Volatility proxies and GARCH models
Visser, M.P.

Citation for published version (APA):

General rights
It is not permitted to download or to forward/distribute the text or part of it without the consent of the author(s) and/or copyright holder(s), other than for strictly personal, individual use, unless the work is under an open content license (like Creative Commons).

Disclaimer/Complaints regulations
If you believe that digital publication of certain material infringes any of your rights or (privacy) interests, please let the Library know, stating your reasons. In case of a legitimate complaint, the Library will make the material inaccessible and/or remove it from the website. Please Ask the Library: http://uba.uva.nl/en/contact, or a letter to: Library of the University of Amsterdam, Secretariat, Singel 425, 1012 WP Amsterdam, The Netherlands. You will be contacted as soon as possible.
Chapter 3

GARCH Parameter Estimation Using High-Frequency Data

GARCH models have a rich tradition in financial economics. The models are standard tools for describing and forecasting time varying volatility. They are also used in risk management, asset allocation, and option pricing.

Much of the empirical work on GARCH models is concerned with modelling daily volatility using close-to-close returns. The basic GARCH(1,1) model (Bollerslev, 1986), applied to daily returns $r_n$, reads

$$r_n = \sigma_n Z_n,$$

$$\sigma_n^2 = \kappa + \alpha r_{n-1}^2 + \beta \sigma_{n-1}^2,$$  \hspace{1cm} (3.1)

(3.2)

where the $Z_n$ are iid, mean zero, unit variance innovations, and the scale factors $\sigma_n$ are positive. Estimation of the parameters ($\kappa, \alpha, \beta$) has become a routine exercise. Common practice is to apply a quasi maximum likelihood estimator (QMLE): maximize the log-likelihood for the returns $r_n$ as if the innovations $Z_n$ have a standard Gaussian distribution. If the $Z_n$ are not Gaussian, the QMLE is still consistent and asymptotically normal, with adjusted standard errors.

Many financial data sets include intraday high-frequency data in addition to the daily close-to-close returns. Such data sets contain more information, so in principle it should be possible to improve classical GARCH procedures based on close-to-close returns. Do high-frequency data give rise to improved estimators of the parameters ($\kappa, \alpha, \beta$) above? How does one include the intraday price movements in the GARCH recursion (3.2) in order to obtain improved forecasts? How large is the improvement? This chapter focuses on the question of parameter estimation. Given the GARCH model (3.1–3.2) for daily returns, the chapter develops theory for using high-frequency data to obtain improved estimators of the parameters ($\kappa, \alpha, \beta$). Despite the now wide availability of high-frequency data, the
parameter estimation for GARCH models has not yet benefitted. To use high-frequency data one needs a model for the price process over the course of the day, but it is unclear how the intraday price movements should be incorporated into the discrete time GARCH model.

This chapter develops a QMLE that uses volatility proxies to estimate the parameters of a discrete time, daily GARCH model. The proxies are functions of intraday data. The resulting estimation theory is fairly easy to apply. It is based on a straightforward model for the intraday price movements. The intraday model is nonparametric; the properties of the estimators are derived without having to know the distribution of the intraday price process. For reasons of exposition the chapter deals with the case of GARCH(1,1) daily returns as given by equations (3.1–3.2). The QML-arguments in the chapter may also be applied to GARCH($p,q$) models, and to other GARCH models.

The principle underlying the classical GARCH QMLE is that the squared daily return is a conditionally unbiased estimator of the squared scale factor:

$$E(r_n^2|\mathcal{F}_{n-1}) = \sigma_n^2,$$

where $\mathcal{F}_{n-1}$ denotes yesterday’s information. It is well known that $r_n^2$ is a noisy estimator of $\sigma_n^2$. The realized variance, obtained by summing squared five-minute returns, is a considerably more accurate (though still noisy) estimator, see Andersen and Bollerslev (1998a). The QMLE of this chapter exploits this accuracy by working with the likelihood for the realized volatility (the square root of the realized variance) instead of a likelihood for the returns $r_n$. We shall also derive the likelihood for other proxies, for example the intraday high-low range, or the absolute value of the maximal decrease of the intraday return process over a fifteen minute interval.

The accuracy of the estimators of the GARCH parameters may be increased by using an appropriate volatility proxy. The chapter relates the asymptotic relative efficiency of the QMLEs to the noisiness of the proxies. The QMLE provided in this chapter reduces to the classical QMLE by inserting the absolute daily return $|r_n|$ as the volatility proxy.

An empirical application to the S&P 500 index data shows a large improvement in the estimator efficiency. The estimated variances for the GARCH autoregression parameters decrease by a factor twenty. A simulation study confirms that substantial improvements of estimator efficiency may be achieved.

Lildholdt (2002) and Brandt and Jones (2006) use intraday data to estimate parameters for discrete time, daily GARCH models. These papers assume an intraday Brownian motion and use the probability distribution of the daily high-low range to estimate the GARCH parameters by maximum likelihood. This chapter does not impose constraints on the distribution of the intraday price process, and allows a large class of volatility proxies including the high-low range.
There is other research that uses high-frequency data for parameter estimation, but not on models that have GARCH(1,1) daily returns. Drost and Nijman (1993) showed that the GARCH model is not closed under temporal aggregation. One can estimate the parameters of a GARCH process with a five-minute time unit and derive the parameters for the implied daily process, but the daily process is no longer GARCH. See also Drost and Werker (1996), Meddahi and Renault (2004), and Meddahi, Renault, and Werker (2006). The approach in this chapter is conceptually different. It preserves the discrete time daily GARCH model and also incorporates the intraday price movements. For overviews of GARCH literature see Bollerslev, Engle, and Nelson (1994), and Andersen, Davis, Kreiss, and Mikosch (2009).

The remainder of the chapter is organized as follows. Section 3.1 introduces the model and discusses volatility proxies. Section 3.2 presents the estimation theory. This part starts by generalizing the classical QMLE based on the returns $r_n$ to a QMLE based on volatility proxies. It also provides an alternative, using the log proxies for estimation, and yielding log-Gaussian QML. Section 3.3 applies the QMLEs to the S&P 500 index data. Section 3.4 uses simulations to compare the estimators based on realized variance with the standard QMLE based on close-to-close returns. Section 3.5 discusses the relation to semimartingale models. Our conclusions are presented in Section 3.6. Appendices 3.A, 3.B, 3.C, and 3.D give a description of the data, background on QML estimation, proofs, and details on simulations.

### 3.1 Model

We use the GARCH(1,1) representation of Drost and Klaassen (1997) to model the daily returns $r_n$:

\[
\begin{align*}
  r_n &= v_n \tau Z_n, \\
  v_n^2 &= 1 + \gamma r_{n-1}^2 + \beta v_{n-1}^2,
\end{align*}
\]

where the innovations $Z_n$ are iid. The system (3.3–3.4) is equivalent to the GARCH equations (3.1–3.2) by writing

\[
\begin{align*}
  \sigma_n &= v_n \tau, \\
  \kappa &= \tau^2, \\
  \alpha &= \gamma \tau^2.
\end{align*}
\]

For identification the second moment is standardized\(^1\) by $\mathbb{E}Z_n^2 = 1$. The system (3.3–3.4) has the property that the standardization of $Z_n$ affects only the parameter $\tau$. This shall be of use when we discuss estimation of $\gamma$ and $\beta$. We shall refer to the parameter $\tau$ as

\(^1\)It is often also assumed that $\mathbb{E}Z_n = 0$, but this is not needed for parameter estimation; $\mathbb{E}r_n \neq 0$ is allowed in general.
the norming parameter and to \( \gamma, \beta \) as the autoregression parameters.

To deal with high-frequency data in the daily GARCH system (3.3–3.4) one needs a model for the intraday price movements. For each trading day \( n \) we introduce a process \( R_n(\cdot) \). This is the continuous time log-return process for that day. For ease of notation we normalize the trading day to the unit time interval. We shall now provide a nonparametric extension of the daily GARCH process to the intraday time interval. Adjust equation (3.3) by replacing the variable \( Z_n \) by a process \( \Psi_n(\cdot) \). This yields the intraday extension

\[
R_n(u) = v_n \tau \Psi_n(u) , \quad 0 \leq u \leq 1 ,
\]

\[
v_n^2 = 1 + \gamma r_{n-1}^2 + \beta v_{n-1}^2 ,
\]

where intraday time \( u \) advances from zero to one. For different days the standard processes \( \Psi_k \) and \( \Psi_n \), \( k \neq n \), are assumed independent and to have the same probability distribution. Precisely, the sequence of standard processes \( \Psi_n(\cdot) \) is assumed iid, and the sample paths of the \( \Psi_n(\cdot) \) are right-continuous and have left limits (cadlag). We impose the standardization \( \mathbb{E} \Psi_n^2(1) = 1 \). Write

\[
r_n \equiv R_n(1) , \quad \text{and} \quad Z_n \equiv \Psi_n(1) ,
\]

to link up with the equations (3.3–3.4). The assumption that the sequence of processes \( (\Psi_n) \) is iid ensures that the sequence of random variables \( (Z_n) \) is iid.

Except for the regularity condition of cadlag sample paths, no conditions are imposed on the standard process \( \Psi \). Equation (3.5) can therefore be a realistic model for the intraday log-return process. The process \( \Psi \) allows for empirically relevant effects such as jumps, intraday periodicity in volatility, intraday volatility persistence, and stochastic spot volatility. The standard process \( \Psi_n \) may have a non-zero mean process; in general \( \mathbb{E} \Psi(1) \neq 0 \).

As a special case of this model take \( \Psi_n(\cdot) = W_n(\cdot) \) the standard Brownian motion. The standard Brownian motion \( W_n(\cdot) \) captures intraday price movements, whereas the scale factor \( v_n \tau \) amplifies the intraday fluctuations in \( W_n(\cdot) \) and is constant over the day. The scaled Brownian motion may be applied for GARCH maximum likelihood estimation using the intraday high-low range, see Lildholdt (2002) and Brandt and Jones (2006). The high-low range is an example of the proxies that we consider in this chapter. In fact, Brownian motion is a deceptive special case. Here \( v_n \tau \) equals \( \sqrt{QV_n} \) (the square root of the day’s quadratic variation), and the scale factor can be calculated after having observed the first ten minutes of the intraday price process. For Brownian motion intraday volatility is constant. This is not the way we like to think of the intraday model (3.5). In general the scale factor \( v_n \tau \) is latent, and the intraday volatility process evolves randomly.

Equation (3.5) is called a scaling model for the intraday return process over the day.
If $\Psi(\cdot)$ is a semimartingale, then so is the log-return process. The estimation theory presented in this chapter does not need the semimartingale assumption. For a discussion on the relation to semimartingale models, see Section 3.5, where we also clarify the relation between quadratic variation and the GARCH scale factor.

Now that the intraday price movements have been incorporated into the daily GARCH model, we may use the intraday information, and construct proxies for the scale factors. Realized volatility is a commonly applied proxy, see for instance Barndorff-Nielsen and Shephard (2002b) and Andersen, Bollerslev, Diebold, and Labys (2003). The daily realized volatility $H_n$ is the square root of the realized variance, which is the sum of the squared returns over intraday intervals,

$$H_n = \left( \sum_k r_{n,k}^2 \right)^{1/2}, \quad (3.7)$$

where $r_{n,k}$ denotes the return over the $k$-th intraday interval of day $n$; in practice five-minute intervals are frequently applied.

The proxy $H_n$ in (3.7) is positive and has the property of positive homogeneity: if the process $R_n(\cdot)$ is scaled by a factor $\alpha \geq 0$, then so is the proxy:

$$H(\alpha R_n) = \alpha H(R_n), \quad \alpha \geq 0, \quad (3.8)$$

with $H_n \equiv H(R_n)$. We shall allow for a large class of proxies by allowing any positive and positively homogeneous statistic. Examples are the intraday high-low range, the absolute power variation, and the absolute return $|r_n|$. In this chapter these proxies serve as measurements of the scale factor $\sigma_n = v_n \tau$. In a recent paper (de Vilder and Visser, 2008) we discussed how to combine proxies into a more efficient one. Visser (2008a) used proxies to forecast volatility. This chapter explores how proxies improve parameter estimation.

We rule out that the degenerate proxy $H_n \equiv 0$ by assuming

$$0 < \mu_2^H \equiv \sqrt{\mathbb{E}H^2(\Psi)} < \infty,$$

and introduce the normalized innovation $Z_H$ (normalization $\mathbb{E}Z_H^2 = 1$) by setting

$$Z_H = H(\Psi) / \mu_2^H.$$

Due to positive homogeneity $H_n = v_n \tau H(\Psi_n)$, yielding the following stochastic system.
for the proxy $H_n$:

$$
H_n = v_n \tau_H Z_{H,n}, \quad (3.9)
$$

$$
v_n^2 = 1 + \gamma r_n^2 + \beta v_{n-1}^2, \quad (3.10)
$$

where $\tau_H = \mu_H^2$. The innovations $Z_{H,n} \geq 0$ are iid and have standardization $E Z_{H,n}^2 = 1$

$$
E(H_n^2 | F_{n-1}) = v_n^2 \tau_H^2. \quad (3.11)
$$

The system given by (3.9–3.10) is similar to equations (3.3–3.4): the parameters $\gamma, \beta$ in (3.10) have the same value as in equation (3.4), so $H_n$ and $r_n$ share the daily factor $v_n$, hence their second moments have identical GARCH dynamics. The intuition is that when the scale factor is large we do not only tend to see large returns $r_n$ (in absolute value), but any volatility proxy $H_n$ tends to be large too. In Section 3.2 we derive the likelihood for the observations $H_n$ yielding a QMLE for the parameter $\theta = (\tau_H, \gamma, \beta)$.

### 3.2 QML Estimation Using a General Volatility Proxy

In this section the classical QMLE based on the returns $r_n$ is generalized to a QMLE based on proxies $H_n$. We shall also apply quasi maximum likelihood to the logarithm of the proxies, since this often gives more efficient estimators. In order to distinguish the two cases we speak of a Gaussian QMLE and a log-Gaussian QMLE. For a brief review of the classical GARCH(1,1) QMLE based on close-to-close returns, see Appendix 3.B.3.

#### 3.2.1 Gaussian QMLE

From Section 3.1 we know that the volatility proxy $H_n$ satisfies

$$
E(H_n^2 | F_{n-1}) = v_n^2 \tau_H^2. \quad (3.11)
$$

Since the distribution of the standard process $\Psi_n$ is unknown, the distribution of $H_n$ is unknown. If one has random variables $H_n$ that satisfy equation (3.11), one may derive a QML estimator for $(\tau_H, \gamma, \beta)$, as we shall show in this section. The scaling model given by equation (3.5) and the positive homogeneity of the class $H$ given by equation(3.8) together ensure that our proxies satisfy (3.9–3.10) and also (3.11). Now that the system (3.9–3.10) and equation (3.11), have been established for the proxies $H_n$, one may forget about the intraday return process $R_n(\cdot) = v_n \tau \Psi_n(\cdot)$ for the remainder of the exposition.

The proxies $H_n$ are positive. Create real-valued, ancillary random variables $y_n$ by randomly attributing signs to the $H_n$. Precisely, the random variables $\text{sgn}_n$, independent
of the GARCH system (3.5–3.6), form an iid sequence of $\{-1, 1\}$-valued random variables with both outcomes having probability $1/2$. Set

$$y_n = \text{sgn}_n H_n,$$

then

$$\mathbb{E}(y_n | \mathcal{F}_{n-1}) = 0, \quad \text{and} \quad \text{var}(y_n | \mathcal{F}_{n-1}) = v_n^2 \tau_H^2.$$ 

Using the QML-notation from Appendix 3.B.1, the variable $y_n$ has a conditional mean $\mu_n = 0$, and a conditional variance $h_n = v_n^2 \tau_H^2$, yielding the Gaussian (quasi) log-likelihood, which up to an unimportant constant equals

$$L_N(\theta; y_1, \ldots, y_N) = -\frac{1}{2} \sum_{n=1}^{N} \left( \log(v_n^2(\gamma, \beta) \, \tau_H^2) + \frac{y_n^2}{v_n^2(\gamma, \beta) \, \tau_H^2} \right),$$

(3.12)

Due to the symmetry of the Gaussian likelihood, one does not need the signed values $(y_n)$ to evaluate the log-likelihood function (3.12); all that is needed$^2$ are the values $(H_n^2)$. So one may estimate the parameter $(\tau_H, \gamma, \beta)$ by QML based on $H_n$ as if the $H_n$ were conditionally mean zero, variance $v_n^2 \tau_H^2$ Gaussian distributed random variables. The special case $H_n = |r_n|$ (in which case $\tau_H = \tau$) yields the classical GARCH QMLE that uses the close-to-close returns $r_n$ (as indicated earlier, the expectation of $r_n$ is irrelevant to the likelihood (3.12); in general $\mathbb{E}r_n \neq 0$ is allowed). We refer to the QMLE that maximizes (3.12) as the Gaussian QMLE (based on $H_n$).

The Bollerslev-Wooldridge (1992) QML covariance matrix $V_0$, Appendix 3.B.1 equation (3.33), may be derived as follows for the Gaussian QMLE. For notational convenience we write

$$\sigma_{H,n} = v_n \tau_H.$$ 

(3.13)

Equation (3.13) suppresses the parameter $\theta$ in $\sigma_{H,n} = \sigma_{H,n}(\theta)$ for $\theta = (\tau_H, \gamma, \beta)$. Define the matrix $G_H$ by

$$G_H(\theta)_{i,j} = \mathbb{E}_0 \left[ \frac{1}{\sigma_{H,0}^4(\theta)} \left( \frac{\partial^2 \sigma_{H,0}^2(\theta)}{\partial \theta_i} \right) \left( \frac{\partial^2 \sigma_{H,0}^2(\theta)}{\partial \theta_j} \right) \right].$$

(3.14)

$^2$A similar argument of the irrelevance of the sign in a Gaussian likelihood may be applied for the estimation of Autoregressive Conditional Duration models (Engle and Russell, 1998), and for Multiplicative Error Models (Engle, 2002).
The matrices $A_0$ and $B_0$ from the appendix now read (to obtain $B_0$ use the independence of $Z_{H,n}$ and $\sigma_{H,n}$ as well as the equality $E(Z_H^2 - 1)^2 = \text{var}(Z_H^2)$),

$$A_0 = \frac{1}{2} G_H(\theta^0), \quad \text{and} \quad B_0 = \frac{1}{4} \text{var}(Z_H^2) G_H(\theta^0).$$

The covariance matrix $V_0 = A_0^{-1} B_0 A_0^{-1}$ now simplifies to the matrix given in (3.16) below.

Equation (3.16) implies that the noisiness of $H_n$, as given by $\text{var}(Z_H^2)$, plays an important role in the asymptotic variance matrix $V_0$. One might have anticipated this property.

The squared proxy $H_n^2 = v_n^2 Z_{H,n}^2$ is used in the likelihood as a conditionally unbiased estimator of the squared scale factor, $E(H_n^2|F_{n-1}) = v_n^2 \tau_H^2$. The quality of $H_n^2$ as an estimator of $v_n^4 \tau_H^4$ is determined by $\text{var}(Z_H^2)$:

$$\text{var}(H_n^2|F_{n-1}) = v_n^4 \tau_H^4 \text{var}(Z_H^2).$$

Let us now provide formal details for the Gaussian QMLE. One obtains the regularity conditions for the Gaussian QMLE by adjusting the five conditions from Appendix 3.B.3 for the QMLE based on close-to-close returns. Both the innovation $Z$ of the returns $(r_n = v_n \tau Z_n)$, and the innovation $Z_H$ of the proxies $(H_n = v_n \tau_H Z_{H,n})$ play a role. One has to adjust the condition $E(Z^4) < \infty$ to $E(Z_H^4) < \infty$, and replace $\tau$ by $\tau_H$ in condition 3.B.3(2). One has to keep $\tau^0$ in condition 3.B.3(3). Let $\Theta$ denote the parameter space.

We have the following assumptions:

(A1) $(Z_n)$ is an iid sequence with $E(Z^2) = 1$,

(A2) $\Theta$ is a compact subspace of the space given by $\tau_H > 0$, $\gamma > 0$, $\beta \in [0,1)$, and $\theta^0 \in \text{int } \Theta$.

(A3) $E \log (\gamma^0 (\tau^0)^2 Z^2 + \beta^0) < 0$,

(A4) $Z^2$ is non-degenerate,

(A5) $E(Z_H^4) < \infty$.

Condition (A1) does not imply $E(Z) = 0$: the daily return $r_n$ may have a non-zero mean. The only condition on $Z_H$ is (A5). Most conditions concern the innovation $Z$ of the close-to-close returns $r_n$. This is because $Z_n$ enters the process $(v_n)$ in equation (3.6), which is driven by the close-to-close returns. Compared with the classical QML estimator based on the close-to-close returns, one may now apply QML if $E(Z_H^4) = \infty$; the conditions require that one applies a proxy $H$ for which $E(Z_H^4) < \infty$. For more background on the conditions
3.2 QML Estimation Using a General Volatility Proxy

(A1) to (A5), see Appendix 3.B.3.

The five conditions above ensure that the likelihood (3.12) yields a consistent and asymptotically normal QMLE. This result is formalized in Theorem 3.1 below. Recall that $\mu_H^2 = \sqrt{E H^2(\Psi)}$.

**Theorem 3.1** Let $\theta^0 = (\tau_H^0, \gamma^0, \beta^0)$ and $\tau_H^0 = \tau^0 \mu_H^2$. Assume conditions (A1) to (A5). Then the Gaussian QMLE $\hat{\theta}_N$ based on $H_n$ is asymptotically normal:

$$\sqrt{N}(\hat{\theta}_N - \theta^0) \overset{d}{\to} \mathcal{N}(0, V_0), \quad N \to \infty,$$

with

$$V_0 = \text{var}(Z_H^2) \, G_H^{-1}(\theta^0).$$

The proof of Theorem 3.1 consists of an adjustment of the proof of Straumann and Mikosch (2006) for the QMLE based on the returns $y_n = r_n$ to the case that $y_n = H_n$, see Appendix 3.C.

Let us recall the notion of asymptotic relative efficiency. If two competing estimators $\hat{\phi}_N^{(1)}$ and $\hat{\phi}_N^{(2)}$ are $\sqrt{N}$-consistent and asymptotically normal estimators of a parameter $\phi$ with asymptotic variances $(\sigma_\phi^{(1)})^2$ and $(\sigma_\phi^{(2)})^2$, then the asymptotic relative efficiency (ARE) is given by

$$ARE = \frac{(\sigma_\phi^{(1)})^2}{(\sigma_\phi^{(2)})^2}.$$

The matrix $G_H$ in (3.14) depends on the parameters $(\tau_H, \gamma, \beta)$ and the distribution of the daily return innovation $Z$. It does not depend on the distribution of $Z_H$. Suppose that one considers QMLEs based on proxies with fixed norming parameter $\tau_H$ (e.g. $\tau_H = \tau$). Then one can speak of the relative efficiency of estimators for the fixed parameters $(\tau_H, \gamma, \beta)$. By equation (3.16) the asymptotic relative efficiency of the QMLEs is determined by the noisiness of $H$, as given by $\text{var}(Z_H^2)$. If one considers proxies with different $\tau_H$ then the asymptotic relative efficiency for $(\gamma, \beta)$ is still determined by $\text{var}(Z_H^2)$, as we shall make precise now.

The following lemma enables one to compare the QML covariance matrices $V_0$ for estimators of $\gamma$ and $\beta$ based on two different proxies $H$. The proof may be found in Appendix 3.C.

**Lemma 3.2** The $(\gamma, \beta)$-block of $G_H^{-1}(\theta^0)$ in Theorem 3.1 does not depend on the proxy $H$.

From Theorem 3.1 and Lemma 3.2 it follows that for finding the optimal Gaussian QMLE
one only has to check a simple criterion. The smaller the variance \( \text{var}(Z_H^2) \) the more efficient the QMLE, irrespective of the parameter values \((\gamma, \beta)\). Corollary 3.3 summarizes.

**Corollary 3.3** Consider two Gaussian QMLEs for \( \gamma \) and \( \beta \) from Theorem 3.1, the first based on proxies \( H'_n \) and the other based on \( H_n \). These estimators have asymptotic relative efficiency

\[
\text{ARE}_{\text{Gaussian}}(H', H) = \frac{\text{var}(Z_H^2)}{\text{var}(Z_{H'}^2)}. \tag{3.17}
\]

As a final remark, the results from this section are not limited to GARCH(1,1). Suppose that instead of the GARCH(1,1) variables \( v_n \) in (3.6) one has a GARCH\((p, q)\) specification, or another GARCH model. One can still introduce the intraday return processes \( R_n(\cdot) = v_n \tau \Psi_n(\cdot) \), and apply the QML-arguments employed in this section.

### 3.2.2 Log-Gaussian QMLE

It is also possible to estimate the parameters \((\gamma, \beta)\) of the GARCH system (3.3–3.4) by a log-Gaussian QMLE. This subsection develops a log-Gaussian QMLE, similarly to the Gaussian QMLE. Readers may prefer to skip Sections 3.2.2 to 3.2.4 upon first reading, and proceed directly to the empirical results from Section 3.3.

The log-Gaussian QMLE applies Gaussian quasi maximum likelihood to the log proxies \( \log(H_n) \). Here we assume that the proxies \( H_n \) are almost surely positive. Applying logarithms to \( H_n \) in (3.9) yields the equation \( \log(H_n) = \log(v_n) + \log(\tau_H) + \log(Z_{H,n}) \).

Define \( \tilde{\tau}_H = \tau_H \exp(\mathbb{E}\log(Z_{H,n})) \), and

\[
U_{H,n} = \frac{\log(Z_{H,n}) - \mathbb{E}\log(Z_{H,n})}{\sqrt{\text{var}(\log(Z_{H,n}))}}.
\]

We now obtain the additive equation

\[
\log(H_n) = \log(v_n) + \log(\tilde{\tau}_H) + \lambda U_{H,n}, \tag{3.18}
\]

with errors \( U_{H,n} \) which are iid mean zero, unit variance. The system (3.18) implies \( \mathbb{E}(\log(H_n)|\mathcal{F}_{n-1}) = \log(v_n) + \log(\tilde{\tau}_H) \), and \( \text{var}(\log(H_n)|\mathcal{F}_{n-1}) = \lambda^2 \). The parameter \( \lambda^2 \) represents the measurement variance of \( \log(H_n) \), as a proxy for the log scale factor. Define \( \lambda^0 \geq 0 \) by

\[(\lambda^0)^2 \equiv \text{var}(\log(Z_H)).\]
Set \( \tilde{\theta} = (\tilde{\tau}_H, \gamma, \beta) \) and define the extended parameter
\[
\eta \equiv (\tilde{\theta}, \lambda) = (\tilde{\tau}_H, \gamma, \beta, \lambda).
\]
The parameters \( \gamma, \beta \) in \( \tilde{\theta} \) are the same as the \( \gamma, \beta \) in the parameter \( \theta \) for the Gaussian QMLE from Section 3.2.1. The additive equation (3.18) fits into the framework of quasi maximum likelihood estimation (see Appendix 3.B.1), setting
\[
y_n = \log(H_n), \quad \mu_n(\eta) = \log(\sigma_{H,n}(\tilde{\theta})), \quad h_n(\eta) = \lambda^2.
\]
We refer to the maximizer \( \hat{\eta}_N \) as the log-Gaussian QMLE. Let us determine the QML covariance matrix \( V_0 \) from Appendix 3.B.1. The matrix \( A_0 \) is block diagonal since the mean and variance functions do not share parameters. Applying
\[
\frac{\partial \mu_n(\eta)}{\partial \eta_i} = \frac{1}{2\sigma_{H,n}^2(\tilde{\theta})} \frac{\partial^2 \sigma_{H,n}^2(\tilde{\theta})}{\partial \eta_i},
\]
one finds that the \( \tilde{\theta} \)-block and the diagonal element for \( \lambda \) of \( A_0 \) satisfy (for \( \lambda^0 > 0 \))
\[
(A_0)_{\tilde{\theta}} = \frac{1}{4(\lambda^0)^2} G_H(\tilde{\theta}^0), \quad (A_0)_{\lambda} = \frac{2}{(\lambda^0)^2},
\]
with \( G_H \) given by equation (3.14). The \( \tilde{\theta} \)-block of \( B_0 \) equals the \( \tilde{\theta} \)-block of \( A_0 \), the diagonal element for \( \lambda \) equals \( (B_0)_{\lambda} = \frac{1}{\lambda^0} \text{var}(U_H^2) \). The off-diagonal \( \tilde{\theta}, \lambda \)-column of \( B_0 \) equals
\[
(B_0)_{\tilde{\theta}, \lambda} = \frac{1}{(\lambda^0)^2} E U_H^3 E \frac{\partial \mu_n}{\partial \tilde{\theta}}(\tilde{\theta}^0)\\',
\]
making use of \( \mu_n(\eta) = \mu_n(\tilde{\theta}) \). The covariance matrix \( V_0 = A_0^{-1} B_0 A_0^{-1} \) divided into \( (\tilde{\theta}, \lambda) \)-blocks now reads
\[
V_0 = 4(\lambda^0)^2 \begin{pmatrix}
G^{-1}_H(\tilde{\theta}^0) & \frac{1}{2} E U_H^3 E \frac{\partial \mu_n}{\partial \tilde{\theta}}(\tilde{\theta}^0)\\' \\
\frac{1}{2} E U_H^3 E \frac{\partial \mu_n}{\partial \tilde{\theta}}(\tilde{\theta}^0) & \frac{1}{16} \text{var}(U_H^2)
\end{pmatrix}.
\]
By the QML theory from Appendix 3.B.1 the log-Gaussian QMLE \( \hat{\eta}_N \) is asymptotically normal under appropriate conditions\(^3\),
\[
\sqrt{N}(\hat{\eta}_N - \eta_0) \overset{d}{\to} N(0, V_0), \quad N \to \infty,
\]
with \( V_0 \) the covariance matrix given by (3.19). By the covariance matrix (3.19) smaller \( (\lambda^0)^2 \) yield more efficient QMLEs for \( \gamma \) and for \( \beta \). As in Corollary 3.3 the asymptotic relative efficiency of two log-Gaussian QMLEs for \( \gamma \) and \( \beta \) based on two different proxies \( H'_n \) and

\(^3\)Condition (A5) may have to be replaced by \( \mathbb{E}(\log(Z_H))^4 < \infty \). We do not prove precise conditions for the log-Gaussian QMLE here.
$H_n$ is given by

$$ARE_{\text{log-Gaussian}}(H', H) = \frac{\text{var}(\log(Z_{H'}))}{\text{var}(\log(Z_H))}.$$  \hspace{1cm} (3.21)

In a recent paper, de Vilder and Visser (2008) defined an optimal proxy $H^*$ as a proxy where the log has minimal variance:

$$\text{var}(\log(H^*(\Psi))) = \inf_H \text{var}(\log(H(\Psi))).$$

Such an optimal proxy yields the most efficient log-Gaussian QMLE for $\gamma$ and $\beta$.

We end this section with a remark regarding practical implementation of the log-Gaussian QMLE. The numerical value of $\hat{\lambda}$ is irrelevant to inference on $\hat{\theta}$, in particular to inference on $\gamma, \beta$. First, the numerical value of $\hat{\lambda}$ does not influence the numerical values of the parameters in $\hat{\theta}$. This is due to the well-known effect that the value of the variance parameter does not influence the value of the mean parameter for Gaussian QML (this is true if the variance function and the mean function do not share parameters). Moreover, the usual “sandwich” QML covariance matrix $\hat{V} = \hat{A}^{-1}\hat{B}\hat{A}^{-1}$ also does not depend on the numerical value of $\hat{\lambda}$ as far as the $\hat{\theta}$-parameters are concerned.

### 3.2.3 Efficiency of log-Gaussian QMLE versus Gaussian QMLE

Let us briefly compare the asymptotic efficiency of $\hat{\gamma}, \hat{\beta}$ for the log-Gaussian and Gaussian QMLE. Comparing the $(\gamma, \beta)$-blocks of $V_0$ in equations (3.16) and (3.19), one finds that the asymptotic relative efficiency of the log-Gaussian and Gaussian QMLEs for $\gamma$ and $\beta$, based on the same proxy $H_n$, is given by (using $\text{var}(\log(Z_H^2)) = 4 (\lambda^0)^2$)

$$ARE(\text{log-Gaussian, Gaussian}) = \frac{\text{var}(\log(Z_H^2))}{\text{var}(Z_H^2)}. \hspace{1cm} (3.22)$$

The log-Gaussian QMLE is the more efficient one if $\text{var}(\log(Z_H^2)) \leq \text{var}(Z_H^2)$. This inequality does not always hold: $\text{var}(Z_H^2)$ tends to be large if $Z_H$ has heavy tails, but $\text{var}(\log(Z_H^2))$ tends to be large if $Z_H$ has values close to zero. In the following example $Z_H$ has a lognormal distribution (and $\mathbb{E}Z_H^2 = 1$).

**Example 3.2.3.1** Let $Z_H$ have a lognormal$(-\sigma^2, \sigma^2)$ distribution. Then $\log(Z_H) \sim \mathcal{N}(-\sigma^2, \sigma^2)$. The $j$-th moment of a lognormal$(\mu, \sigma^2)$ equals $e^{j\mu + j^2\sigma^2/2}$, so $\mathbb{E}Z_H^2 = 1$ and $\text{var}(Z_H^2) = e^{4\sigma^2} - 1$. Apply relation (3.22) to find

$$ARE(\text{log-Gaussian, Gaussian}) = \frac{4\sigma^2}{e^{4\sigma^2} - 1}.$$
3.2 QML Estimation Using a General Volatility Proxy

Since $4\sigma^4 \leq e^{4\sigma^2} - 1$ the log-Gaussian QMLE is more efficient for all values of $\sigma^2$. In this example the log-Gaussian QMLE is the exact maximum likelihood estimator.

3.2.4 Relative Error of Volatility Extraction

Sometimes one is interested in the quality of the estimator of the scale factor $\sigma_{H,n} = v_n \tau_H$, for a fixed day $n$. The extraction $\theta \rightarrow \hat{\sigma}_{H,n}(\theta)$, with initialization $\hat{v}_0$ is a function of $\theta$. To simplify the notation we omit the hat on $\sigma_{H,n}$ in this section. If we plug in the estimator $\hat{\theta}_N$, we obtain the estimated extraction $\sigma_{H,n}(\hat{\theta}_N)$. The asymptotic distribution of $\sigma_{H,n}(\hat{\theta}_N)$ for $N \rightarrow \infty$ may be found by the Delta method. Let the row vector $\hat{\sigma}_{H,n}$ denote the derivative of $\sigma_{H,n}$ with respect to $\theta$. The matrix $V_0$ denotes the asymptotic covariance matrix for $\hat{\theta}_N$. The Delta method gives

$$\sqrt{N}(\sigma_{H,n}(\hat{\theta}_N) - \sigma_{H,n}(\theta^0)) \rightarrow N(0, \hat{\sigma}_{H,n}(\theta^0) V_0 \hat{\sigma}_{H,n}(\theta^0)^t), \quad N \rightarrow \infty,$$

for fixed $n$. It is natural to look at the relative error of $\sigma_{H,n}$:

$$\text{re}(\sigma_{H,n}) = \frac{\sigma_{H,n}(\hat{\theta}_N)}{\sigma_{H,n}(\theta^0)} - 1.$$

The relative error itself is not observed. One may estimate its variance by

$$\frac{1}{\sigma^2_{H,n}} \overline{\text{var}}(\sigma_{H,n}),$$

where $\overline{\text{var}}(\sigma_{H,n})$ denotes the empirical counterpart of the variance as implied by equation (3.23). The estimate (3.24) does not depend on $\hat{\tau}_H$, see formula (3.38) in Appendix 3.C. So the asymptotic variance of the relative error is proportional to $\text{var}(Z_H^2)$ and $\text{var}(\log(Z_H))$, for Gaussian and log-Gaussian estimation.

For practical implementation one needs the derivatives $\hat{\sigma}_{H,n}(\hat{\theta}_N)$. One has $h_n(\theta) = \sigma^2_{H,n}(\theta)$. The analytical derivatives $h_n$ in $\theta = \hat{\theta}_N$ are available from the optimization procedure, so one may estimate the variance $\overline{\text{var}}(\sigma_{H,n})$ in equation (3.24) by a straightforward application of the Delta method, making use of the chain rule:

$$\hat{\sigma}_{H,n}(\hat{\theta}_N) = \frac{1}{2\sigma_{H,n}(\hat{\theta}_N)} \hat{h}_n(\hat{\theta}_N).$$

Of course, if one wishes to construct a confidence interval for $v_n \tau$, instead of $v_n \tau_H$, one has to carry out estimation based on a proxy that has $\tau_H \equiv \tau$, for instance the absolute returns $|r_n|$.  

3.3 Empirical Efficiency Gains for the S&P 500 Index

This section examines the empirical increase in efficiency resulting from the use of volatility proxies in estimating the GARCH(1,1) parameters $\gamma$ and $\beta$. The analysis is carried out for both the Gaussian and the log-Gaussian QMLE. The estimates in this section use 1001 days of S&P 500 index tick data over the period 1992–1995. For a description of the data, see Appendix 3.A. We use this time period, since it is a fairly stable period without clear structural breaks in the level of volatility, see Figure 3.3 in Appendix 3.A. We try to avoid structural breaks, since it is well known that GARCH parameter estimation may break down in the presence of such breaks. Parameter estimators may no longer be consistent, and the volatility persistence tends to be overestimated if the level of volatility has a change-point, see Mikosch and Starica (2004), and Hillebrand (2005).

Figure 3.1 gives an impression of the empirical efficiency gains. It shows four 95% confidence ellipses for estimates of $(\gamma, \beta)$, based on the absolute daily return $|r_n|$ and on three other proxies. The confidence regions are calculated using Bollerslev-Wooldridge (1992) QML covariances. The ellipses may be considered to give a first, and reasonable, indication of the differences in accuracy between estimators based on different proxies. The estimators all make use of GARCH(1,1) dynamics, which is generally found to give a...
3.3 Empirical Efficiency Gains for the S&P 500 Index

A good description of the daily volatility process (see e.g. Andersen and Bollerslev, 1998a, and Hansen and Lunde, 2005a). The confidence regions for the QML estimators do not rely on specific assumptions for the intraday price process. They are valid for any standard process $\Psi$. The figure suggests that there are great differences in accuracy for different proxies. We shall use the results from Section 3.2 to gain insight into these differences.

Table 3.1 provides an efficiency factor that expresses the efficiency gain with respect to the standard GARCH(1,1) QMLE (as $1/\text{ARE}$). The relative efficiency of the QMLE using the proxy $H$ is determined by the variance of $Z^2_\hat{H}$ or the variance of its logarithm. We consider four volatility proxies $H$. For each proxy $H$ we estimate the parameters by both the Gaussian and the log-Gaussian QMLE. We then use the standardized residuals, $\hat{Z}_{H,n}$, to compare the quality of the estimators. The proxy $H^{(\hat{w})}$ combines the sum of the ten-minute highs, the sum of the ten-minute lows and the sum of the ten-minute absolute returns as

$$H^{(\hat{w})}_n = (\text{RAV}10\text{HIGH}_n)^{1.04}(\text{RAV}10\text{LOW}_n)^{0.72}(\text{RAV}10_{n})^{-0.76}.$$  (3.26)

The exponents add to one so as to ensure that $H^{(\hat{w})}_n$ is homogeneous of degree one. It is a good proxy for S&P 500 daily volatility$^4$, see de Vilder and Visser (2008).

<table>
<thead>
<tr>
<th>$H$</th>
<th>Gaussian $\text{\bar{v}ar}(Z^2_\hat{H})$</th>
<th>eff. factor</th>
<th>log-Gaussian $\text{\bar{v}ar}(\log(Z^2_\hat{H}))$</th>
<th>eff. factor</th>
</tr>
</thead>
<tbody>
<tr>
<td>$</td>
<td>r_n</td>
<td>$</td>
<td>3.34</td>
<td>1</td>
</tr>
<tr>
<td>hl</td>
<td>1.41</td>
<td>2.4</td>
<td>0.68</td>
<td>4.9</td>
</tr>
<tr>
<td>$RV^{(81)}$</td>
<td>0.48</td>
<td>7.0</td>
<td>0.25</td>
<td>13.2</td>
</tr>
<tr>
<td>$H^{(\hat{w})}$</td>
<td>0.23</td>
<td>14.8</td>
<td>0.17</td>
<td>20.1</td>
</tr>
</tbody>
</table>

Table 3.1: Empirical QMLE relative efficiency for four volatility proxies: absolute return, high-low, realized volatility using 81 five-minute intervals (see (3.7)), and $H^{(\hat{w})}$ (see (3.26)). The table reports $\text{\bar{v}ar}(Z^2_\hat{H})$ and $\text{\bar{v}ar}(\log(Z^2_\hat{H}))$, see Sections 3.2.1 and 3.2.2. The numbers are based on residuals of GARCH(1,1) estimation of the S&P 500 index over 1992-01-01 to 1995-12-31, or 1001 observations. The Gaussian QMLE with $H = |r|$ is the classical GARCH QMLE. The efficiency factor is the gain with respect to the classical GARCH(1,1) QMLE, expressed as $1/\text{ARE}$, so $2.4=3.34/1.41$.

Moving down in Table 3.1 from absolute returns $|r_n|$ to $H^{(\hat{w})}$ reveals an efficiency gain by a factor 15 for the Gaussian QMLE. The log-Gaussian QMLE based on $H^{(\hat{w})}$ yields an efficiency gain by a factor 20. This means that estimation of $(\gamma, \beta)$ based on $\log(H^{(\hat{w})})$ needs roughly 20 times fewer days of observations than the usual QMLE that uses squared close-to-close returns to obtain the same precision for the parameter estimates.

$^4$The ten-minute high is obtained by the difference of the maximum of $R_n(\cdot)$ and the starting value of $R_n(\cdot)$ over the ten-minute interval in question. The lows are obtained similarly, and made positive by taking absolute values. The weights are obtained by a procedure that minimizes the variance of the error of the proxy.
are no entries for $H = |r|$ for the log-Gaussian QMLE since these would involve taking the log of zeros. The table reflects the differences in the confidence regions shown in Figure 3.1. In the figure the estimate based on $|r_n|$ is situated below and to the left of the other estimates. The simulations in Section 3.4 show a similar effect, suggesting that the classical QMLE based on $|r_n|$ suffers more severely from finite-sample bias than the estimators based on the other proxies.

The log-Gaussian QMLE outperforms the Gaussian QMLE for the proxies $hl$, $RV^{(81)}$, and $H^{(w)}$. One possible interpretation is that these proxies are closer to having the distribution of a lognormal random variable than to the absolute value of a Gaussian random variable. Indeed, in empirical research it has been found that log realized volatility and the log high-low range may have a distribution that is fairly symmetric and approximately Gaussian, see for instance Andersen, Bollerslev, Diebold, and Ebens (2001), and Alizadeh, Brandt, and Diebold (2002).

Let us briefly examine the error in the extraction of the scale factors. We apply the Delta method from Section 3.2.4 to obtain the standard errors of the relative error in the extraction. Table 3.2 lists these standard errors for the final scale factors $\sigma_{H,n}$, $n = N = 1001$. The first entry, 3.8%, suggests that the interval $\hat{\sigma}_{H,1001} \pm 7.6\%$ encloses the true $\sigma_{H,1001}$ with probability 95%. The log-Gaussian QMLE that uses $H^{(w)}$ gives a more than 4 times tighter interval.

We also checked what Tables 3.1 and 3.2 would look like for the larger sample over the years 1988–2006, $n = 1, \ldots, 4575$, (ignoring possible structural breaks). The results for the larger sample are similar to the patterns in Tables 3.1 and 3.2, though the efficiency gains in Table 3.1 get more pronounced: instead of a factor 20 for the log-Gaussian QMLE based on $H^{(w)}$, the estimates suggest a gain by a factor more than 40.

One should not interpret these percentages as typical for this GARCH(1,1) process: they depend on the path of the process before $n = 1001$.

<table>
<thead>
<tr>
<th>$H$</th>
<th>Gaussian s.e.($re_N$) %</th>
<th>log-Gaussian s.e.($re_N$) %</th>
</tr>
</thead>
<tbody>
<tr>
<td>$</td>
<td>r</td>
<td>$</td>
</tr>
<tr>
<td>$hl$</td>
<td>2.2</td>
<td>1.7</td>
</tr>
<tr>
<td>$RV^{(81)}$</td>
<td>1.2</td>
<td>1.0</td>
</tr>
<tr>
<td>$H^{(w)}$</td>
<td>0.9</td>
<td>0.8</td>
</tr>
</tbody>
</table>

Table 3.2: Estimates of the standard error of the relative error in $\hat{\sigma}_{H,1001}$. The quantities reported are $100 \times \text{s.e.}(\hat{\sigma}_{H,N})/\hat{\sigma}_{H,N}$, see also equation (3.24). Numbers are based on the same volatility proxies and data as in Table 3.1.
3.4 Simulation Study

The estimates for the S&P 500 in Section 3.3 use one sample path only. To explore the finite-sample properties of the QML estimators, and determine whether the asymptotic efficiency gains as described in Theorem 3.1 are also present in finite samples, we perform simulations. For simulations of the classical QMLE based on close-to-close returns for the GARCH(1,1) parameters, see also Bollerslev and Wooldridge (1992), Lumsdaine (1995), Fiorentini, Calzolari, and Panattoni (1996), and Straumann (2005). The simulations in this chapter focus on the difference between the inference based on the close-to-close returns $H_n = |r_n|$ and inference by the square root of realized variance

$$H_n = RV_n^{(m)} = \sqrt{RQV_n^{(m)}} = \left(\sum_{k=1}^{m} r_{n,k}^2\right)^{1/2}.$$

In principle the simulations can be used to analyze any positive and positively homogeneous proxy. The simulations here focus on the realized variance with $m = 81$ intervals. This corresponds to the five-minute realized volatility, which is a widely applied proxy in practice. To generate the realized variance one has to simulate the process $\Psi_n(\cdot)$ at $(m + 1)$ equidistant points in $[0, 1]$. In the scaling model the standard process $\Psi(\cdot)$ may be any process. To generate sample paths, we have to choose a specific process $\Psi(\cdot)$. A Brownian motion will not do: the realized volatility using $m = 81$ intervals would yield unrealistically precise parameter estimates (the Gaussian QMLE would have a gain in efficiency by a factor 81).

We shall consider a diffusion with a diffusion coefficient that is random and varies in time. For simplicity we assume that the logarithm of the diffusion coefficient is a stationary process defined by an Ornstein-Uhlenbeck equation (see also Scott, 1987, and Wiggins, 1987). The resulting process is continuous on $[0, 1]$. For the log of the diffusion coefficient we have

$$dY_n(u) = -\delta(Y_n(u) - \mu)du + \sigma_Y dB_n^{(2)}(u). \tag{3.27}$$

The process $\Psi_n(\cdot)$ used in the simulations is given by

$$d\Psi_n(u) = \exp(Y_n(u)) dB_n^{(1)}(u), \quad u \in [0, 1]. \tag{3.28}$$

The Brownian motions $B_n^{(1)}$ and $B_n^{(2)}$ are uncorrelated, and $\Psi_n(0) = 0$. We sample $Y(0)$ from its stationary distribution. For $\mu = -\sigma^2_Y/(2\delta)$, the realized variance $RQV_n^{(m)}(\Psi_n)$ for all $m$, as well as the quadratic variation over the unit interval, have expectation 1, see
Appendix 3.D. Choose
\[ \delta = \frac{1}{2}, \quad \sigma_Y = \frac{1}{4}, \quad \mu = -\frac{1}{16}. \]
Then the return innovations \( Z_n \) satisfy
\[ \mathbb{E}Z_n^2 = 1, \quad \text{var}(Z_n^2) \approx 2.77. \]
For the realized volatility we take \( m = 81 \) intervals, yielding innovations \( Z_{H,n} \) that satisfy
\[ \mathbb{E}Z_{H,n}^2 = 1, \quad \text{var}(Z_{H,n}^2) \approx 0.27, \quad \text{var}(\log(Z_{H,n}^2)) \approx 0.24. \quad (3.29) \]
The simulations consist of 10000 replications. First generate 10000 sets of 2500 days of realizations of \( \Psi_n(\cdot) \). For each sequence \( (\Psi_n), n = 1, \ldots, 2500 \), we generate the paths \( (v_n, \tau) \) for five different configurations \( (\gamma, \beta) \), fixing \( \tau = 1 \). One may now examine the finite-sample properties of the GARCH(1,1) QMLEs \( (\hat{\gamma}, \hat{\beta}) \) for sample lengths \( N = 250, 500, 1000, 2500 \). Figure 3.2 shows the estimates for 1000 such paths for \( (\gamma, \beta) = (0.05, 0.9) \) and sample length \( N = 1000 \) days. The left figures are based on absolute returns as a volatility proxy, the right figures are based on the realized volatility \( RV_n^{(81)} \). The estimates that use the theoretically more efficient proxy \( RV_n^{(81)} \) are indeed more concentrated around the true parameter value, and have no outliers. Part (a1) shows that for \( H_n = |r_n| \) in some instances the optimization routine finds a maximum of the likelihood at a value \( \hat{\gamma} < 0 \), which is outside of the admissible parameter region.\(^6\) Using the more efficient proxy \( H_n = RV_n^{(81)} \) resolves this problem, see part (b1).

Table 3.3 provides a more complete overview of the finite-sample properties than Figure 3.2. The first two rows list 100 \( \times \) the bias and 100 \( \times \) the root mean square error (RMSE) of \( \hat{\gamma} \) for \( (\tau, \gamma, \beta) = (1, 0.05, 0.9) \). The first four columns in the first row contain the biases for the return based GARCH(1,1) QMLE, for increasing sample sizes. The next eight columns give this bias using the volatility proxy \( RV_n^{(81)} \), first for the Gaussian and then for the log-Gaussian QMLE.

While the small-sample biases of \( \hat{\gamma}, \hat{\beta} \) tend to be substantial for the return based QMLE, they are moderate to negligible for the realized volatility based QMLE. The asymptotic relative efficiencies with respect to the usual GARCH(1,1) QMLE may be deduced from equation (3.29) and equations (3.17) and (3.21). For the square root of realized variance this yields an efficiency factor \( 2.77/0.27 \approx 10 \) for the Gaussian QMLE and efficiency factor 11 for the log-Gaussian QMLE. So one may expect that the RMSE for \( H_n = RV_n \) is more than a factor three smaller for large samples. This factor reflects

\[ ^6 \text{In theory a negative } \hat{\gamma} \text{ could lead to negative values of } \hat{\sigma}_n^2. \text{ We made sure that the in-sample values } \hat{\sigma}_n^2 \text{ are strictly positive by assigning a large negative number to the likelihood for when a negative value of } \hat{\sigma}_n^2 \text{ occurs.} \]
the difference in RMSE between using returns or realized volatility, for \( N = 2500 \). For smaller sample sizes the efficiency gain is larger, suggesting that return based estimation suffers more from small-sample effects. The quality of the parameter estimates using 250 observations of realized volatility resembles using somewhere between 1000–2500 close-to-close returns. As predicted by the asymptotic efficiency factors for \( RV_n \) calculated above (11 versus 10), the log-Gaussian QMLE does slightly better than the Gaussian QMLE in these simulations (recall that for the S&P data the log-Gaussian QMLE appears substantially more efficient).

![Diagram](image-url)

**Figure 3.2**: Scatters for \((\hat{\gamma}, \hat{\beta})\) plane, 1000 sample paths \((\tau = 1, \gamma = 0.05, \beta = 0.9)\). The intraday process \(\Psi(\cdot)\) is given by equations (3.28) and (3.27) with \((\delta = 0.5, \sigma_Y = 0.25, \mu = -0.0625)\). Upper and lower left: estimates based on absolute returns (Gaussian QMLE). Upper and lower right: realized volatility (Gaussian QMLE). Figure (a) leaves out four points where \(\hat{\gamma} > 0.2\). Figures (b) and (d) contain all points.
<table>
<thead>
<tr>
<th>param</th>
<th>$N=250$</th>
<th>$N=500$</th>
<th>$N=1000$</th>
<th>$N=2500$</th>
<th>$N=250$</th>
<th>$N=500$</th>
<th>$N=1000$</th>
<th>$N=2500$</th>
<th>$N=250$</th>
<th>$N=500$</th>
<th>$N=1000$</th>
<th>$N=2500$</th>
</tr>
</thead>
<tbody>
<tr>
<td>$\gamma$ 0.05 bias</td>
<td>-4.8</td>
<td>-1.4</td>
<td>-0.2</td>
<td>-0.1</td>
<td>0.3</td>
<td>0.0</td>
<td>-0.0</td>
<td>-0.0</td>
<td>0.2</td>
<td>0.0</td>
<td>-0.0</td>
<td>-0.0</td>
</tr>
<tr>
<td>RMSE</td>
<td>13.6</td>
<td>8.3</td>
<td>3.8</td>
<td>1.7</td>
<td>3.0</td>
<td>1.4</td>
<td>0.9</td>
<td>0.5</td>
<td>2.7</td>
<td>1.3</td>
<td>0.9</td>
<td>0.5</td>
</tr>
<tr>
<td>$\beta$ 0.9 bias</td>
<td>-4.0</td>
<td>-4.7</td>
<td>-2.7</td>
<td>-0.9</td>
<td>-1.1</td>
<td>-0.4</td>
<td>-0.2</td>
<td>-0.1</td>
<td>-1.0</td>
<td>-0.4</td>
<td>-0.2</td>
<td>-0.1</td>
</tr>
<tr>
<td>RMSE</td>
<td>22.5</td>
<td>17.3</td>
<td>10.3</td>
<td>4.0</td>
<td>5.2</td>
<td>2.5</td>
<td>1.6</td>
<td>0.9</td>
<td>4.8</td>
<td>2.3</td>
<td>1.5</td>
<td>0.9</td>
</tr>
<tr>
<td>$\gamma$ 0.15 bias</td>
<td>-1.7</td>
<td>-0.7</td>
<td>-0.3</td>
<td>-0.2</td>
<td>0.0</td>
<td>-0.1</td>
<td>-0.1</td>
<td>-0.0</td>
<td>-0.0</td>
<td>-0.1</td>
<td>-0.0</td>
<td>-0.0</td>
</tr>
<tr>
<td>RMSE</td>
<td>13.6</td>
<td>8.1</td>
<td>5.3</td>
<td>3.3</td>
<td>4.1</td>
<td>2.5</td>
<td>1.7</td>
<td>1.0</td>
<td>3.9</td>
<td>2.3</td>
<td>1.6</td>
<td>1.0</td>
</tr>
<tr>
<td>$\beta$ 0.8 bias</td>
<td>-5.1</td>
<td>-2.2</td>
<td>-1.0</td>
<td>-0.4</td>
<td>-0.4</td>
<td>-0.2</td>
<td>-0.1</td>
<td>-0.0</td>
<td>-0.3</td>
<td>-0.1</td>
<td>-0.1</td>
<td>-0.0</td>
</tr>
<tr>
<td>RMSE</td>
<td>17.4</td>
<td>8.4</td>
<td>4.7</td>
<td>2.7</td>
<td>3.1</td>
<td>2.0</td>
<td>1.3</td>
<td>0.8</td>
<td>2.9</td>
<td>1.9</td>
<td>1.3</td>
<td>0.8</td>
</tr>
<tr>
<td>$\gamma$ 0.35 bias</td>
<td>-0.2</td>
<td>-0.2</td>
<td>-0.1</td>
<td>-0.1</td>
<td>0.0</td>
<td>-0.1</td>
<td>-0.0</td>
<td>-0.0</td>
<td>-0.1</td>
<td>-0.1</td>
<td>-0.0</td>
<td>-0.0</td>
</tr>
<tr>
<td>RMSE</td>
<td>21.3</td>
<td>13.3</td>
<td>9.2</td>
<td>5.7</td>
<td>6.4</td>
<td>4.2</td>
<td>2.8</td>
<td>1.8</td>
<td>6.0</td>
<td>3.9</td>
<td>2.7</td>
<td>1.7</td>
</tr>
<tr>
<td>$\beta$ 0.6 bias</td>
<td>-3.4</td>
<td>-1.3</td>
<td>-0.7</td>
<td>-0.3</td>
<td>-0.2</td>
<td>-0.1</td>
<td>-0.1</td>
<td>-0.0</td>
<td>-0.2</td>
<td>-0.1</td>
<td>-0.1</td>
<td>-0.0</td>
</tr>
<tr>
<td>RMSE</td>
<td>13.5</td>
<td>8.0</td>
<td>5.2</td>
<td>3.2</td>
<td>3.5</td>
<td>2.3</td>
<td>1.6</td>
<td>1.0</td>
<td>3.3</td>
<td>2.2</td>
<td>1.5</td>
<td>0.9</td>
</tr>
<tr>
<td>$\gamma$ 0.25 bias</td>
<td>0.3</td>
<td>0.1</td>
<td>0.0</td>
<td>0.0</td>
<td>0.0</td>
<td>-0.0</td>
<td>-0.0</td>
<td>-0.0</td>
<td>0.0</td>
<td>-0.0</td>
<td>-0.0</td>
<td>-0.0</td>
</tr>
<tr>
<td>RMSE</td>
<td>18.3</td>
<td>10.4</td>
<td>7.0</td>
<td>4.3</td>
<td>4.8</td>
<td>3.2</td>
<td>2.2</td>
<td>1.3</td>
<td>4.5</td>
<td>3.0</td>
<td>2.0</td>
<td>1.3</td>
</tr>
<tr>
<td>$\beta$ 0.6 bias</td>
<td>-5.9</td>
<td>-2.5</td>
<td>-1.2</td>
<td>-0.5</td>
<td>-0.4</td>
<td>-0.2</td>
<td>-0.1</td>
<td>-0.0</td>
<td>-0.4</td>
<td>-0.2</td>
<td>-0.1</td>
<td>-0.0</td>
</tr>
<tr>
<td>RMSE</td>
<td>20.8</td>
<td>12.3</td>
<td>7.7</td>
<td>4.7</td>
<td>5.1</td>
<td>3.3</td>
<td>2.3</td>
<td>1.4</td>
<td>4.8</td>
<td>3.2</td>
<td>2.2</td>
<td>1.3</td>
</tr>
<tr>
<td>$\gamma$ 0.05 bias</td>
<td>-6.8</td>
<td>-1.9</td>
<td>0.4</td>
<td>0.4</td>
<td>0.8</td>
<td>0.2</td>
<td>0.0</td>
<td>0.0</td>
<td>0.7</td>
<td>0.2</td>
<td>0.1</td>
<td>0.0</td>
</tr>
<tr>
<td>RMSE</td>
<td>26.5</td>
<td>18.8</td>
<td>9.1</td>
<td>2.2</td>
<td>3.5</td>
<td>1.7</td>
<td>1.1</td>
<td>0.6</td>
<td>3.4</td>
<td>1.5</td>
<td>1.0</td>
<td>0.6</td>
</tr>
<tr>
<td>$\beta$ 0.8 bias</td>
<td>-10.0</td>
<td>-10.3</td>
<td>-6.7</td>
<td>-3.0</td>
<td>-3.0</td>
<td>-1.3</td>
<td>-0.6</td>
<td>-0.2</td>
<td>-2.8</td>
<td>-1.1</td>
<td>-0.5</td>
<td>-0.2</td>
</tr>
<tr>
<td>RMSE</td>
<td>37.9</td>
<td>33.5</td>
<td>24.1</td>
<td>13.2</td>
<td>15.0</td>
<td>7.9</td>
<td>4.8</td>
<td>2.8</td>
<td>14.2</td>
<td>7.3</td>
<td>4.5</td>
<td>2.6</td>
</tr>
<tr>
<td>$\gamma$ 0.05 bias</td>
<td>-5.4</td>
<td>-2.2</td>
<td>-0.9</td>
<td>-0.4</td>
<td>-0.6</td>
<td>-0.1</td>
<td>-0.1</td>
<td>-0.1</td>
<td>-0.6</td>
<td>-0.2</td>
<td>-0.1</td>
<td>-0.1</td>
</tr>
<tr>
<td>RMSE</td>
<td>7.7</td>
<td>4.7</td>
<td>3.1</td>
<td>1.9</td>
<td>2.9</td>
<td>2.0</td>
<td>1.2</td>
<td>0.6</td>
<td>2.7</td>
<td>1.9</td>
<td>1.1</td>
<td>0.6</td>
</tr>
<tr>
<td>$\beta$ 0.94 bias</td>
<td>-0.6</td>
<td>-1.8</td>
<td>-0.9</td>
<td>-0.2</td>
<td>-0.6</td>
<td>-0.2</td>
<td>-0.1</td>
<td>-0.0</td>
<td>-0.6</td>
<td>-0.2</td>
<td>-0.1</td>
<td>-0.0</td>
</tr>
<tr>
<td>RMSE</td>
<td>14.5</td>
<td>8.5</td>
<td>3.5</td>
<td>1.2</td>
<td>2.1</td>
<td>1.0</td>
<td>0.6</td>
<td>0.4</td>
<td>2.0</td>
<td>0.9</td>
<td>0.6</td>
<td>0.3</td>
</tr>
</tbody>
</table>

Table 3.3: Sampling distributions of GARCH(1,1) QMLE, using 10000 replications. The intraday process $\Psi_n(\cdot)$ is given by equations (3.28) and (3.27) with $\delta = 0.5$, $\sigma_Y = 0.25$, $\mu = -0.0625$. All simulations set the norming parameter $\tau = 1$. From top to bottom there are six panels of different parameters ($\gamma, \beta$). For each parameter setting the table gives $100 \times$ the bias and $100 \times$ the root mean squared error of $\hat{\gamma}$ and $\hat{\beta}$, for different lengths of the time series: 250, 500, 1000, 2500. The case $H_n = |r_n|$, using the Gaussian QML is the classical GARCH(1,1) QMLE. The cases $H_n = RV^{(m=81)}_n$ give the results for the realized volatility based on 81 intraday returns as a volatility proxy, using the Gaussian and the log-Gaussian QMLE, see Sections 3.2.1 and 3.2.2.
3.5 Relation to Semimartingale Models

In recent years the use of high-frequency data for volatility measurement and modelling has developed rapidly. A common approach is to assume that the continuous time log-return process is a semimartingale. In semimartingale theory the quadratic variation plays a central role. It is interpreted as squared volatility. The quadratic variation (QV) is the limit of the sum of squared intraday returns as the lengths of the sampling intervals approach zero, see for instance Protter (2005). For more statistical discussions, see Barndorff-Nielsen and Shephard (2002b) and Andersen, Bollerslev, Diebold, and Labys (2003). Alizadeh, Brandt and Diebold (2002), Barndorff-Nielsen and Shephard (2002a), and Haug, Klüppelberg, Lindner, and Zapp (2007) estimated parameters for particular parametric continuous time semimartingales.

In the scaling model introduced in Section 3.1 the log-return process is a semimartingale if Ψ(·) is a semimartingale. This might suggest that one should look at the quadratic variation, rather than bothering with likelihood methods. For standard Brownian motion quadratic variation over the unit time interval equals 1, so \( QV_n \equiv QV(R_n) = v_n^2 \tau^2 \). In general however, we do not have this exact relationship. If \( Ψ_n(·) \) is a martingale (hence \( R_n(·) \) too), the quadratic variation of \( R_n(·) \) is merely an unbiased estimator of the squared scale factor (or the conditional variance of the daily return),

\[
\mathbb{E}(QV_n | F_{n-1}) = \text{var}(r_n | F_{n-1}) = v_n^2 \tau^2.
\]

A common interpretation (e.g. Andersen, Bollerslev, Diebold, and Labys, 2003) is that quadratic variation is a natural (ex-post) estimator for the squared scale factor (or conditional variance). Another insightful interpretation is that the squared scale factor is the best forecaster of quadratic variation. Quadratic variation and scale factor are two different concepts for the fluctuations in the intraday price process. Quadratic variation is a measure pertaining to the sample path fluctuations during the trading day. The scale factor is a latent variable that determines the expected fluctuations during the trading day. Since \( v_n \tau \) is latent, one has to rely on statistical techniques for parameter estimation and extraction of the scale factors. In the case of intraday Brownian motion one happens to be in the fortunate circumstance of having a perfect proxy, \( H_n = v_n \tau H \) (here \( Z_{H,n} \equiv 1 \)). In this case the QML estimation introduced in this chapter yields perfect parameter estimates.

It should be noted that high-frequency data do not allow one to calculate quadratic variation to any desired degree of precision. In empirical applications it turns out that the use of intervals below a certain time unit (in the order of five-minutes for liquid equity markets) does not improve the estimates. On small time scales the price process shows a square saw-tooth effect that cannot be interpreted as a semimartingale. See Hansen
and Lunde (2006b). It may be tempting to posit that quadratic variation is observed, up to “small” noise. Barndorff-Nielsen and Shephard (2002b) showed that realized variance may actually have quite large noise in practical situations (with confidence intervals for volatility larger than the level of volatility itself). In practice volatility cannot be regarded as observable.

### 3.6 Conclusions

This chapter has developed a GARCH quasi maximum likelihood estimator that uses intraday volatility proxies, as a generalization of the classical QMLE based on daily close-to-close returns. One may achieve a substantial efficiency gain by using a volatility proxy different from the absolute or squared close-to-close return. The chapter analysed the situation where the daily close-to-close returns \( r_n \) satisfy the GARCH(1,1) system

\[
\begin{align*}
  r_n &= v_n \tau Z_n, \\
  v_n^2 &= 1 + \gamma r_{n-1}^2 + \beta v_{n-1}^2.
\end{align*}
\]

The intraday return process \( R_n(\cdot) \) is incorporated into this system by the scaling model. For each day \( n \) one has \( R_n(u) = v_n \tau \Psi_n(u), \quad u \in [0, 1] \), where the sequence of processes \( \Psi_n(\cdot) \) is iid. The setup does not impose assumptions on the process \( \Psi(\cdot) \). One obtains sharp estimators \( \hat{\gamma}, \hat{\beta} \) by making use of a suitable volatility proxy \( H(R_n) \), such as the realized volatility. For the S&P 500 index data the estimated variances of the estimators decrease by a factor 20 for the proxy \( H(\hat{w}) \) in (3.26).

The intraday model introduced in the chapter and the resulting QML theory can be used for other discrete time volatility models. It would be interesting to apply the methods of this chapter to asymmetric GARCH models, or to models where the variable \( v_n \) is driven by statistics different from the squared return \( r_{n-1}^2 \). For instance Andersen, Bollerslev, Diebold, and Labys (2003) showed that realized volatilities are useful for forecasting.

A good parameter estimation is important for financial processes. It gives better predictions for future market behaviour. A sharp estimation procedure may also clear up fundamental questions around the stationarity of certain financial processes. Do parameters change over time? Is this change slow or abrupt? We hope that the results in this chapter will help to obtain answers to these questions in the future.
Appendices

3.A Data

Our data set is the U.S. Standard & Poor’s 500 stock index future, traded at the Chicago Mercantile Exchange (CME), for the period 1st of January, 1988 until May 31st, 2006. The data were obtained from Nexa Technologies Inc. (www.tickdata.com). The futures trade from 8:30 A.M. until 15:15 P.M. Central Standard Time. Each record in the set contains a timestamp (with one second precision) and a transaction price. The tick size is $0.05 for the first part of the data and $0.10 from 1997–11–01. The data set consists of 4655 trading days. We remove sixty four days for which the closing hour is 12:15 P.M. (early closing hours occur on days before a holiday). Sixteen more days are removed, either because of too late first ticks, too early last ticks, or a suspiciously long intraday no-tick period. These removals leave us with a data set of 4575 days with nearly 14 million price ticks, on average more than 3 thousand price ticks per day, or 7.5 price ticks per minute.

There are four expiration months: March, June, September, and December. We use the most actively-traded contract: we roll to a next expiration as soon as the tick volume for the next expiration is larger than for the current expiration.

Figure 3.3 gives an impression of the course of volatility over the years 1988–2006. It gives an impression of the cumulative squared volatility. The shape of the graphs is the same, but the total increase in the figure based on realized variance is smaller since it does not take into account the overnight return. The growth of cumulative volatility is low in certain periods and high in other periods. The years 1992–1995 form a period without clear qualitative changes in the slope. The empirical analysis in Section 3.3 is based on these four years.

3.B Quasi Maximum Likelihood

This section contains background for the QML theory presented in Section 3.2. Sections 3.B.1 and 3.B.2 discuss QML estimation and the regularity conditions. Section 3.B.3 briefly discusses the special case of the standard GARCH(1,1) QMLE.

3.B.1 Principle of QML

The estimation method used in this chapter is quasi maximum likelihood (QML). Let us briefly describe the principle of Gaussian quasi maximum likelihood estimation, as discussed in Bollerslev and Wooldridge (1992). The observations \((y_n)\) form a stationary
sequence adapted to the filtration \( (\mathcal{F}_n) \). The conditional mean and variance functions \( \mu_n(\theta), h_n(\theta) \) are parameterized by a finite dimensional parameter \( \theta \) and there is a true value \( \theta^0 \in \Theta \) in the sense that

\[
\mu_n(\theta^0) = \mathbb{E}(y_n|\mathcal{F}_{n-1}), \quad h_n(\theta^0) = \text{var}(y_n|\mathcal{F}_{n-1}),
\]

for all \( n \). The likelihood of the sample \( (y_1, \ldots, y_N) \) is a function of \( \theta \). The parameter \( \theta \) may be estimated by maximizing the Gaussian likelihood, even if the true conditional probability distribution of \( y_n \) is not Gaussian. The likelihood is constructed then as if \( y_n \) is \( \mathcal{N}(\mu_n, h_n) \), and is called quasi-likelihood. The maximizer is still consistent and asymptotically normal, but with adjusted standard errors. The residual function \( \varepsilon_n(\theta) = \varepsilon_n(y_n, \theta) \) denotes the standardized \( y_n \),

\[
\varepsilon_n(\theta) = \frac{y_n - \mu_n(\theta)}{\sqrt{h_n(\theta)}}.
\]

This leads to a log-likelihood (precisely, a quasi-conditional log-likelihood)

\[
L_N(\theta) = \sum_{n=1}^{N} l_n(\theta),
\]

where, by the Gaussian likelihood,

\[
l_n(\theta) = -\frac{1}{2} \left[ \log(2\pi) + \log(h_n(\theta)) + \varepsilon_n(\theta)^2 \right].
\]
The QMLE \( \hat{\theta}_N \) denotes the maximizer of the log-likelihood. Under regularity (see Appendix 3.B.2) the QMLE is asymptotically normal,

\[
\sqrt{N}(\hat{\theta}_N - \theta^0) \overset{d}{\to} \mathcal{N}(0, V_0), \quad N \to \infty,
\]

where

\[
V_0 = A_0^{-1}B_0A_0^{-1}.
\]

The matrices \( A_0 \) and \( B_0 \) are given by the expected Hessian and the expectation of the outer product of the scores (which is the covariance matrix of the scores):

\[
(A_0)_{i,j} = -\mathbb{E} \left( \frac{\partial^2 l_0(\theta^0)}{\partial \theta_i \partial \theta_j} \right), \quad (B_0)_{i,j} = \mathbb{E} \left( s_{0,i}(\theta^0)s_{0,j}(\theta^0) \right),
\]

where, using stationarity, the expectation is taken at time \( n = 0 \). The scores \( s_{n,i}(\theta) \) are given by

\[
s_{n,i}(\theta) = \frac{\partial l_n(\theta)}{\partial \theta_i} = \frac{\varepsilon_n(\theta)}{\sqrt{h_n(\theta)}} \left( \frac{\partial \mu_n(\theta)}{\partial \theta_i} \right) + \frac{\varepsilon_n^2(\theta) - 1}{2h_n(\theta)} \left( \frac{\partial h_n(\theta)}{\partial \theta_i} \right).
\]

The expected Hessian \( A_0 \) may be expressed using first derivatives only,

\[
(A_0)_{i,j} = \mathbb{E} \left[ \frac{1}{l_{10}(\theta^0)} \left( \frac{\partial \mu_0(\theta^0)}{\partial \theta_i} \right) \left( \frac{\partial \mu_0(\theta^0)}{\partial \theta_j} \right) + \frac{1}{2h_{10}(\theta^0)} \left( \frac{\partial h_0(\theta^0)}{\partial \theta_i} \right) \left( \frac{\partial h_0(\theta^0)}{\partial \theta_j} \right) \right].
\]

If the true conditional probability distribution is Gaussian, the QMLE reduces to the Gaussian maximum likelihood estimator and the information matrix equality \( A_0 = B_0 \) holds, so \( V_0 \) reduces to \( A_0^{-1} \), and the QMLE is efficient.

### 3.B.2 QML Regularity Conditions

Bollerslev and Wooldridge (1992) provided abstract regularity conditions allowing for additional regressors \( (x_n) \), and without assuming stationarity for \( (y_n) \). We restate these conditions below, assuming stationarity, and leaving out \( x_n \). The scores \( s_n \) are row vectors. Let \( \tilde{l}_n \) denote the Hessian of \( l_n(\theta) \), so \( \tilde{l}_n = \dot{s}_n \). We first state the definition of the Uniform Weak Law of Large Numbers, as given by Wooldridge (1990, Definition A.1). A sequence of random functions \( q_n(y_n, \theta) \) satisfies the UWLLN if

\[
\sup_{\theta \in \Theta} \left| N^{-1} \sum_{n=1}^{N} q_n(y_n, \theta) - \mathbb{E} q_n(y_n, \theta) \right| \overset{P}{\to} 0, \quad N \to \infty.
\]
The QML regularity conditions are:

1. $\Theta$ is compact, has nonempty interior and $\theta^0 \in \operatorname{int} \Theta$,
2. The mean and variance functions $\mu_n, h_n$ are measurable functions of the data for all $\theta \in \Theta$, are twice continuously differentiable with respect to $\theta$ on $\operatorname{int} \Theta$, and the conditional variance is nonsingular (with probability one), for all $\theta \in \Theta$,
3. (a) $(l_n(\theta))$ satisfies the UWLLN,
   (b) $\theta^0$ is the identifiably unique maximizer (Bates and White, 1985) of $\theta \to \mathbb{E} l_n(\theta)$,
4. (a) The Hessians $(\tilde{l}_n(\theta))$ satisfy the UWLLN,
   (b) The expected Hessian $A_0 = \mathbb{E}\tilde{l}_n(\theta^0)$ is positive definite,
5. (a) The expected outer product $B_0 = \mathbb{E}s_n'(\theta^0)$ is positive definite,
   (b) $\frac{1}{\sqrt{N}}B_0^{-1/2}\sum s_n'(\theta^0) \xrightarrow{d} \mathcal{N}(0, I_p), \quad N \to \infty$,
6. The outer product of the scores $(s_n' s_n(\theta))$ satisfies the UWLLN.

### 3.B.3 QML Regularity Conditions for GARCH(1,1)

The verification of the conditions for asymptotic normality of the quasi maximum likelihood estimator given in Appendix 3.B.2, has to be carried out on a case-by-case basis. The GARCH(1,1) system (3.3–3.4) corresponds to $y_n = r_n, \mu_n(\theta) = 0, h_n(\theta) = v_n^2(\gamma, \beta) \tau^2$, with $\theta = (\tau, \gamma, \beta)$. In the case of a GARCH type process a problem is that one cannot evaluate the exact likelihood for a given parameter $\theta$, since the unobservable variables $v_n(\theta)$ have to be replaced by approximations $\tilde{v}_n(\theta)$. The unobserved $v_n$ is approximated by the GARCH recursion, with initialization $\tilde{v}_0^2 > 0$. There are several papers on the Gaussian QMLE for GARCH(1,1) including Lee and Hansen (1994), Lumsdaine (1996), Berkes, Horvath, and Kokoszka (2003), and Francq and Zakoïan (2004). The QMLE $\hat{\theta}_N$ satisfies the asymptotic normality of equation (3.32) and one may consistently estimate the covariance matrix $V_0$ by using the empirical counterparts of $A_0$ and $B_0$; we refer to Straumann and Mikosch (2006) for the following regularity conditions, see also the monograph of Straumann (2005). The observations $y_1, \ldots, y_N$ are part of a stationary sequence $(y_n)$ that satisfies (cf. (3.3–3.4))

\begin{align}
y_n &= v_n \tau Z_n \\
v_n^2 &= 1 + \gamma \tau^2 v_{n-1}^2 Z_{n-1}^2 + \beta v_{n-1}^2,
\end{align}

where

1. $(Z_n)$ is an iid sequence with $\mathbb{E}Z^2 = 1$,.
3.C Proofs

Proof of Theorem 3.1. The proof of Theorem 3.1 in this chapter applies the extensive likelihood theory of Straumann and Mikosch (2006). The theory in Straumann and Mikosch treats the Gaussian QMLE based on close-to-close returns \( r_n \) for general GARCH-type models. The proof below is a direct adjustment of their proof; we shall provide a detailed description of the adjustments that we make.

The asymptotic normality of the usual GARCH(1,1) QMLE follows from Theorem 8.1 of Straumann and Mikosch (2006). The proof of that theorem relies on their more general Theorem 7.1. We extend the assumptions needed to invoke Theorem 8.1 in Straumann and Mikosch, check that this set of assumptions establishes asymptotic normality of the Gaussian QMLE in this chapter and then remove the redundant elements of the assumptions. We collected the conditions for the usual Gaussian QMLE based on close-to-close returns as conditions 3.B.3(1) to 3.B.3(5) in our Appendix 3.B.3. First extend these assumptions by duplication: copy the conditions for \( \tau \) and \( Z \) to \( \tau_H \) and \( Z_H \) : assume \( \tau_H > 0 \) and add to each condition for \( Z \) the same condition for \( Z_H \). We now have a set of (temporary) conditions (D1) to (D5), concerning both \( Z \) and \( Z_H \):
(D1) \((Z_n)\) is an iid sequence with \(\mathbb{E}Z^2 = 1\), \((Z_{H,n})\) is an iid sequence with \(\mathbb{E}Z^2_{H,n} = 1\), (D2) \(\tau_H > 0\), \(\tau > 0\), \(\gamma > 0\), \(\beta \in [0,1)\), (D3) \(\mathbb{E} \log (\gamma^0(\tau^0)^2Z^2 + \beta^0) < 0\), and \(\mathbb{E} \log (\gamma^0(\tau_H^0)^2Z^2_{H,n} + \beta^0) < 0\), (D4) \(Z^2, Z^2_{H,n}\) are non-degenerate, (D5) \(\mathbb{E}Z^4 < \infty\), \(\mathbb{E}Z^4_{H} < \infty\).

Under conditions (D1) to (D4) the usual GARCH model satisfies the consistency conditions (C1) to (C4) of Straumann and Mikosch, pp. 2473 (for a verification, see their Section 5.2). Let us first verify that the Gaussian QMLE in this chapter is consistent, in analogy to Theorem 4.1 in Straumann and Mikosch. Let \(L_{H,N}(\theta) = \sum_{n=1}^{N} l_{H,n}(\theta)\) denote the log-likelihood (modulo an additive constant), where

\[
l_{H,n}(\theta) = -\frac{1}{2} \left( \log(h_n(\theta)) + H^2_n/h_n(\theta) \right)
\]

and \(h_n(\theta) = v^2_n(\gamma, \beta)\tau_H^2\). Recall that the classical GARCH(1,1) QMLE corresponds to \(H_n = |r_n|\), and \(h_n(\theta) = v^2_n(\gamma, \beta)\tau^2\). It is important to note that the innovation \(Z_{H,n}\) is independent of \(h_n(\theta)\) and \(v_n\) and satisfies \(\mathbb{E}Z^2_{H,n} = 1\). The function \(L(\theta) = \mathbb{E} l_{H,0}(\theta)\) equals

\[
L(\theta) = -\frac{1}{2} \mathbb{E} \left( \log(h_0(\theta)) + \frac{v^2_0(\gamma, \beta)(\tau_H^0)^2}{h_0(\theta)} \right).
\]

One has \(2(L(\theta) - L(\theta_0)) - 1 = \mathbb{E}(\log(h_0(\theta^0)/h_0(\theta)) - h_0(\theta^0)/h_0(\theta))\). Since \(\log(x) - x \leq -1\) with equality if and only if \(x = 1\), \(L(\theta)\) is maximized if and only if \(h_0(\theta) = h_0(\theta^0)\) if and only if \(\theta = \theta^0\) (the latter equivalence follows from Lemma 5.4, Straumann and Mikosch). One may now follow the proof of Theorem 4.1 of Straumann and Mikosch, pp. 2473 part i, to obtain that \(L_{H,N}/N\) converges to \(L\) uniformly. The rest of the proof of Theorem 4.1 needs no adjustment and shows that the QMLE converges almost surely to \((\tau_H^0, \gamma^0, \beta^0)\).

Straumann and Mikosch, Section 7, treat the asymptotic normality of their general QMLE under their assumptions \((N1)\) to \((N4)\), see Theorem 7.1. One may follow their exposition, replacing \(X\) by \(H\), until the second display on pp. 2488, for which we may write

\[
\hat{L}_{H,n}(\theta^0) = \sum_{n=1}^{N} \hat{l}_{H,n}(\theta^0) = \frac{1}{2} \sum_{n=1}^{N} \frac{\hat{h}_n(\theta^0)}{h_n(\theta^0)} (Z^2_{H,n} - 1),
\]

where \(\hat{l}_{H,n}(\theta^0)\) is a martingale difference sequence since \(Z_{H,n}\) is independent of \(\mathcal{F}_{n-1}\) and \(\mathbb{E}Z^2_{H,n} = 1\). Accordingly, as in Straumann and Mikosch, one may apply the central limit theorem for martingale differences, assuming \(\mathbb{E}Z^4_{H} < \infty\). So, an application of Theorem 7.1 to the Gaussian QMLE in this chapter needs \(\mathbb{E}Z^4_{H} < \infty\), which is satisfied by (D5).

Under conditions 3.B.3(1) to 3.B.3(5) from Appendix 3.B.3, the standard GARCH
model satisfies condition (N1) to (N4) of Straumann and Mikosch, see also their Theorem 8.1. For the Gaussian QMLE of this chapter we have to establish (N1) to (N4) under our duplicated conditions (D1) to (D5). The only conditions that is left for reexamination is condition N3.iv: \( \mathbb{E}[\|l_0\|_\Theta^2 < \infty \) (N3.iii is never used, see also Straumann (2005)). Since \( \hat{\theta}_N \) converges almost surely to \( \theta^0 \), it suffices to verify \( \mathbb{E}[\|l_0\|_{\mathcal{V}(\theta_0)} < \infty \) where \( \mathcal{V}(\theta_0) \) denotes a neighbourhood of \( \theta^0 \). Let us follow the lines of Section 8 of Straumann and Mikosch. We may write

\[
\mathbb{E}[H_0^2/h_0|\mathcal{V}(\theta_0)] = \mathbb{E}[h_0(\theta^0)/h_0(\theta)|\mathcal{V}(\theta_0)] E Z_{H,0}^{\nu^2}.
\]

From Francq and Zakoïan (2004) one obtains \( \mathbb{E}[h_0(\theta^0)/h_0(\theta)|\mathcal{V}(\theta_0)] < \infty \) (see the first display below their equation (4.25), pp. 622). Therefore

\[
\mathbb{E}[H_0^2/h_0|\mathcal{V}(\theta_0)] < \infty, \quad 0 \leq \nu \leq 1.
\]

One may now follow the arguments of Straumann and Mikosch to establish their condition N.3.iv. This establishes the asymptotic normality of the Gaussian QMLE of Theorem 3.1 in this chapter.

It remains to remove the redundant elements from conditions (D1) to (D5), and establish conditions (A1) to (A5) from Section 3.2.1. The assumption \( EZ_H^2 < \infty \) and equation (3.11) already imply that \( (Z_{H,n}) \) is an iid sequence with \( EZ_H^2 = 1 \), yielding (A1). One should read condition (2) from Appendix 3.B.3 as a description of the parameter space. This does not need \( \tau > 0 \), since we optimize \( L_H \) over \( \tau_H \), not \( \tau \). Furthermore \( \tau > 0 \) is equivalent to \( \tau_H > 0 \) by \( \tau_H = \mu^H \tau \), hence (A2). Condition (D3) was used for establishing stationarity, ergodicity, and invertibility of \( (v_n) \). These properties do not rely on the innovations \( Z_{H,n} \), yielding (A3). Condition (D4) was used for establishing that \( v_n \) is uniquely determined by \( \theta \), again a property that does not depend on \( Z_H \), hence (A4). Condition (D5) was used for establishing asymptotic normality. Condition (D5) is needed to obtain a finite variance in the application of the martingale difference central limit theorem to the derivative of \( L_{H,N} \), which only requires \( EZ_H^4 < \infty \), and not \( EZ^4 < \infty \), see the arguments above.

**Proof of Lemma 3.2.** Differentiation yields \( \frac{\partial \sigma_{H,0}(\theta)}{\partial H} = 2\tau_H v_0^2(\theta), \frac{\partial \sigma_{H,0}(\theta)}{\partial \gamma} = \tau_H \frac{\partial \sigma_{H,0}(\theta)}{\partial \gamma}, \) and 

\[
G_H(\theta) = \mathbb{E}
\begin{pmatrix}
\frac{4}{\tau_H} & 2 \frac{\partial \sigma_{H,0}^2(\theta)}{\partial H} & 2 \frac{\partial \sigma_{H,0}^2(\theta)}{\partial \gamma}
\end{pmatrix}
\begin{pmatrix}
\frac{1}{v_0} & \frac{\partial \sigma_{H,0}^2(\theta)}{\partial \gamma} & \frac{\partial \sigma_{H,0}^2(\theta)}{\partial \gamma}
\end{pmatrix}
\begin{pmatrix}
\frac{1}{v_0} & \frac{\partial \sigma_{H,0}^2(\theta)}{\partial \gamma} & \frac{\partial \sigma_{H,0}^2(\theta)}{\partial \gamma}
\end{pmatrix}
\theta.
\]

(3.36)
The lower right block of the inverse of a matrix

\[ A = \begin{pmatrix} A_{11} & A_{12} \\ A_{21} & A_{22} \end{pmatrix}, \]

equals \( C^{-1} = (A_{22} - A_{21}A_{11}^{-1}A_{12})^{-1} \). So the \((\gamma, \beta)\) block of \( G^{-1} \) equals the inverse of the \(2 \times 2\) matrix given by

\[
(G^{-1})_{\gamma, \beta} = \left( \begin{array}{cc}
\frac{1}{\tau^2} \left( \frac{\partial^2}{\partial \gamma^2} \right)^2 & \frac{1}{\tau^2} \left( \frac{\partial^2}{\partial \gamma \partial \beta} \right)^2 \\
\frac{1}{\tau^2} \left( \frac{\partial^2}{\partial \beta^2} \right)^2 & \frac{1}{\tau^2} \left( \frac{\partial^2}{\partial \gamma \partial \beta} \right)^2
\end{array} \right)^{-1}.
\]

\[ = \left( \begin{array}{cc}
\text{var}(\frac{1}{\tau^2} \frac{\partial^2}{\partial \gamma^2}) & \text{cov}(\frac{1}{\tau^2} \frac{\partial^2}{\partial \gamma \partial \beta}, \frac{1}{\tau^2} \frac{\partial^2}{\partial \beta^2}) \\
\text{cov}(\frac{1}{\tau^2} \frac{\partial^2}{\partial \gamma \partial \beta}, \frac{1}{\tau^2} \frac{\partial^2}{\partial \beta^2}) & \text{var}(\frac{1}{\tau^2} \frac{\partial^2}{\partial \beta^2})
\end{array} \right)^{-1}.
\]

(3.37)

Formula (3.37) does not depend on \( H \). ■

**On the relative error** \( \text{re}(\sigma_{H,n}) \) in Section 3.2.4. Let \( h_n(\theta) = \sigma_{H,n}^2(\theta) = v_n^2 \tau_H^2 \). The derivative of \( h_n \) is given by

\[
\hat{h}_n = \left( 2 \tau_H v_n^2 \tau_H^2 (v_n^2 - 1 + \beta \frac{\partial^2}{\partial \gamma^2} + \frac{\partial^2}{\partial \gamma \partial \beta} \frac{\partial^2}{\partial \beta^2}) \right).
\]

The Gaussian QMLE \( \hat{\theta}_N \) has asymptotic variance \( V_0 = \text{var}(Z_H^2)G^{-1}_H(\theta^0) \). The asymptotic variance \((N \to \infty)\) of \( h_n(\hat{\theta}_N) \) is given by \( V_{h_n} = \hat{h}_n(\theta^0)V_0 \hat{h}_n(\theta^0)' \). Partition the matrix \( G \) in (3.36) into \( \tau_H \) and \((\gamma, \beta)\) blocks. Using partitioned inverses one finds that the asymptotic variance of \( h_n \) (for fixed \( n \)) equals

\[ V_{h_n} = c \text{var}(Z_H^2)\tau_H^4, \]

where \( c \) is a constant that does not depend on \( H \). The asymptotic variance of \( \sigma_{H,n} = \sqrt{h_n} \) may be obtained by the Delta method using formula (3.25):

\[ V_{\sigma_{H,n}} = \frac{1}{4v_n^2(\theta^0)\tau_H^4} V_{h_n}. \]

(3.38)

One sees that the parameter \( \tau_H \) drops out. ■
3.D Realized Variance of Ornstein-Uhlenbeck Log-Volatility

Consider the intraday process $\Psi(\cdot)$ from Section 3.4. The accompanying process $Y(\cdot)$ satisfies

$$Y(u) = \exp(-\delta u)Y(0) + \mu(1 - \exp(-\delta u)) + \sigma_{Y} \int_{s=0}^{u} \exp(-\delta(u - s)) \, dB^{(2)}(s).$$

Simulation of the process $Y$ is straightforward since $Y(u + \Delta)|Y(u)$ has a normal distribution with mean

$$\exp(-\delta \Delta)Y(u) + (1 - \exp(-\delta \Delta))\mu,$$

and variance

$$\sigma_{Y}^{2} \frac{1}{2\delta}(1 - \exp(-2\delta \Delta)),$$

see for instance Glasserman (2003). The process $Y(\cdot)$ has a stationary version which is normally distributed with mean $\mu$ and variance $\sigma_{Y}^{2}/(2\delta)$. We sample $Y(0)$ from the stationary distribution: this yields a simple expression for the expectation of the realized quadratic variation using step size $\Delta$. We shall use that the expectation of the squared increment in $\Psi(\cdot)$ equals the expectation of the increment in the quadratic variation. The expected increment in $QV$ equals

$$\mathbb{E}QV[u, u + \Delta] = \int_{s=u}^{u+\Delta} \mathbb{E} \exp(2Y(s))ds$$

$$= \int_{s=u}^{u+\Delta} \mathbb{E} \exp(2Y(0))ds$$

$$= \exp(2\mu + 2\frac{\sigma_{Y}^{2}}{2\delta}) \Delta.$$ 

So, for $\mu = -\sigma_{Y}^{2}/(2\delta)$, the quadratic variation over the unit interval, $\Delta = 1$, has expectation 1. This implies that the realized variance $RQV^{(m)}$ has expectation 1 for all $m$. We simulate a realized variance based on $m = 81$ intervals. Each of those intervals is divided into 10 subintervals using equally spaced grid points. The simulation of the process $Y$ on all grid points is exact. The value of $\Psi(\cdot)$ on each grid point is obtained by Euler discretization.