARCH$ models will be used throughout the following.

Remark 2 *Assuming without loss of generality that $s > 0$, we notice*

$$\begin{aligned} \operatorname{cov}(Z_t, Z_{t+s}) &= E(Z_t Z_{t+s}) = E(E(\varepsilon_t \varepsilon_{t+s}/(\sigma_t \sigma_{t+s}) | F_{t+s-1})) \\ &= E(\varepsilon_t E(\varepsilon_{t+s} | F_{t+s-1})/(\sigma_t \sigma_{t+s})) = 0, \end{aligned}$$

and this, in conjunction with (4) and (5), shows that

$$Z_t \sim N(0,1)$$

implies

$$Z_t \sim i.i.d.\ N(0,1).$$

Remark 3 *The inequality restrictions (3) are imposed to guarantee*

$$\sigma_t^2 > 0, \tag{6}$$

almost surely. In the case of $GARCH(1,1)$ model with

$$\sigma_t^2 = \omega + \alpha_1 \varepsilon_{t-1}^2 + \beta_1 \sigma_{t-1}^2, \tag{7}$$

obviously, the sufficient but not necessary conditions for (6) are

$$\omega > 0, \qquad \alpha_1 \geq 0, \qquad \text{and } \beta_1 \geq 0. \tag{8}$$

On the other hand, subsequent substitutions show that

$$\begin{aligned} \sigma_t^2 &= \omega + \alpha_1 \varepsilon_{t-1}^2 + \beta_1 \left(\omega + \alpha_1 \varepsilon_{t-2}^2 + \beta_1 \sigma_{t-2}^2\right) \\ &\vdots \\ &= \omega (1-\beta_1)^{-1} + \alpha_1 \varepsilon_{t-1}^2 + \alpha_1 \beta_1 \sum_{j=0}^{\infty} \beta_1^j \varepsilon_{t-j-2}^2, \end{aligned}$$

where $\omega > 0$, $\alpha_1 \geq 0$ and $0 \leq \beta_1 < 1$ are necessary and also sufficient conditions for (6) assuming that the sum $\sum_{j=0}^{\infty} \beta_1^j \varepsilon_{t-j-2}^2$ converges.

2.2 Stationarity Condition

Bollerslev(1986) established the necessary and sufficient condition for existence of a weakly stationary solution ε_t to the $GARCH$ model. To state the theorem below, we first recall the definition of weak stationarity.

Definition 3 *The process ε_t, $t \in Z$, is weakly stationary if*

(i) $E(\varepsilon_t)$ *is independent of t, and*

(ii) $\operatorname{cov}(\varepsilon_{t+h}, \varepsilon_t)$ *is independent of t for each h.*

Throughout the following, whenever we use the term stationary we shall mean weakly stationary as in Definition 3, unless we specifically indicate otherwise.

Theorem 1 *The $GARCH(p,q)$ process as defined in (1) and (2) is stationary with $E(\varepsilon_t) = 0$, $\operatorname{var}(\varepsilon_t) = \omega \left(1 - \sum_{i=1}^{q} \alpha_i - \sum_{j=1}^{p} \beta_j\right)^{-1}$ and $\operatorname{cov}(\varepsilon_t, \varepsilon_s) = 0$ for $t \neq s$ if and only if*

$$\sum_{i=1}^{q} \alpha_i + \sum_{j=1}^{p} \beta_j < 1. \tag{9}$$

The proof can be found in [1], and hence is omitted here. However, we illustrate basic ideas of this proof in the following with the often used $GARCH(1,1)$ model.

Example 1 *Now an analogue of Theorem 1 can be formulated. For the $GARCH(1,1)$ process given by (7),(8) a necessary and sufficient condition for stationarity is*

$$\alpha_1 + \beta_1 < 1. \tag{10}$$

Proof: *First, notice that, the process ε_t obtained from $GARCH(1,1)$ model satisfying*

$$\varepsilon_t = \sigma_t Z_t, \qquad Z_t \sim i.i.d.N(0,1),$$

and use recursive substitution to show that

$$\begin{aligned}
\sigma_t^2 &= \omega + \alpha_1 Z_{t-1}^2 \sigma_{t-1}^2 + \beta_1 \sigma_{t-1}^2 \\
&= \omega + \alpha_1 Z_{t-1}^2 \left(\omega + \alpha_1 Z_{t-2}^2 \sigma_{t-2}^2 + \beta_1 \sigma_{t-2}^2\right) \\
&\quad + \beta_1 \left(\omega + \alpha_1 Z_{t-2}^2 \sigma_{t-2}^2 + \beta_1 \sigma_{t-2}^2\right) \\
&\vdots \\
&= \omega \sum_{k=0}^{\infty} M(t,k),
\end{aligned} \tag{11}$$

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

$$\alpha_1^a \beta_1^b \prod_{l=1}^{a} Z_{t-S_l}^2,$$

for

$$a+b=k, \qquad 1 \leq S_1 < S_2 < \ldots < S_a \leq k.$$

Since Z_t *is i.i.d., the expected values of M(t,k) do not depend on t, that is,*

$$E\left(M\left(t,k\right)\right) = E\left(M\left(s,k\right)\right) \text{ for all } k,\ t,\ s. \tag{12}$$

Moreover,

$$\begin{aligned} M\left(t,0\right) &= 1, \\ M\left(t,1\right) &= \alpha_1 Z_{t-1}^2 + \beta_1, \\ M\left(t,2\right) &= \left(\alpha_1 Z_{t-1}^2 + \beta_1\right)\left(\alpha_1 Z_{t-2}^2 + \beta_1\right), \end{aligned}$$

and in general

$$M\left(t,k+1\right) = \left(\alpha_1 Z_{t-1}^2 + \beta_1\right) M\left(t-1,k\right),$$

which yields, together with (12), that

$$\begin{aligned} E\left(M\left(t,k+1\right)\right) &= \left(\alpha_1+\beta_1\right) E\left(M\left(t,k\right)\right) \\ &\vdots \\ &= \left(\alpha_1+\beta_1\right)^{k+1} E\left(M\left(t,0\right)\right) \\ &= \left(\alpha_1+\beta_1\right)^{k+1}. \end{aligned} \tag{13}$$

Finally by (1), (11) and (13),

$$\begin{aligned} E\left(\varepsilon_t^2\right) &= \omega E\left(\sum_{k=0}^{\infty} M\left(t,k\right)\right) \\ &= \omega \sum_{k=0}^{\infty} E\left(M\left(t,k\right)\right) \\ &= \omega \sum_{k=0}^{\infty} 1 \Big/ \left(\alpha_1+\beta_1\right)^{k+1}, \end{aligned}$$

it then follows that

$$E\left(\varepsilon_t^2\right) = \omega\left(1-\alpha_1-\beta_1\right)^{-1}$$

if and only if (10) holds and ε_t^2 *converges almost surely. Then* $E\left(\varepsilon_t\right)=0$ *and* $\operatorname{cov}\left(\varepsilon_t,\varepsilon_s\right)=0$ *for* $t \neq s$ *follows by symmetry.*

$\square$

2.3 ARMA Representation

As pointed out in [2, Theorem 12.9], the GARCH process ε_t can be interpreted as an ARMA process in ε_t^2, however, whose proof is omitted. Here we review the theorem and will present the proof in full detail later.

Theorem 2 *Let ε_t be a stationary $GARCH\,(p,q)$ process with $E\left(\varepsilon_t^4\right) = const < \infty$ and $Z_t = \varepsilon_t/\sigma_t$. It holds that*

(i) $\eta_t = \sigma_t^2\left(Z_t^2 - 1\right)$ *is white noise.*

(ii) ε_t^2 *is an $ARMA\,(m,p)$ process with*

$$\varepsilon_t^2 = \omega + \sum_{i=1}^{m} \gamma_i \varepsilon_{t-i}^2 - \sum_{j=1}^{p} \beta_j \eta_{t-j} + \eta_t, \tag{14}$$

where $m = max\,(p,q)$, $\gamma_i = \alpha_i + \beta_i$, $\alpha_i = 0$ when $i > q$, and $\beta_i = 0$ when $i > p$.

We first state Lemma 1 and the proof from [4] for easy reference.

Lemma 1 *For $\beta_j \geq 0$, $1 \leq j \leq p-1$ and $\beta_p \neq 0$, relation*

$$\beta_1 + \beta_2 + \ldots + \beta_p < 1 \tag{15}$$

is equivalent to

$$|\zeta_i| > 1 \text{ for all } 1 \leq i \leq l, \tag{16}$$

where $\zeta_1, \zeta_2, \ldots, \zeta_l$ stand for the solutions of

$$B\,(z) = 1 - \beta_1 z - \beta_2 z^2 - \ldots - \beta_p z^p = 0 \tag{17}$$

with multiplicities $\nu_1, \ldots, \nu_l$.

Proof: Let us assume first that $\beta_1 + \beta_2 + \ldots + \beta_p \geq 1$. Since $B\,(0) = 1$ and

$$B\,(1) = 1 - (\beta_1 + \beta_2 + \ldots + \beta_p) \leq 0,$$

we have at least one solution of (17) in the interval $(0,1]$, contradicting (16).

Let us assume now that (15) holds. Then, for any $|z| \leq 1$, we have

$$\begin{aligned} |B\,(z)| &\geq 1 - \left(\beta_1 |z| + \beta_2 |z|^2 + \ldots + \beta_p |z|^p\right) \\ &\geq 1 - (\beta_1 + \beta_2 + \ldots + \beta_p) > 0, \end{aligned}$$

and therefore (16) must be true.

□

As a preliminary step we prove the following result:

Lemma 2 *Let ε_t be a stationary GARCH (p,q) process. Then there exists a constant C and a sequence of constants $\{\psi_j\}$ such that*

$$\sigma_t^2 = C + \sum_{j=0}^{\infty} \psi_j \varepsilon_{t-j}^2 \tag{18}$$

for all t.

Proof: Let $\varphi(z) = 1 - \sum_{j=1}^{p} \beta_j z^j$ and $\theta(z) = \sum_{i=1}^{q} \alpha_i z^i$. Using the backward shift operator B, (2) is equivalently written as

$$\varphi(B)\,\sigma_t^2 = \omega + \theta(B)\,\varepsilon_t^2. \tag{19}$$

By Theorem 1 and $\alpha_i \geq 0$, $\beta_j \geq 0$, it follows easily from (9) that $\sum_{j=1}^{p} \beta_j < 1$. Since this relation entails, by Lemma 1, that all roots of $\varphi(z) = 0$ lie outside of the unit circle, we have

$$\frac{1}{\varphi(z)} = \sum_{j=0}^{\infty} d_j z^j \text{ for } |z| \leq 1,$$

and $\sum_{j=0}^{\infty} |d_j| < \infty$. Now we can define $\varphi^{-1}(B)$ as

$$\varphi^{-1}(B) := \frac{1}{\varphi(B)} = \sum_{j=0}^{\infty} d_j B^j.$$

Applying the operator $\varphi^{-1}(B)$ to both sides of (19) gives

$$\begin{aligned} \sigma_t^2 &= \varphi^{-1}(1)\,\omega + \varphi^{-1}(B)\,\theta(B)\,\varepsilon_t^2 \\ &= \underbrace{\omega\left(1 - \sum_{j=1}^{p} \beta_j\right)^{-1}}_{const.} + \sum_{j=0}^{\infty} \psi_j \varepsilon_{t-j}^2, \end{aligned}$$

in which ψ_j's are found from the power series expansion of $\varphi^{-1}(B)\,\theta(B)$, i.e.,

$$\psi_j = \alpha_j + \sum_{k=1}^{p} \beta_k \psi_{j-k},\ j \geq 0,$$

where $\alpha_0 := 0$, $\alpha_j := 0$ for $j > q$, and $\psi_j := 0$ for $j < 0$. The proof is complete.

□

We now turn to the proof of Theorem 2.

PROOF: To prove (i) first note that, by Definition 1,

$$E\left(\varepsilon_t|F_{t-1}\right)=0, \quad \operatorname{var}\left(\varepsilon_t|F_{t-1}\right)=\sigma_t^2.$$

This gives

$$E\left(\varepsilon_t^2|F_{t-1}\right)=\operatorname{var}\left(\varepsilon_t|F_{t-1}\right)+E^2\left(\varepsilon_t|F_{t-1}\right)=\sigma_t^2. \tag{20}$$

Note also that, in view of (4) and (5),

$$E\left(Z_t^2\right)=\operatorname{var}\left(Z_t\right)+E^2\left(Z_t\right)=1. \tag{21}$$

We decompose the proof into three parts. Firstly, using $Z_t=\varepsilon_t/\sigma_t$, we show η_t is a zero-mean process:

$$\begin{aligned}
E\left(\eta_t\right) &= E\left(\sigma_t^2\left(Z_t^2-1\right)\right)=E\left(E\left(\sigma_t^2\left(Z_t^2-1\right)|F_{t-1}\right)\right)\\
&= E\left(\sigma_t^2 E\left(\left(Z_t^2-1\right)|F_{t-1}\right)\right)=E\left(\sigma_t^2 E\left(\varepsilon_t^2/\sigma_t^2|F_{t-1}\right)\right)-E\left(\sigma_t^2\right)\\
&= E\left(E\left(\varepsilon_t^2|F_{t-1}\right)\right)-E\left(\sigma_t^2\right)=E\left(\sigma_t^2\right)-E\left(\sigma_t^2\right)=0.
\end{aligned}$$

Next we prove that $\operatorname{var}\left(\eta_t\right)$ does not depend on t. By (21) and $\varepsilon_t=\sigma_t Z_t$,

$$\begin{aligned}
\operatorname{var}\left(\eta_t\right) &= E\left(\eta_t^2\right)-E^2\left(\eta_t\right)=E\left(\eta_t^2\right)=E\left(\sigma_t^4\left(Z_t^2-1\right)^2\right)\\
&= E\left(E\left(\sigma_t^4\left(Z_t^2-1\right)^2|F_{t-1}\right)\right)=E\left(\sigma_t^4 E\left(\left(Z_t^4-2Z_t^2+1\right)|F_{t-1}\right)\right)\\
&= E\left(\sigma_t^4 E\left(\varepsilon_t^4/\sigma_t^4|F_{t-1}\right)-2\sigma_t^4 E\left(Z_t^2|F_{t-1}\right)+\sigma_t^4\right)\\
&= E\left(\varepsilon_t^4-\sigma_t^4\right),
\end{aligned}$$

in which we substitute (18) to obtain

$$\begin{aligned}
\operatorname{var}\left(\eta_t\right) &= E\left(\varepsilon_t^4\right)-E\left(C+\sum_{j=0}^{\infty}\psi_j\varepsilon_{t-j}^2\right)^2\\
&= E\left(\varepsilon_t^4\right)-\sum_{j=0}^{\infty}\sum_{k=0}^{\infty}\psi_j\psi_k E\left(\varepsilon_{t-j}^2\varepsilon_{t-k}^2\right)-2C\sum_{j=0}^{\infty}\psi_j E\left(\varepsilon_{t-j}^2\right)-C^2.
\end{aligned}$$

Therefore we will need to prove that

$$E\left(\varepsilon_{t-j}^2\right) \text{ and } E\left(\varepsilon_{t-j}^2\varepsilon_{t-k}^2\right) \text{ are independent of t.} \tag{22}$$

Note that under Definition 1, ε_t is normally distributed, together with condition of stationarity, as it can be shown, which yields

$$\left(\varepsilon_t,\varepsilon_{t+h}\right)'\sim\left(\varepsilon_1,\varepsilon_{1+h}\right)' \quad \text{for all } t \text{ and } h,$$

that is, $\left(\varepsilon_t,\varepsilon_{t+h}\right)$ and $\left(\varepsilon_1,\varepsilon_{1+h}\right)$ have the same joint distributions.

We already know from mathematical definition of expected value that two variables with the same probability distribution will have the same expected value, from which we readily obtain (22).

Finally we show that the process η_t is uncorrelated. By (21),

$$\begin{aligned}
\operatorname{cov}(\eta_t, \eta_{t+s}) &= E\left(\left(\eta_t - E(\eta_t)\right)\left(\eta_{t+s} - E(\eta_{t+s})\right)\right) \\
&= E(\eta_t \eta_{t+s}) = E\left(E\left(\sigma_t^2\left(Z_t^2 - 1\right)\sigma_{t+s}^2\left(Z_{t+s}^2 - 1\right)|F_{t+s-1}\right)\right) \\
&= E\left(\sigma_t^2\left(Z_t^2 - 1\right)\sigma_{t+s}^2 E\left(\left(Z_{t+s}^2 - 1\right)|F_{t+s-1}\right)\right) = 0,
\end{aligned}$$

without loss of generality, for $s > 0$. The proof of (i) is thus complete.

We now prove (ii). It follows from rearranging that

$$\begin{aligned}
\varepsilon_t^2 &= \sigma_t^2 Z_t^2 = \sigma_t^2 + \sigma_t^2\left(Z_t^2 - 1\right) \\
&= \omega + \sum_{i=1}^{q} \alpha_i \varepsilon_{t-i}^2 + \sum_{j=1}^{p} \beta_j \sigma_{t-j}^2 + \eta_t \\
&= \omega + \sum_{i=1}^{q} \alpha_i \varepsilon_{t-i}^2 + \sum_{k=1}^{m} \beta_k \varepsilon_{t-k}^2 - \sum_{k=1}^{m} \beta_k \sigma_{t-k}^2 Z_{t-k}^2 + \sum_{j=1}^{p} \beta_j \sigma_{t-j}^2 + \eta_t \\
&= \omega + \sum_{i=1}^{m} (\alpha_i + \beta_i)\varepsilon_{t-i}^2 - \sum_{j=1}^{p} \beta_j \sigma_{t-j}^2\left(Z_{t-j}^2 - 1\right) + \eta_t \\
&= \omega + \sum_{i=1}^{m} \gamma_i \varepsilon_{t-i}^2 - \sum_{j=1}^{p} \beta_j \eta_{t-j} + \eta_t,
\end{aligned}$$

where $m = max(p, q)$, $\gamma_i = \alpha_i + \beta_i$, $\alpha_i = 0$ when $i > q$, and $\beta_i = 0$ when $i > p$. It remains to prove that ε_t^2 is stationary. By Theorem 1 we have

$$\sum_{i=1}^{m} \gamma_i = \sum_{i=1}^{q} \alpha_i + \sum_{i=1}^{p} \beta_i < 1,$$

where this inequality entails, by Lemma 1, that all roots of

$$\phi(z) = 1 - \gamma_1 z - \gamma_2 z^2 - \ldots - \gamma_m z^m = 0$$

lie outside of the unit circle, which is equivalent to the causality condition described in [3], that is,

$$\phi(z) \neq 0 \quad \text{for all } |z| \leq 1,$$

from which by the existence and uniqueness theorem in [3] we can conclude that ε_t^2 is a stationary solution of equations (14) and that is also causal, completing the proof of (ii).

□

As an immediate consequence of Theorem 2 and Lemma 2 we have:

Corollary 1 *Let ε_t be a stationary $GARCH\,(p,q)$ process. Then for the ARMA(m,p) process ε_t^2 satisfying (14), there exists a constant*

$$C = \omega \left(1 - \sum_{j=1}^{m} \gamma_j\right)^{-1},$$

and a sequence of constants

$$\psi_j = -\beta_j + \sum_{k=1}^{m} \gamma_k \psi_{j-k},\ j \geq 0, \tag{23}$$

where $\beta_0 := -1,\ \beta_j := 0$ for $j > p$, and $\psi_j := 0$ for $j < 0$, such that

$$\varepsilon_t^2 = C + \sum_{j=0}^{\infty} \psi_j \eta_{t-j}$$

for all t.

Example 2 *For the GARCH(1,1) model in Example 1, ε_t^2 is an ARMA(1,1) process with*

$$\begin{aligned}
\varepsilon_t^2 &= \left(\omega + \alpha_1 \varepsilon_{t-1}^2 + \beta_1 \sigma_{t-1}^2\right) + \eta_t \\
&= \left(\omega + (\alpha_1 + \beta_1)\,\varepsilon_{t-1}^2 - \beta_1\left(\sigma_{t-1}^2 Z_{t-1}^2 - \sigma_{t-1}^2\right)\right) + \eta_t \\
&= \omega + (\alpha_1 + \beta_1)\,\varepsilon_{t-1}^2 - \beta_1 \eta_{t-1} + \eta_t.
\end{aligned}$$

3 Parameter estimation

3.1 Yule-Walker estimator

Based on the $ARMA$ representation of $GARCH$ processes, parameter estimation can be carried out conveniently by utilizing existing Yule-Walker equations. We first outline the idea of Yule-Walker estimation with $ARCH\,(q)$ model as a special case of $GARCH\,(p,q)$ where $p = 0$. Consider the stationary $ARCH\,(q)$ process $\bar{\varepsilon}_t$ defined by

$$\sigma_t^2 = \omega + \alpha_1 \bar{\varepsilon}_{t-1}^2 + \ldots + \alpha_q \bar{\varepsilon}_{t-q}^2,$$

whose $AR\,(q)$ representation in $\bar{\varepsilon}_t^2$ is

$$\bar{\varepsilon}_t^2 - \alpha_1 \bar{\varepsilon}_{t-1}^2 - \ldots - \alpha_q \bar{\varepsilon}_{t-q}^2 = \omega + \eta_t, \tag{24}$$

where η_t has the same definition as in Theorem 2.

Now we describe the main steps in the Yule-Walker estimation procedure in more detail.

Step 1. Transform a nonzero mean process to a zero-mean one. For a stationary process we shall use the notation $\mu := E\left(\bar{\varepsilon}_t^2\right)$. Taking the expectation on both sides of (24), we have

$$\mu = \omega \left(1 - \sum_{i=1}^{q} \alpha_i\right)^{-1}.$$

Now we denote $\bar{\varepsilon}_t^2 - \mu$ as ε_t^2 to obtain a zero-mean, stationary process satisfying

$$\varepsilon_t^2 - \alpha_1 \varepsilon_{t-1}^2 - \ldots - \alpha_q \varepsilon_{t-q}^2 = \eta_t. \tag{25}$$

Step 2. Express ε_t^2 in terms of $\eta_s, s \leq t$. It can be obtained from Corollary 1 that

$$\varepsilon_t^2 = \sum_{j=0}^{\infty} \psi_j \eta_{t-j}, \qquad \psi_0 = 1, \qquad \psi_j = \sum_{k=1}^{q} \alpha_k \psi_{j-k} \text{ for } j \geq 1, \tag{26}$$

where $\psi_j := 0$ for $j < 0$.

Step 3. Derive the Yule-Walker equations. Multiplying each side of (25) by ε_{t-k}^2, $k = 0, 1, \ldots, q$, taking expectations and using (26) to evaluate the right-hand side of (25), we obtain the Yule-Walker equations

$$\gamma(k) - \alpha_1 \gamma(k-1) - \ldots - \alpha_q \gamma(k-q) = \begin{cases} \sigma^2, & k = 0 \\ 0, & 0 < k \leq q \end{cases}, \tag{27}$$

where $\gamma(k) := \operatorname{cov}\left(\varepsilon_t^2, \varepsilon_{t-k}^2\right)$ and $\sigma^2 := \operatorname{var}(\eta_t)$.

Step 4. Solve for the Yule-Walker estimators $\hat{\alpha}_1, \ldots, \hat{\alpha}_q$ and $\hat{\sigma}^2$, by calculating the $q+1$ linear equations (27) with sample estimates

$$\hat{\gamma}(k) = n^{-1} \sum_{t=k+1}^{n} \left(\varepsilon_t^2 - \hat{\mu}\right)\left(\varepsilon_{t-k}^2 - \hat{\mu}\right), \tag{28}$$

where $\hat{\mu} = n^{-1} \sum_{t=1}^{n} \varepsilon_t^2$.

Remark 4 *Stationarity is an important part required in the Yule-Walker estimation. It is necessary for the existence of $MA(\infty)$ representation of ε_t^2 in Step 2. It also guarantees that we can consistently estimate $\gamma(k)$ in Step 4.*

Next we extend the previous result to a more general case where $p > 0$. Since the steps involved are analogous, we skip some details in subsequent description and concentrate on the the key points.

We start with the transformation of (14), which as shown in Theorem 2 is the ARMA representation of a stationary $GARCH\,(p,q)$ model, to have a zero-mean process

$$\varepsilon_t^2 - \sum_{i=1}^{m} \gamma_i \varepsilon_{t-i}^2 = \eta_t - \sum_{j=1}^{p} \beta_j \eta_{t-j},$$

and by Lemma 1, we write ε_t^2 in the form $\varepsilon_t^2 = \sum_{j=0}^{\infty} \psi_j \eta_{t-j}$, where ψ_j's are found directly from (23). Proceeding as in Step 3, we find the Yule-Walker equations

$$\gamma\,(k) - \gamma_1 \gamma\,(k-1) - \ldots - \gamma_m \gamma\,(k-m) = \sigma^2 \sum_{j=k}^{p} (-\beta_j)\, \psi_{j-k}, \tag{29}$$

for $0 \leq k \leq m+p$, from which the unknown coefficients can be solved with the sample covariances.

Although the Yule-Walker estimation can be adapted to ARMA models, the corresponding equations are nonlinear in the unknown coefficients, as the following example reveals.

Example 3 *Again consider the $GARCH\,(1,1)$ model in Example 1, the Yule-Walker equations obtained from (29) for $k = 0,1,2$ are*

$$\begin{aligned}
\hat{\gamma}\,(0) - \left(\hat{\alpha}_1 + \hat{\beta}_1\right) \hat{\gamma}\,(1) &= \left(1 - \hat{\alpha}_1 \hat{\beta}_1\right) \hat{\sigma}^2, \\
\hat{\gamma}\,(1) - \left(\hat{\alpha}_1 + \hat{\beta}_1\right) \hat{\gamma}\,(0) &= -\hat{\beta}_1 \hat{\sigma}^2, \\
\hat{\gamma}\,(2) - \left(\hat{\alpha}_1 + \hat{\beta}_1\right) \hat{\gamma}\,(1) &= 0.
\end{aligned}$$

Simple algebra shows that

$$\hat{\beta}_1^2 - \frac{\rho_2^2 - 2\rho_1\rho_2 + 1}{\rho_2 - \rho_1} \hat{\beta}_1 + 1 = 0, \qquad \hat{\alpha}_1 = \rho_2 - \hat{\beta}_1,$$

where $\rho_1 := \hat{\gamma}\,(1)/\hat{\gamma}\,(0)$ and $\rho_2 := \hat{\gamma}\,(2)/\hat{\gamma}\,(1)$.

We solve the quadratic equation for $\hat{\beta}_1$, which would require much computing time and lead to possible nonexistence and nonuniqueness of solution.

3.2 Maximum likelihood estimator

A more common approach to parameter estimation is Maximum Likelihood method. Here we review the general estimation strategy of ML technique. For a given set of observations $(x_1, x_2, \ldots, x_n)$ drawn from a probability distribution associated with a known density function f parameterized by θ, the likelihood function $L\,(\theta)$ is

$$L\,(\theta|\,(x_1, x_2, \ldots, x_n)) = f\,((x_1, x_2, \ldots, x_n)\,|\theta)\,,$$

viewed as a function of θ with $x_1, x_2, \ldots, x_n$ fixed. A ML estimator of θ is then defined as

$$\hat{\theta} = \arg\max_{\theta \in \Theta} L(\theta). \tag{30}$$

In other words, the ML estimator is the value of θ that maximizes the probability of the observed sample.

For likelihood maximization one frequently uses numerical iterative methods, where a likelihood function is often replaced by its natural logarithm for computational convenience, which of course does not change the location of the maximum. Many algorithms exist for solving such problem, here we briefly recall score algorithm:

Let θ_j denote the parameter estimates after the jth iteration. θ_{j+1} is then calculated from

$$\theta_{j+1} = \theta_j + J^{-1}(\theta_j) \nabla L(\theta_j), \tag{31}$$

with the gradient

$$\nabla L = \frac{\partial L}{\partial \theta}, \tag{32}$$

and the Fisher Information matrix

$$J = E\left(-\frac{\partial^2 L}{\partial \theta \partial \theta^T}\right). \tag{33}$$

Let us now look at the application of ML estimation in the case $p = 0$ and then, as before, we turn to the $GARCH(p,q)$ model with $p > 0$.

For the $ARCH(q)$ model defined by $\sigma_t^2 = \omega + \sum_{i=1}^{q} \alpha_i \varepsilon_{t-i}^2$, assuming that ε_t is conditionally normally distributed, we have

$$P(\varepsilon_t | F_{t-1}) = \frac{1}{\sqrt{2\pi}\sigma_t} \exp\left\{-\frac{1}{2}\frac{\varepsilon_t^2}{\sigma_t^2}\right\}.$$

The log-likelihood function L of parameter vector $\theta = (\omega, \alpha_1, \ldots, \alpha_q)^T$ can be written as

$$L(\theta) = \sum_{t=q+1}^{n} l_t(\theta), \tag{34}$$

where

$$\begin{aligned} l_t(\theta) &= \log P(\varepsilon_t | F_{t-1}) \\ &= -\frac{1}{2}\log(2\pi) - \frac{1}{2}\log\sigma_t^2 - \frac{1}{2}\frac{\varepsilon_t^2}{\sigma_t^2}. \end{aligned}$$

Therefore, the derivatives of l_t are given by

$$\begin{aligned} \frac{\partial l_t}{\partial \theta} &= \frac{1}{2\sigma_t^2}\frac{\partial \sigma_t^2}{\partial \theta}\left(\frac{\varepsilon_t^2}{\sigma_t^2} - 1\right), \\ \frac{\partial^2 l_t}{\partial \theta \partial \theta^T} &= \frac{1}{2\sigma_t^2}\left(\frac{\varepsilon_t^2}{\sigma_t^2} - 1\right)\frac{\partial^2 \sigma_t^2}{\partial \theta \partial \theta^T} + \frac{1}{\sigma_t^4}\left(\frac{1}{2} - \frac{\varepsilon_t^2}{\sigma_t^2}\right)\frac{\partial \sigma_t^2}{\partial \theta}\frac{\partial \sigma_t^2}{\partial \theta^T}, \end{aligned}$$

where

$$\frac{\partial \sigma_t^2}{\partial \theta} = \left(1, \varepsilon_{t-1}^2, \ldots, \varepsilon_{t-q}^2\right)^T.$$

This allows us to construct the gradient

$$\nabla L = \frac{1}{2} \sum_{t=q+1}^{n} \frac{1}{\sigma_t^2} \frac{\partial \sigma_t^2}{\partial \theta} \left(\frac{\varepsilon_t^2}{\sigma_t^2} - 1 \right), \tag{35}$$

and the Fisher Information matrix

$$J = \frac{1}{2} \sum_{t=q+1}^{n} E \left(\frac{1}{\sigma_t^4} \frac{\partial \sigma_t^2}{\partial \theta} \frac{\partial \sigma_t^2}{\partial \theta^T} \right) \tag{36}$$

employed in the updating scheme (31).

Example 4 *As an application, let us consider an $ARCH\left(1\right)$ model with $\sigma_t^2 = \omega + \alpha_1 \varepsilon_{t-1}^2$. For the optimization of the log-likelihood function $L\left(\omega, \alpha_1\right)$, we carry out the iterations (31) with*

$$\nabla L = \frac{1}{2} \sum_{t=2}^{n} \frac{1}{\sigma_t^2} \left(\frac{\varepsilon_t^2}{\sigma_t^2} - 1 \right) \begin{bmatrix} 1 \\ \varepsilon_{t-1}^2 \end{bmatrix},$$

and

$$J = \frac{1}{2} \sum_{t=2}^{n} E \left(\frac{1}{\sigma_t^4} \right) \begin{bmatrix} 1 & \varepsilon_{t-1}^2 \\ \varepsilon_{t-1}^2 & \varepsilon_{t-1}^4 \end{bmatrix},$$

where $E\left(\sigma_t^{-4}\right)$ is estimated by $(n-1)^{-1} \sum_{t=2}^{n} \left(\omega + \alpha_1 \varepsilon_{t-1}^2\right)^{-2}$.

Having made this introduction, we continue to the $GARCH\left(p, q\right)$ model, which could be handled in a similar way. Under the assumption of conditionally normal distribution, we write the log-likelihood function L with the same notation as in (34) of parameter vector $\theta = \left(\omega, \alpha_1, \ldots, \alpha_q, \beta_1, \ldots, \beta_p\right)^T$. When computing the ML estimator, the score-algorithm (31) is still valid with the gradient (35) and the Fisher Information matrix (36), however, in which

$$\frac{\partial \sigma_t^2}{\partial \theta} = \upsilon_t + \sum_{j=1}^{p} \beta_j \frac{\partial \sigma_{t-j}^2}{\partial \theta}, \tag{37}$$

where

$$\upsilon_t = \left(1, \varepsilon_{t-1}^2, \ldots, \varepsilon_{t-q}^2, \sigma_{t-1}^2, \ldots, \sigma_{t-p}^2\right)^T.$$

The only difference from the $ARCH\left(q\right)$ regression model is the inclusion of the recursive part in (37).

Example 5 *Now consider* $GARCH\,(1,1)$ *process with* $\sigma_t^2 = \omega + \alpha_1 \varepsilon_{t-1}^2 + \beta_1 \sigma_{t-1}^2$, *to estimate the unknown coefficients* $\theta = (\omega, \alpha_1, \beta_1)^T$, *we use iteration (31) with* J *and* ∇L *as defined above, where*

$$\frac{\partial \sigma_t^2}{\partial \theta} = \upsilon_t + \beta_1 \frac{\partial \sigma_{t-1}^2}{\partial \theta}, \quad \upsilon_t = \left(1, \varepsilon_{t-1}^2, \sigma_{t-1}^2\right)^T. \tag{38}$$

Remark 5 *From [2] it follows that, under the conditions*

$$E\left(Z_t | F_{t-1}\right) = 0, \quad E\left(Z_t^2 | F_{t-1}\right) = 1, \quad E\left(Z_t^4 | F_{t-1}\right) < \infty, \tag{39}$$

and strict stationarity of ε_t, *the ML estimator* $\hat{\theta}$ *is strongly consistent and asymptotically normally distributed:*

$$\sqrt{n}\left(\hat{\theta} - \theta\right) \to N_{p+q+1}\left(0, J^{-1} I J^{-1}\right)$$

with

$$I = E\left(\frac{\partial l_t}{\partial \theta} \frac{\partial l_t}{\partial \theta^T}\right), \qquad J = -E\left(\frac{\partial^2 l_t}{\partial \theta \partial \theta^T}\right).$$

Moreover replacing (39) with a stronger condition

$$Z_t \sim N\left(0, 1\right),$$

as proved in [2] that $I = J$, *the asymptotic variance can be simplified to* J^{-1}, *that is,*

$$\sqrt{n}\left(\hat{\theta} - \theta\right) \to N_{p+q+1}\left(0, J^{-1}\right).$$

4 Simulated example

To evaluate the performance of the estimators discussed, a simulation study has been carried out. We would like to know if the computed parameter estimates are as expected for data series generated from the following $GARCH\,(1,1)$ process:

$$\omega = 0.1, \qquad \alpha_1 = 0.1, \qquad \beta_1 = 0.8.$$

It was found in [5] that at least 200-300 observations were required to reliably model a GARCH process, but in fact many more are needed as shown in Figure 1. We see that as the number of observations increases from 200 to 800 the precision of ML estimates increases dramatically with a fixed number of simulations (200 times). Therefore in the following our simulations will be performed with a sample size of 800.

We also need to assess the necessary number of simulation runs at each parameter estimation. The result is given in Figure 2. It appears that the effect of an increase in number on the estimation accuracy is trivial, especially when the

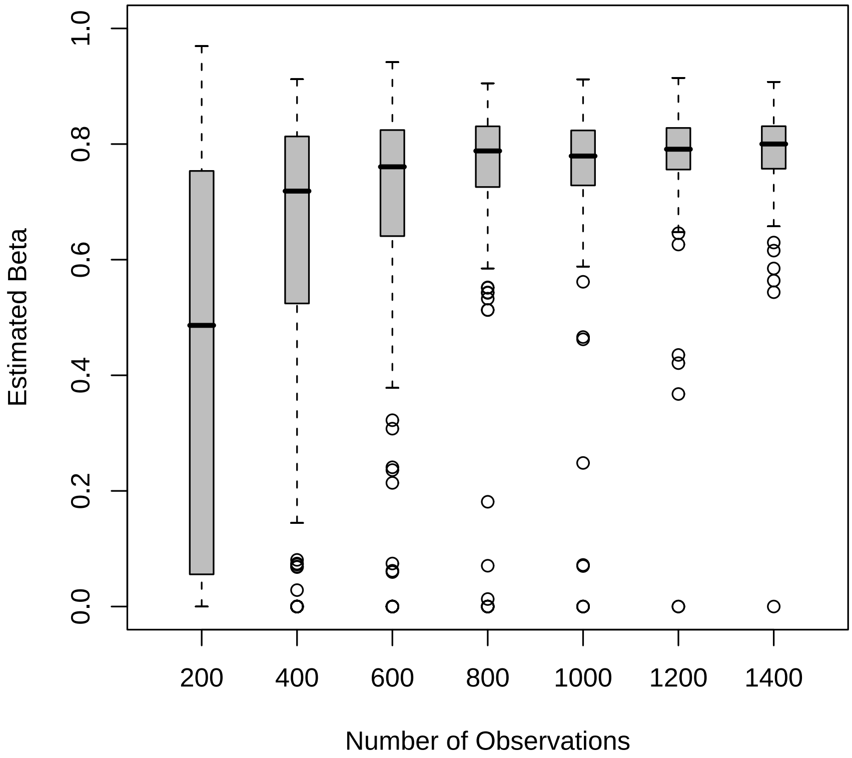


Figure 1: A box-plot of estimated β_1.

run number reaches 400. To achieve better accuracy even slightly, thereafter the number of simulations will be 400 instead of 200 in the above experiment.

Now an experiment designed to examine estimation performance is implemented by varying the value of β_1 from 0.1 to 0.9 while fixing values of other parameters, with 400 simulations on a sample of 800 observations.

Figure 3 contains three box-plots of ML estimates of ω, α_1 and β_1, respectively in the order in which they appear. It is clear that ML method does not work well in the case of $\beta_1 = 0.9$, i.e., $\alpha_1 + \beta_1 = 1$, which shows that violating the stationary condition discussed in Example 1 does indeed cause a serious problem in the ML estimation of our working model. From the box-plot of estimated β_1, we also see that the IQR, that is inter-quartile range calculated by subtracting the first quartile from the third quartile, is decreasing evidently with rising β_1

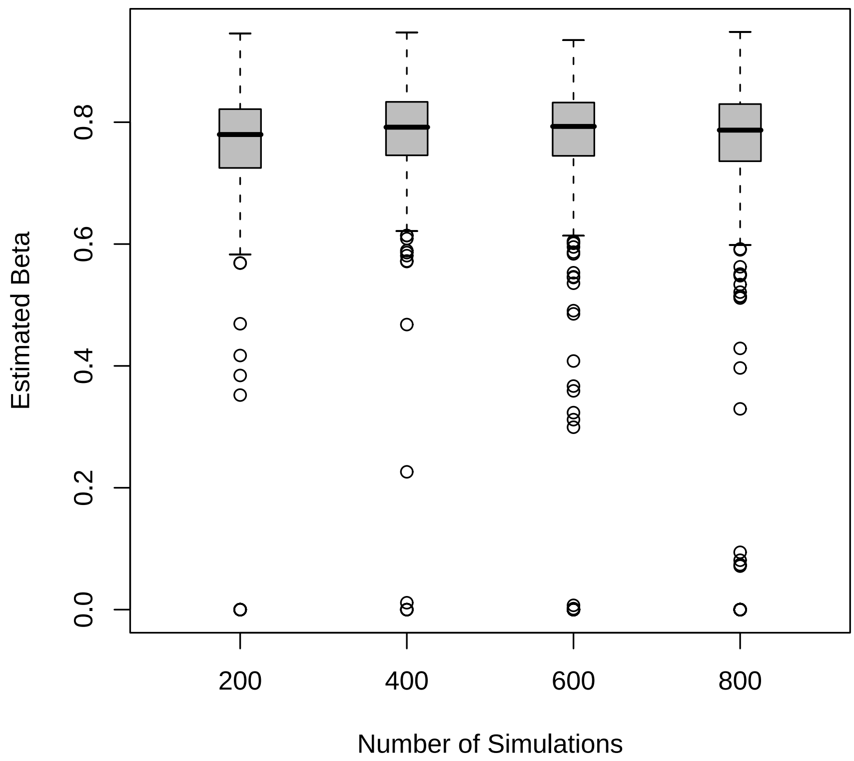


Figure 2: A box-plot of estimated β_1.

from 0.4 to 0.8, indicating that the middle 50% of the estimates is more closely clustered.

Yule-Walker estimation is applied within each same data set by solving a simple quadratic equation obtained in Example 3. Table 1 records the number n of no real roots in the interval $(0, 1)$ out of 400 simulations in each case, which indicates, as pointed out in Example 3, that sometimes the roots of this Yule-Walker equation make no sense to estimate parameters. The box-plots of Yule-Walker estimators in Figure 4 confirms that this method does not perform well for $ARMA\,(p, q)$ models with $q > 0$, however it can be used as preliminary estimation. Notice that, in the case of $\beta_1 = 0.9$, the performance of Yule-Walker method for estimating β_1 is not so sensitive to the violation of $\alpha_1 + \beta_1 < 1$ that holds for stationarity of our working model.

β_1	0.1	0.2	0.3	0.4	0.5	0.6	0.7	0.8	0.9
n	184	144	99	109	75	60	65	67	113

Table 1: List of n based on 400 simulations in each case.

Furthermore we consider ML estimation based on the ARMA representation of squared GARCH process also from the same data set. Figure 5 gives the box-plots of corresponding estimated parameters, and shows that ML method works relatively well for α_1 and β_1 even in the case of $\beta_1 = 0.9$, while it yields considerably more unreliable estimates of ω and the associated IQR is 12.0788 to 34.0176 when β_1 reaches 0.9.

Figure 6 presents a combination of box-plots of estimated β_1 for $\beta_1 = 0.1$, 0.3, 0.5, 0.7 and 0.9 from Figures 3-5, which allows a preliminary rough comparison of these three estimators. According to the box-plot presentation, ML estimators in both cases achieve similar performances except for the $\beta_1 = 0.9$ case, and show better estimation properties than Yule-Walker estimators. This result is also confirmed by associated comparative plots for the 50% confidence interval and arithmetic mean presented in Figures 7 and 8, respectively. From Figure 8, we also see that, ML method performs better on original $GARCH\,(1,1)$ than the corresponding $ARMA\,(1,1)$ under stationary conditions, in the sense of having much closer means to true values of β_1 for most cases.

Having observed these estimators in response to our stationary critical condition $\alpha_1 + \beta_1 = 1$, we may take this a little further by $\alpha_1 + \beta_1$ exceeding this critical value 1, which means we vary β_1 over 0.91, 0.92 and 0.93 under the same experimental procedure as described previously. Box-plots summarizing estimated β_1 values are presented in Figure 9, showing that the estimation fails completely when $\alpha_1 + \beta_1$ exceeds the critical value beyond the 1% level. The result clearly indicates that, neglecting the acceptable difference from theoretical value, a stationary process is needed to get a reliable estimation result.

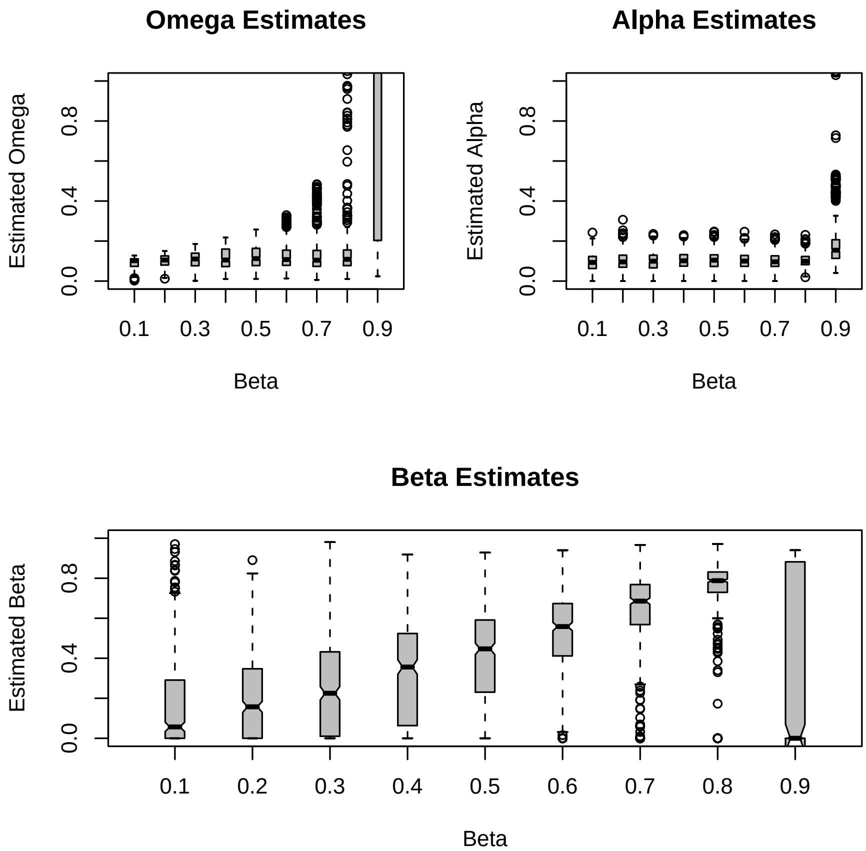


Figure 3: Box-plots of ML estimators based on $GARCH\,(1,1)$ models for $\omega = 0.1$, $\alpha_1 = 0.1$ and $\beta_1 = 0.1, 0.2, \ldots, 0.9$.

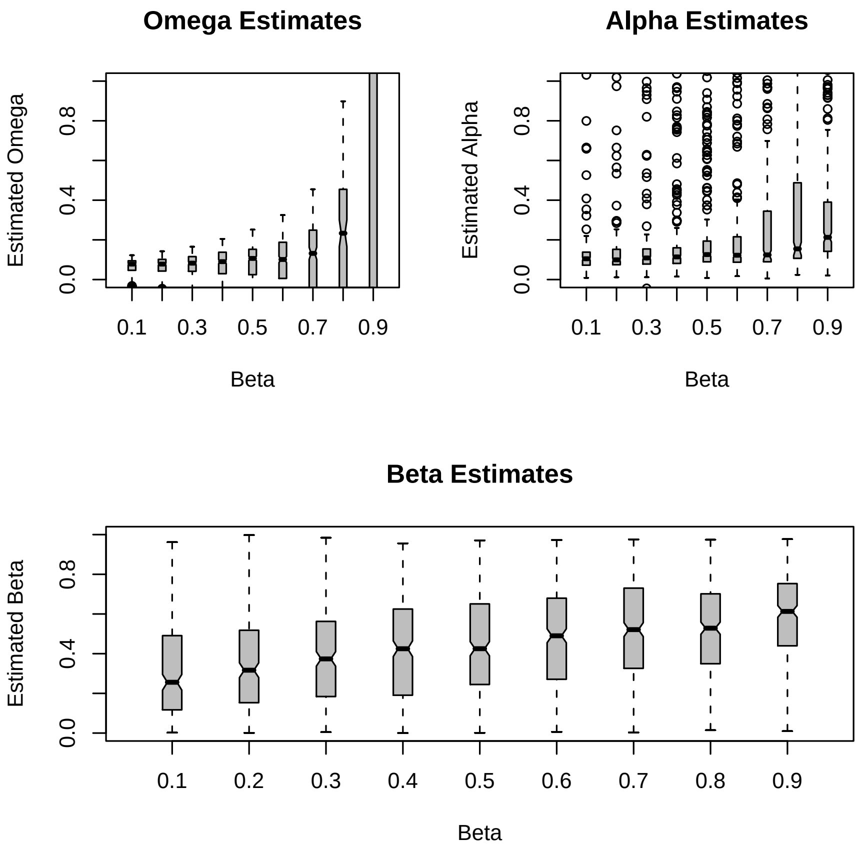


Figure 4: Box-plots of Yule-Walker estimators based on ARMA representation of $GARCH\,(1,1)$ models for $\omega = 0.1$, $\alpha_1 = 0.1$ and $\beta_1 = 0.1, 0.2, \ldots, 0.9$.

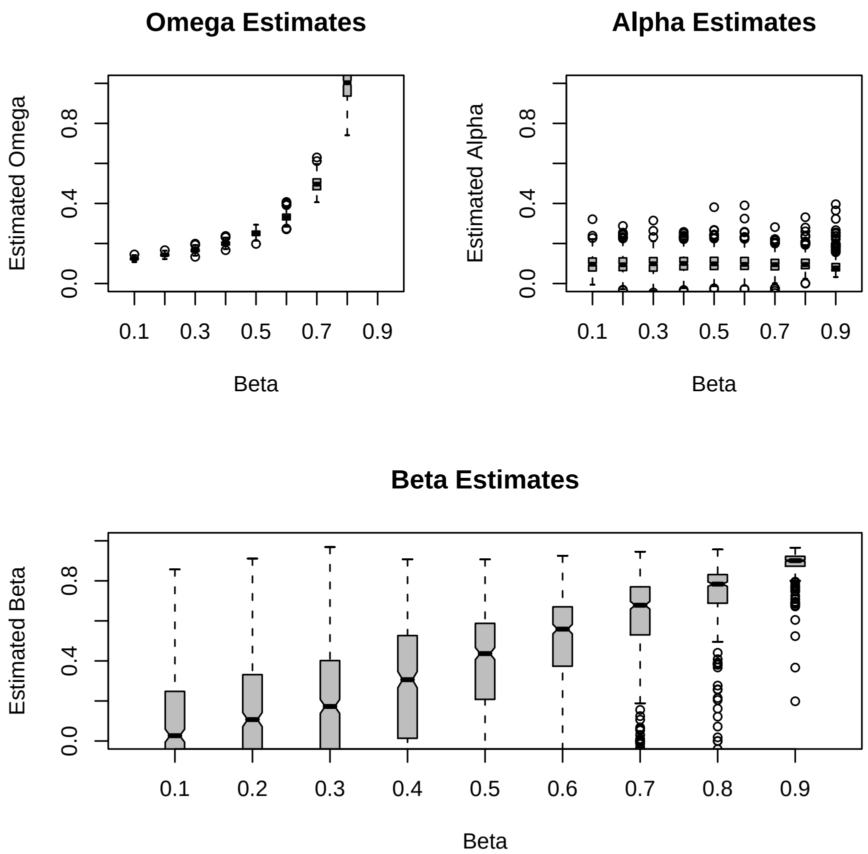


Figure 5: Box-plots of ML estimators based on ARMA representation of $GARCH\,(1,1)$ models for $\omega = 0.1$, $\alpha_1 = 0.1$ and $\beta_1 = 0.1, 0.2, \ldots, 0.9$.

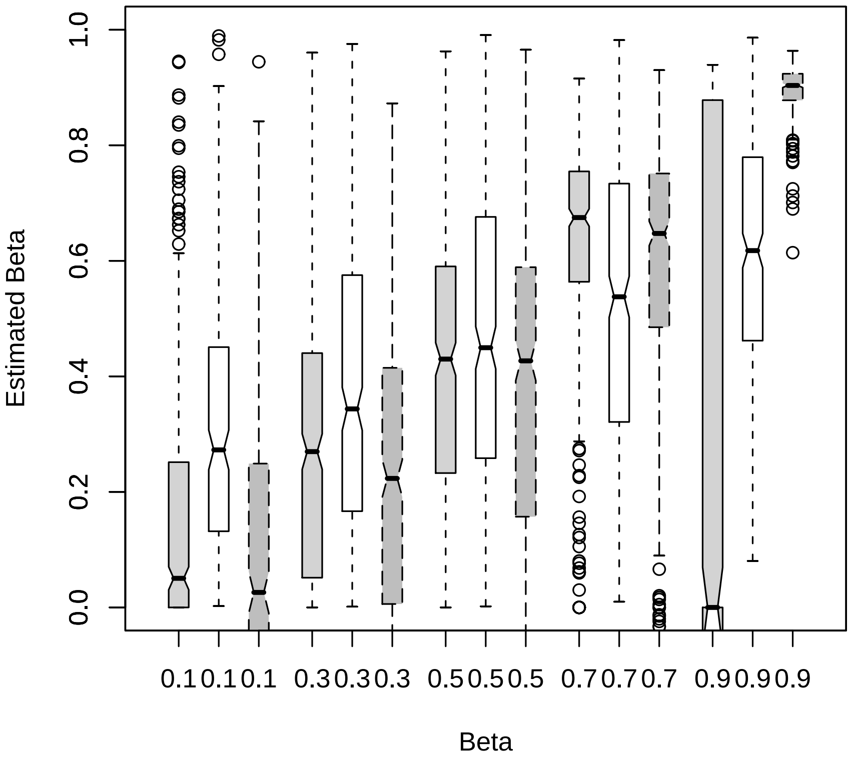


Figure 6: A combination of Figures 3-5 for estimated β_1 plotted in the order of ML estimator on GARCH, Yule-Walker and ML estimators based on squared GARCH for each case.

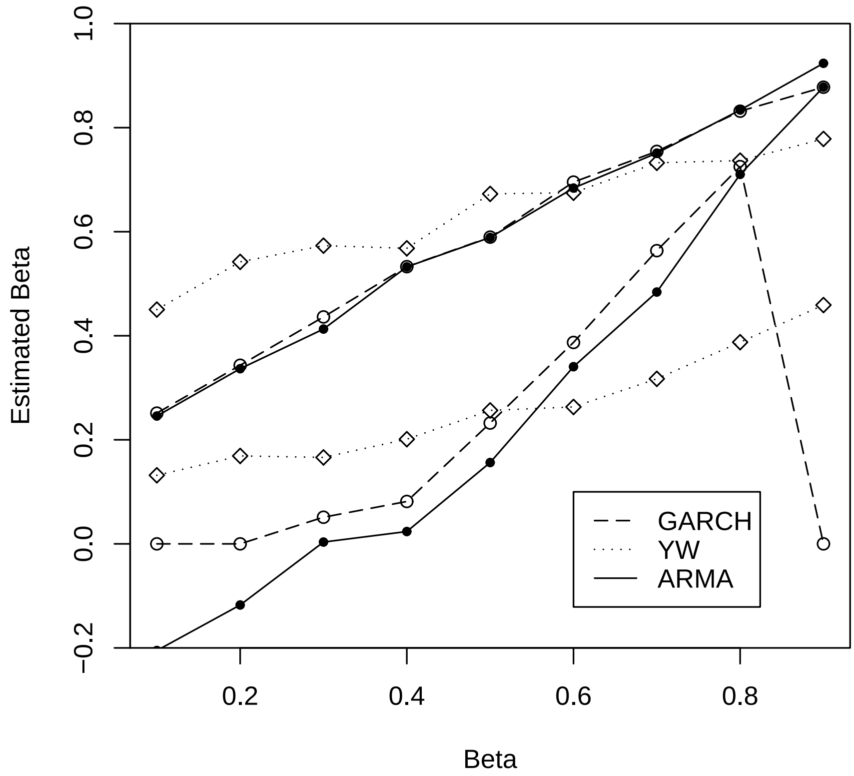


Figure 7: A plot of 50% confidence intervals for estimated β_1. Here GARCH denotes ML estimator on GARCH, YW and ARMA present Yule-Walker and ML estimators based on squared GARCH, respectively.

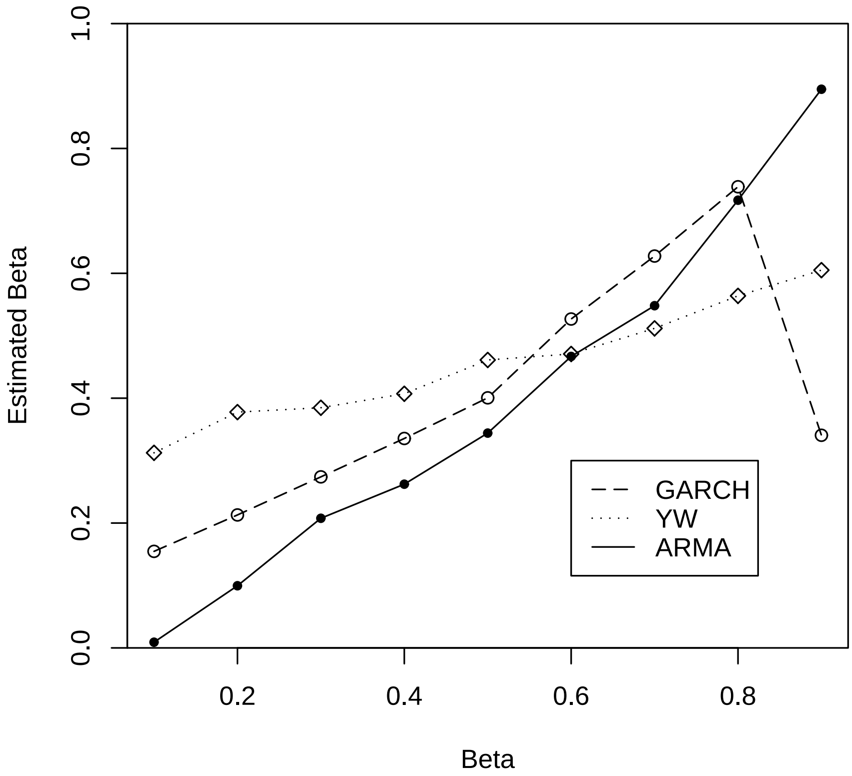


Figure 8: A plot of arithmetic means for estimated β_1. Here GARCH denotes ML estimator on GARCH, YW and ARMA present Yule-Walker and ML estimators based on squared GARCH, respectively.

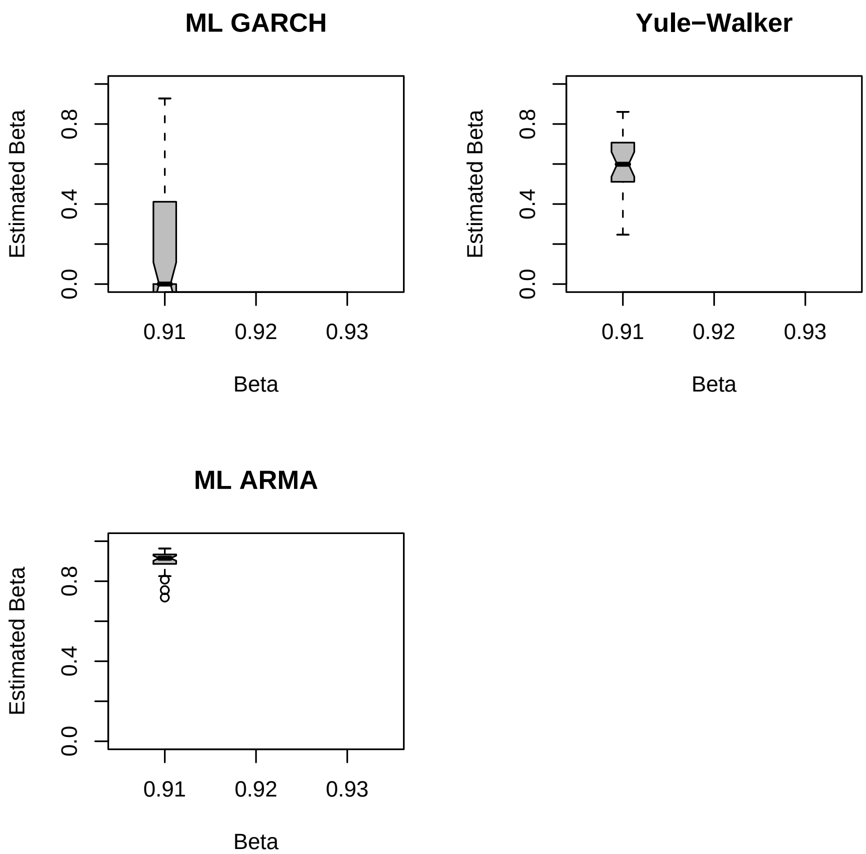


Figure 9: Box-plots of estimated β_1 based on $GARCH\,(1,1)$ models for $\omega = 0.1$, $\alpha_1 = 0.1$ and $\beta_1 = 0.91, 0.92$ and 0.93.

5 Application to Nordea stock price

An empirical study on stock price behavior of Nordea Bank has been performed by applying the class of GARCH models to stock return series. The daily stock price data are collected from The Nordic Exchange for the period 2000-10-16 to 2007-01-10.

The plot of closing prices in top row in Figure 10 suggests that there are three different series in our sample, the first one as Period 1 is up to the minimal price in 2003-03-11, thereafter Period 2 as the price climes until 2006-04-05, and then for the remaining data is Period 3. Therefore we split the data into three corresponding samples and analyze separately. Notice that, in Period 3, the sample size of 189 might cause our analysis lack the precision to provide reliable answers.

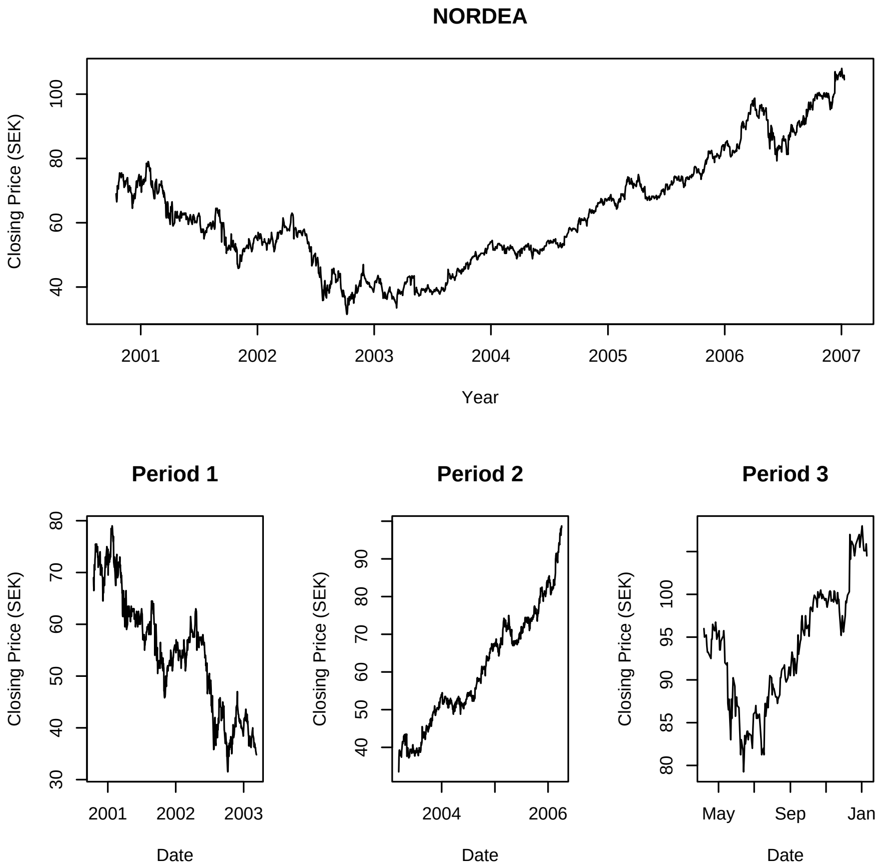


Figure 10: Plots of stock prices of Nordea from 2000-10-16 to 2007-01-10.

The daily stock returns R_t in our case is calculated as the difference in natural logrithm of the closing price P_t for two consecutive trading days,

$$R_t = \log(P_t) - \log(P_{t-1}).$$

Table 2 presents a range of statistics for returns of these three samples. It can be observed that all distributions are slightly skewed, as skewness is a measure of lack of symmetry, showing insignificant deviation from symmetry. The later two periods 2 and 3 exhibit a higher level of kurtosis than 3, which, as pointed out in [2], is one of the stylized facts of financial asset returns. Moreover we can see that the p-values of KPSS test are greater than 0.1, which fails to reject the stationary null at the 1% significance level in all three cases. The Ljung-Box test statistic for autocorrelation check has also been calculated, and it only strongly rejects the null hypothesis of no autocorrelation for Period 3 since the p-value is less than the significance level of 1%, which together with its first order autocorrelation coefficient -0.252, indicating that the daily returns of Period 3 is first order serially correlated.

In order to generate an uncorrelated series for Period 3, we build an $AR(1)$ model of daily returns, and obtain

$$R_t = 0.0005 - 0.2521 R_{t-1} + r_t.$$

Basic statistical properties of residuals r_t obtained from above regression model is presented in the last column in Table 2. It appears that the value of kurtosis is close to that of the original data and the distribution is still slightly skewed, while the first order autocorrelation and the Ljung-Box test indicate that the $AR(1)$ transformation of the returns provides an uncorrelated series of residuals.

Statistics	Period 1	Period 2	Period 3	Residuals r_t
Sample size	600	772	188	188
Mean	-0.001141	0.001400	0.000451	-0.000013
Variance	0.000768	0.000202	0.000286	0.000268
Skewness	-0.000793	-0.506408	0.080548	-0.084196
Kurtosis	1.609951	9.630865	3.170765	3.325100
KPSS test p-value	> 0.1	> 0.1	> 0.1	> 0.1
Ljung-Box test statistic	3.57	4.3208	12.1264	0.009
p-value	0.05883	0.03765	0.0004971	0.9244
First order autocorrelation	-0.077	-0.075	-0.252	-0.007

Table 2: A range of statistics for return series in Periods 1-3 and residuals from AR transformation to Period 3.

We are now in a position to apply the class of GARCH models to returns R_t for Periods 1 and 2, and residuals r_t of the $AR(1)$ model for Period 3. Figure 11 gives the associated ACF and PACF plots, which not surprisingly reveals no hint about the order in which these models should be created. It is just confirming that indeed these are uncorrelated series.

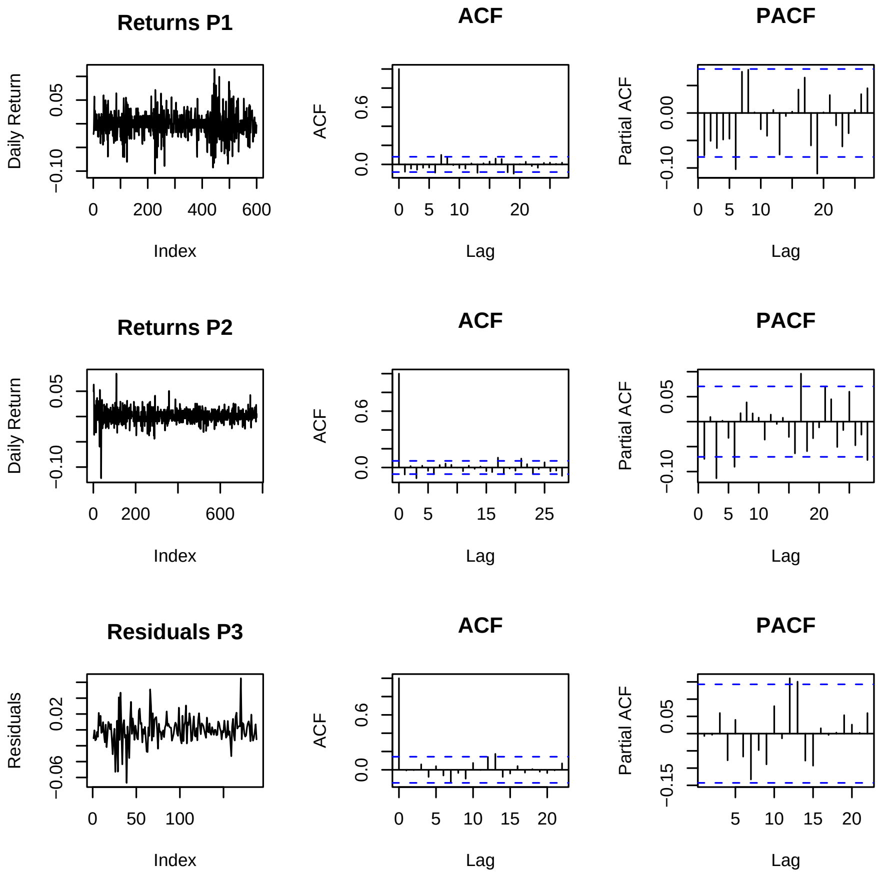


Figure 11: Plots of data to be fitted and corresponding ACF, PACF.

As we know from Theorem 2 if ε_t follows a GARCH process, then ε_t^2 follows an ARMA model. Thus the ACF and PACF of squared returns R_t^2 and squared residuals r_t^2 are plotted in Figure 12 in order to determine the order (p, q) of the model, however, they are really not adding that much information for order selection here. Motivated by comments in [2] that in practical applications GARCH with smaller orders often sufficiently describe the data and in most cases $GARCH(1, 1)$ is sufficient, we hence consider four different combinations of $p = 0, 1$ and $q = 1, 2$ for each period.

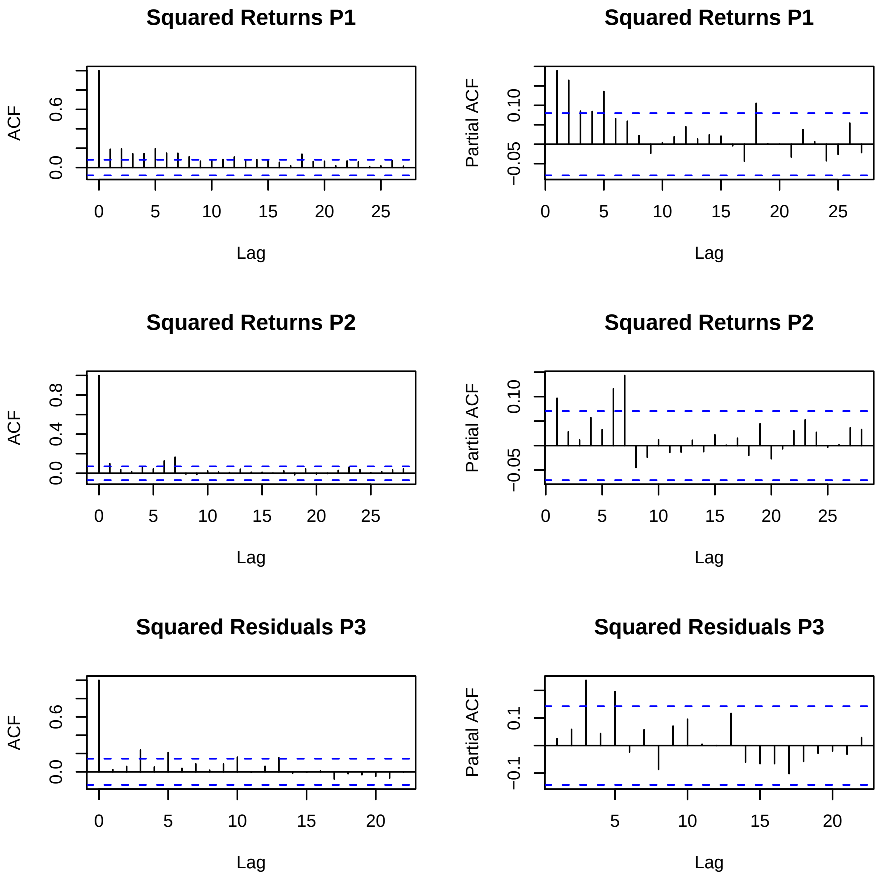


Figure 12: ACF, PACF Plots of squared return series in Periods 1-2 and squared residuals in Period 3.

In order to test for the relative fit of the various models, we calculate Bayesian information criterion (abbreviated BIC),

$$BIC = -2\ln(L) + k\ln(n),$$

where L denotes the maximized value of the likelihood function for the estimated model, k represents the number of free parameters to be estimated and n is the number of observations. According to this criterion, the model with the lowest BIC value fits the data best. Moreover, as a goodness-of-fit measure of departure from normality, the Jarque-Bera test statistic is also calculated by

$$JB = \frac{n}{6}\left(S^2 + \frac{(K-3)^2}{4}\right),$$

based on sample skewness S and sample kurtosis K. Since samples from a normal distribution have an expected skewness of 0 and an expected excess kurtosis of 0, any deviation from this increases the Jarque-Bera statistic.

Model	Log-Likelihood	BIC Value	Jarque-Bera stats.	p-value
		Period 1		
$ARCH(1)$	1859.749	-2605.816	58.0172	2.521e-13
$ARCH(2)$	1869.857	-2621.474	43.0094	4.577e-10
$GARCH(1,1)$	1889.757	-2659.435	31.202	1.677e-07
$GARCH(1,2)$	1887.312	-2649.985	24.9657	3.791e-06
		Period 2		
$ARCH(1)$	2910.165	-4390.028	4258.518	$<$ 2.2e-16
$ARCH(2)$	2911.913	-4388.714	4374.246	$<$ 2.2e-16
$GARCH(1,1)$	2948.735	-4460.52	377.6917	$<$ 2.2e-16
$GARCH(1,2)$	2949.495	-4457.228	370.4134	$<$ 2.2e-16
		Period 3		
$ARCH(1)$	675.9263	-997.6968	96.075	$<$ 2.2e-16
$ARCH(2)$	673.0288	-988.5032	110.6344	$<$ 2.2e-16
$GARCH(1,1)$	685.7861	-1012.180	110.4373	$<$ 2.2e-16
$GARCH(1,2)$	680.5356	-998.2803	103.4722	$<$ 2.2e-16

Table 3: Maximum log-likelihoods and results of goodness-of-fit tests for GARCH models for each period.

The maximized log-likelihood for each model as well as results of above goodness-of-fit tests are summarized in Table 3 by sample. Clearly in all three cases, BIC values are the lowest for the $GARCH(1,1)$, indicating that this model generally outperforms other GARCH models. Notice that differences in BIC values among all models appear rather small relative to the high-level BICs, which might imply that no significant difference in performance in data fitting. This will be confirmed to some extend by the results from our bootstrap analysis in the following. We construct bootstrap confidence intervals for the original data series by generating new series from the fitted model. The basic idea is as follows:

Let $\hat{\omega}$ and $\hat{\alpha}_1$ be the parameter estimates from the regression model $ARCH(1)$ fitted for the data series $e_1, \ldots, e_N$. For other three models the construction procedure is almost the same except different conditional variance forms.

(1) Generate a series of N values from the standard normal distribution, that is, $Z_t \sim N(0,1)$.

(2) Generate an equal number of values $\{\hat{e}_t\}_{t=1}^N$ by $\hat{e}_t = Z_t\sqrt{h_t^2}$ for $h_t^2 = \hat{\omega} + \hat{\alpha}_1\hat{e}_{t-1}^2$.

(3) Repeat steps (1)-(2) 200 times to obtain data series $\hat{e}_{i1},\ldots,\hat{e}_{iN}$ for $i = 1,\ldots,200$.

(4) Generate N bootstrap samples of $\{\hat{e}_{1,j},\ldots,\hat{e}_{200,j}\}$, $j = 1,\ldots,N$. For each sample, data are ordered from smallest to largest as $\{\tilde{e}_{1,j},\ldots,\tilde{e}_{200,j}\}$.

(5) The 5th and 195th values, that is $\{\tilde{e}_{5,j}\}$ and $\{\tilde{e}_{195,j}\}$, represent the lower and upper limits of the 95% confidence interval.

Figures 13-15 plot the 95% confidence intervals for each period by model, respectively. It can be observed that both GARCH models fit the data better to some extent, while $GARCH\,(1,1)$ does not exhibit clearly detectable superior performance to $GARCH\,(1,2)$ in first two cases.

To look further into the extent to which two GARCH models can fit the dynamics of our data in the first two periods, we simulate stock prices $\hat{P}_t$ from the iterations $\hat{P}_t = e^{\hat{R}_t}\hat{P}_{t-1}$, where $\hat{R}_t$ is return series generated from the fitted models by taking bootstrap steps (1)-(2). The comparative plots of two simulated price series in Figures 16 and 17 almost coincide, conforming again that the difference in model performance is very small.

Finally we present the parameter estimates of $GARCH\,(1,1)$ and $GARCH\,(1,2)$ in Table 4. As for the stationarity discussed, $\sum \alpha_i + \beta_1$ in both models are all less than one, though rather close to, the stationary condition holds for the fitted models.

Model	$\hat{\omega}$	$\hat{\alpha}_1$	$\hat{\alpha}_2$	$\hat{\beta}_1$	$\sum \alpha_i + \beta_1$
			Period 1		
$GARCH\,(1,1)$	4.714e-05	1.418e-01	-	7.984e-01	0.9402
$GARCH\,(1,2)$	6.856e-05	1.026e-01	9.872e-02	7.151e-01	0.91642
			Period 2		
$GARCH\,(1,1)$	7.542e-06	9.585e-02	-	8.680e-01	0.96385
$GARCH\,(1,2)$	7.105e-06	6.669e-02	2.031e-02	8.771e-01	0.9641
			Period 3		
$GARCH\,(1,1)$	1.581e-05	1.147e-01	-	8.326e-01	0.9473
$GARCH\,(1,2)$	3.137e-05	2.129e-08	2.598e-01	6.808e-01	0.9406

Table 4: $GARCH\,(1,1)$ and $GARCH\,(1,2)$ parameter estimates.

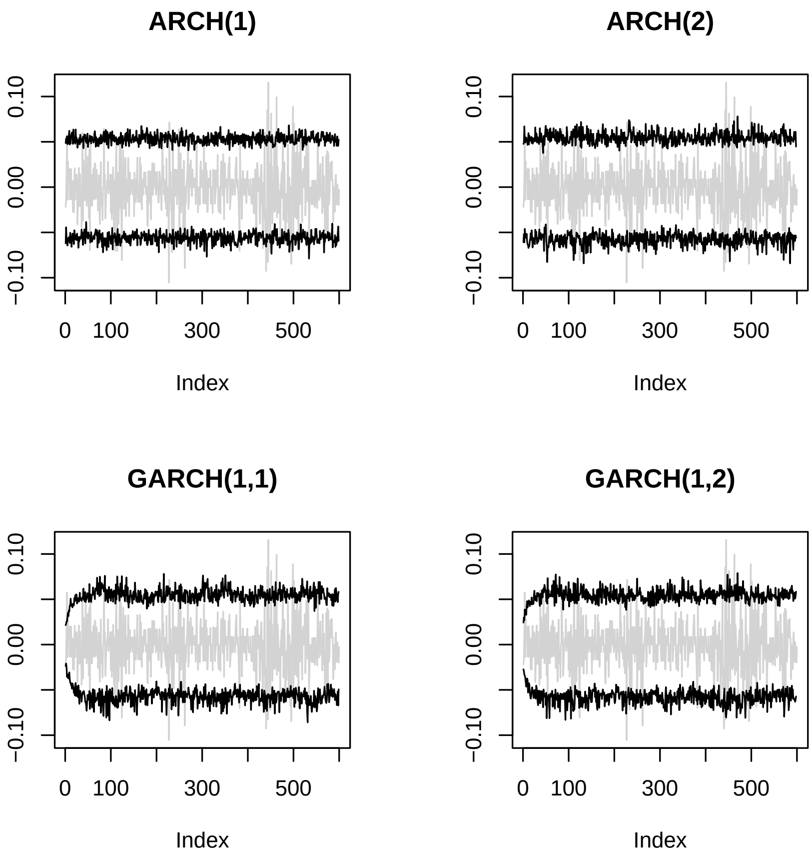


Figure 13: A plot of 95% bootstrap confidence intervals for returns in Period 1.

5.1 Remarks on Nordea application

In Figures 16 and 17, the simulated path moves on distinctive patterns chosen from a number of different simulation results sampled from standard normal distribution in the first step of bootstrap, based on normal distribution assumption. However the Jarque-Bera test statistic listed in Table 3 provides clear evidence to reject the null hypothesis of normality for the distribution of the residuals, as a rule of thumb, which implies that the data to be fitted is not normally distributed. This might cause the unstable performance of our fitted models, which varies over such a large range as shown, regarding the match between data and simulation.

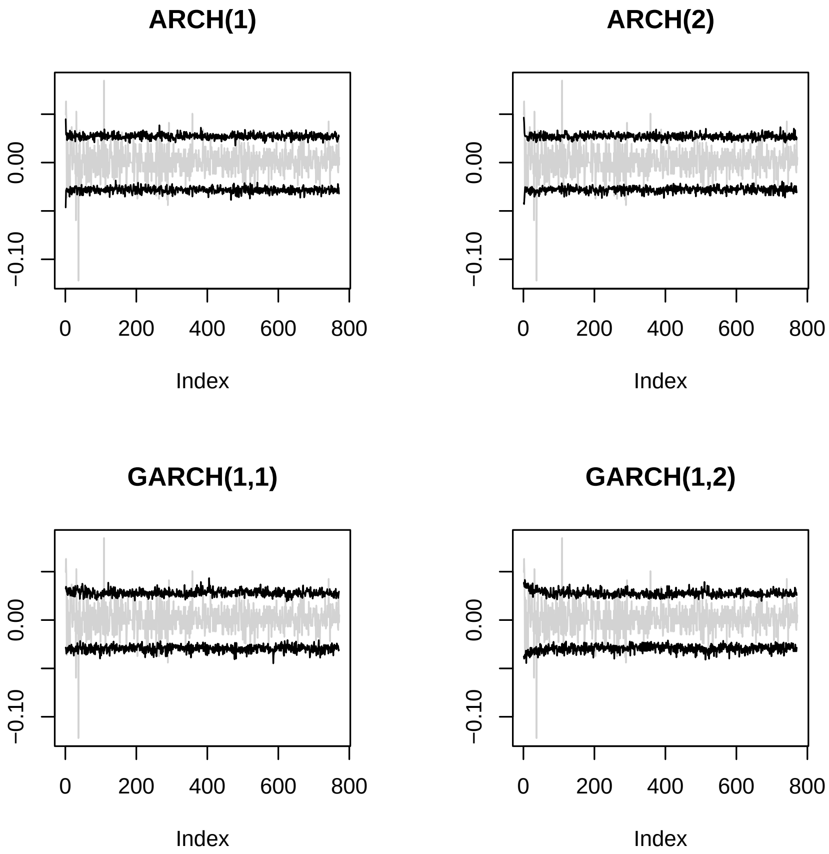


Figure 14: A plot of 95% bootstrap confidence intervals for returns in Period 2.

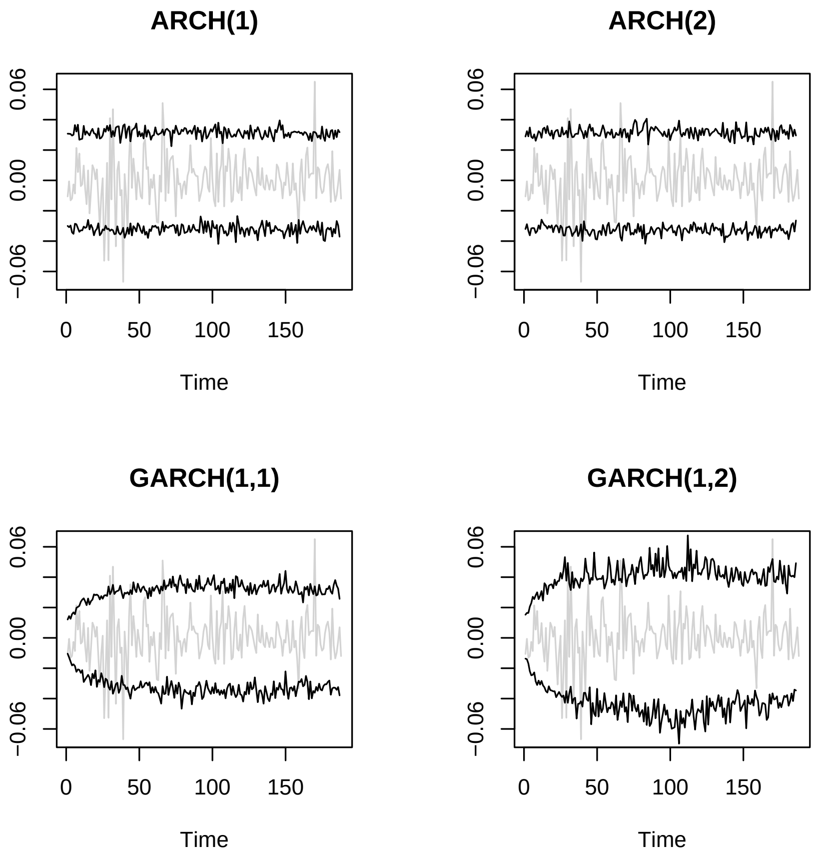


Figure 15: A plot of 95% bootstrap confidence intervals for residuals from $AR\,(1)$ transformation to Period 3.

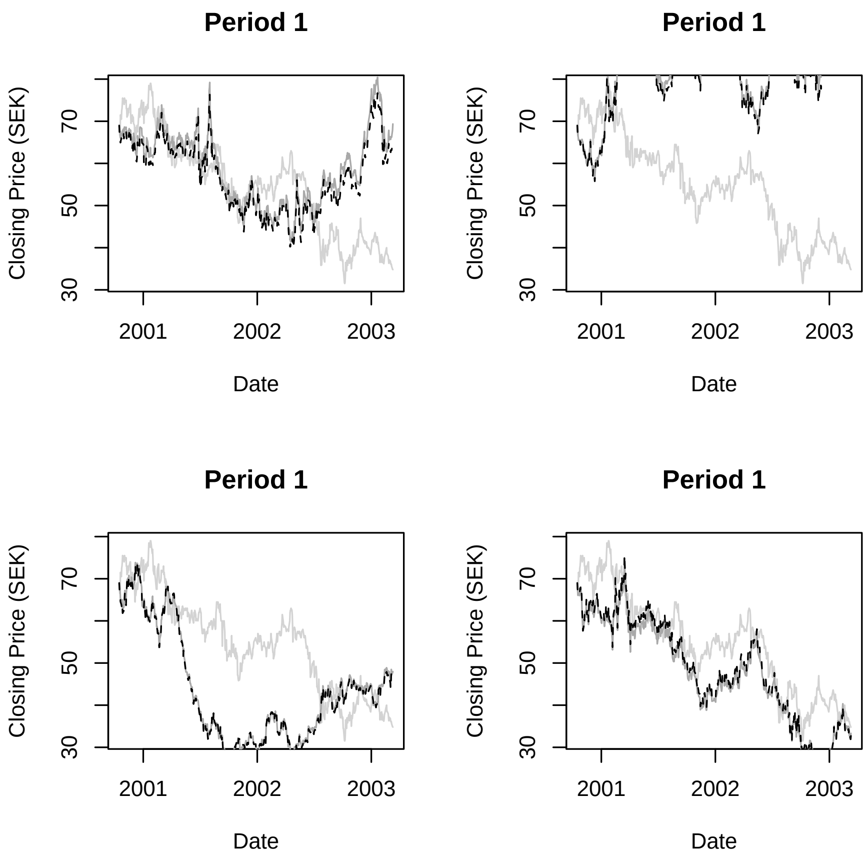


Figure 16: Plots of simulated stock prices for Period 1 from fitted models $GARCH\,(1,1)$ and $GARCH\,(1,2)$.

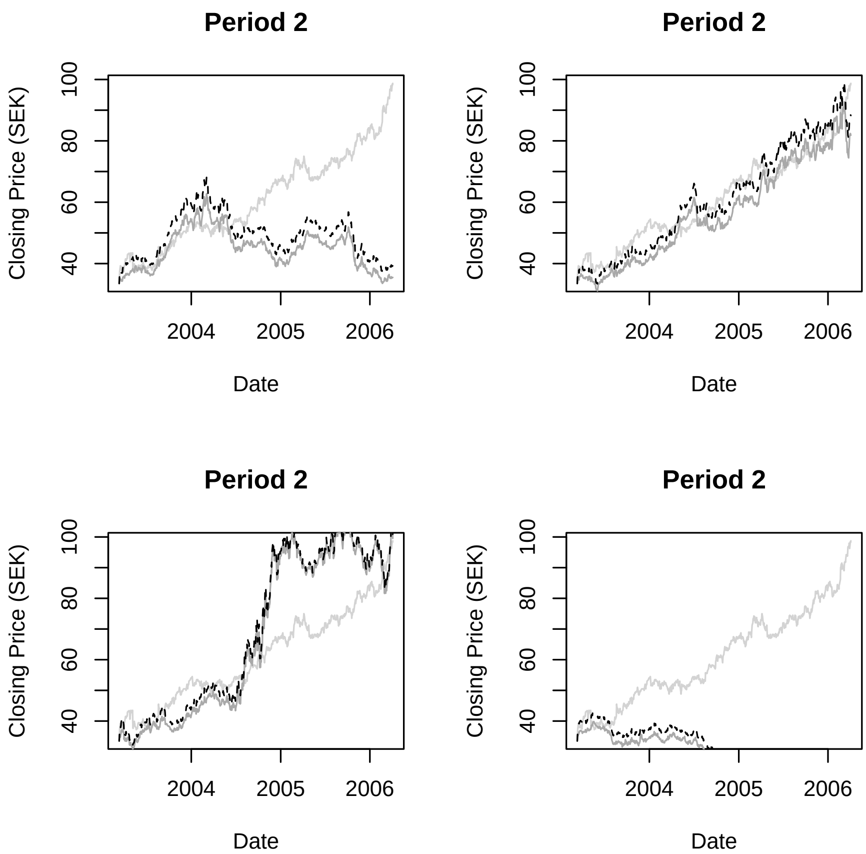


Figure 17: Plots of simulated stock prices for Period 2 from fitted models $GARCH\,(1,1)$ and $GARCH\,(1,2)$.

References

[1] Bollerslev, T. (1986) Generalized Autoregressive Conditional Heteroskedasticity, J. Econometrics, 31, 307-327.

[2] Franke, J., Hardle, W. and Hafner, C. (2004) Statistics of Financial Markets, Springer-Verlag, Germany.

[3] Brockwell, P.J. and Davis, R.A. (2002) Introduction to Time Series and Forecasting, 2nd Edition, Springer-Verlag, New York.

[4] Berkes, I., Horvath, L., and Kokoszka, P. (2003) GARCH processes: structure and estimation, J. Bernoulli, 9, 201-227.

[5] George Levy (2001) An Introduction to GARCH Models in Finance, Financial Engineering News, 22.