Econometric Theory, **28**, 2012, 1037–1064.
doi:10.1017/S0266466612000047

RANK-BASED ESTIMATION FOR GARCH PROCESSES

BETH ANDREWS
Northwestern University

We consider a rank-based technique for estimating generalized autoregressive conditionally heteroskedastic (GARCH) model parameters, some of which are scale transformations of conventional GARCH parameters. The estimators are obtained by minimizing a rank-based residual dispersion function similar to the one given in Jaeckel (1972, *Annals of Mathematical Statistics* 43, 1449–1458). They are useful for GARCH order selection and preliminary estimation. We give a limiting distribution for the rank estimators that holds when the true parameter vector is in the interior of its parameter space and when some GARCH parameters are zero. The limiting theory is used to show that the rank estimators are robust, can have the same asymptotic efficiency as maximum likelihood estimators, and are relatively efficient compared to traditional Gaussian and Laplace quasi-maximum likelihood estimators. The behavior of the estimators for finite samples is studied via simulation, and we use rank estimation to fit a GARCH model to exchange rate log-returns.

1. INTRODUCTION

Observed time series processes frequently appear uncorrelated, yet exhibit volatility clustering. Volatility clustering is the tendency of observations relatively small in absolute value to be followed by other small observations, and the tendency of observations relatively large in absolute value to be followed by other large observations. Hence, these series appear uncorrelated but dependent. Nonlinear models with time-dependent conditional variances, most notably generalized autoregressive conditionally heteroskedastic (GARCH) models, are often used to describe time series with these features. GARCH models were first developed for modeling inflation rates (Engle, 1982; Bollerslev, 1986) and have also appeared for analyzing the returns of exchange rates (Bollerslev, 1987; Engle and González-Rivera, 1991; Shephard, 1996) and stock prices (Bollerslev, 1987; Shephard, 1996; Fan and Yao, 2003). Applications for GARCH models are not limited to finance, however. Time series processes exhibiting GARCH-type behavior have also appeared, for example, in speech signals (Abramson and Cohen, 2008), daily and monthly mean temperatures (Campbell and Diebold, 2005; Romilly, 2005; Huang, Shiu, and Lin, 2008), wind speeds (Ewing, Kruse,

The author is very grateful to editor Peter C.B. Phillips, co-editor Pentti Saikkonen, and two anonymous referees for their helpful comments. This work was supported in part by NSF grant DMS0806104. Address correspondence to Beth Andrews, Department of Statistics, Northwestern University, Evanston, IL 60208, USA; e-mail: bandrews@northwestern.edu.

© Cambridge University Press 2012

https://doi.org/10.1017/S0266466612000047 Published online by Cambridge University Press

and Thompson, 2008), and atmospheric carbon dioxide concentrations (Hoti, McAleer, and Chan, 2005).

In this paper, we consider a rank-based technique for estimating GARCH model parameters, some of which are scale transformations of conventional GARCH parameters. The rank (R) estimators are obtained by minimizing the sum of mean-corrected model residuals weighted by a function of residual rank. They are similar to the R-estimators proposed by Jaeckel (1972) for estimating linear regression parameters and can be used for GARCH order selection and preliminary estimation. As discussed in Jurečková and Sen (1996), R-estimators are, in general, robust and relatively efficient, and results in this paper indicate that this is true in the case of GARCH estimation. The technique is robust because the R-estimators are $n^{1/2}$-consistent (n represents sample size) under general conditions, and, because the weight function can be chosen so that R-estimators have the same asymptotic efficiency as maximum likelihood (ML) estimators, it is also relatively efficient. In addition, R-estimation dominates traditional techniques such as Gaussian and Laplace quasi-ML (QML) with respect to asymptotic efficiency. We therefore recommend that R-estimation be used for preliminary GARCH estimation and order selection when the noise distribution is unknown. Once a R-estimate has been found, corresponding model residuals can be used to identify an appropriate noise distribution, and then the conventional GARCH parameters can be estimated via ML.

Another rank-based technique for estimating the parameters of conditionally heteroskedastic processes is given in Mukherjee (2007). However, the class of models considered in Mukherjee (2007) includes autoregressive conditionally heteroskedastic (ARCH) but not GARCH models, and $n^{1/2}$-consistency for the ARCH parameter estimates is established only when the noise distribution has a finite fourth moment. Although it is traditional to assume the noise distribution is Gaussian when fitting an ARCH/GARCH model to an observed series, many series appear to have noise distributions that are heavier tailed than Gaussian (Bollerslev, 1987; Engle and González-Rivera, 1991; Shephard, 1996; Fan and Yao, 2003). Given the many applications for GARCH models, it is therefore important that robust statistical theory be developed. In this paper, we consider both ARCH and GARCH models and show that, when the true parameter vector lies in the interior of the parameter space, higher order moment conditions are not required for the $n^{1/2}$-consistency of R-estimators of ARCH model parameters.

The R-estimators are introduced in Section 2, and in Section 3.1 we show that as sample size $n \to \infty$, they converge in distribution to the minimizer of a random quadratic function on a convex space. This limiting result holds when the parameter vector lies in the interior of its parameter space, in which case it follows that the estimators are asymptotically normal, and also when some GARCH parameters are zero, and hence the parameter vector lies on the boundary of its parameter space. In Section 3.2 we show that the limiting distribution for the R-estimators can be used for GARCH order selection and confidence interval estimation. Proofs of the lemmas used to establish the results of Section 3 are in

https://doi.org/10.1017/S0266466612000047 Published online by Cambridge University Press

the Appendix. The quality of the asymptotic approximations for finite samples is studied via simulation in Section 4.1, and we use R-estimation to fit a GARCH model to exchange rate log-returns in Section 4.2.

2. PRELIMINARIES

A series $\{X_t\}_{t=-\infty}^{\infty}$ is a GARCH(p, q) process if

$$X_t = \sigma_t Z_t, \tag{2.1}$$

where $\{Z_t\}$ is a sequence of independent and identically distributed (i.i.d.) random variables with $\mathbb{E}\{Z_t\} = 0$ and $\mathrm{Var}\{Z_t\} = 1$, $\{\sigma_t\}$ is a nonnegative process satisfying

$$\sigma_t^2 = \alpha_{00} + \sum_{i=1}^{p} \alpha_{0i} X_{t-i}^2 + \sum_{j=1}^{q} \beta_{0j} \sigma_{t-j}^2, \tag{2.2}$$

and Z_t is independent of $\{X_{t-k}, k \geq 1\}$ for every t (Bollerslev, 1986). The parameter α_{00} is positive, $\alpha_{0i}, i \in \{1, \ldots, p\}$, and $\beta_{0j}, j \in \{1, \ldots, q\}$, are nonnegative, and σ_t^2 represents $\mathrm{Var}\{X_t | X_s, s < t\}$. When $q = 0$, $\{X_t\}$ is an ARCH(p) process (Engle, 1982). In the case of a GARCH(1, 1) process, $\mathbb{E}\{\ln(\alpha_{01} Z_t^2 + \beta_{01})\} < 0$ is necessary and sufficient for the stationarity and ergodicity of $\{X_t\}$ (Nelson, 1990) and, for $p \geq 2$ and $q \geq 2$, a stationary, ergodic solution to (2.1)–(2.2) exists if and only if the top Lyapunov exponent associated with the $(p+q-1) \times (p+q-1)$ matrices $\{\mathbf{A}_{0t}\}$, where

$$\mathbf{A}_{0t} := \begin{bmatrix} \alpha_{01} Z_t^2 + \beta_{01} & \beta_{02} & \cdots & \beta_{0,q-1} & \beta_{0q} & \alpha_{02} & \cdots & \alpha_{0,p-1} & \alpha_{0p} \\ 1 & 0 & \cdots & 0 & 0 & 0 & \cdots & 0 & 0 \\ 0 & 1 & \cdots & 0 & 0 & 0 & \cdots & 0 & 0 \\ \vdots & \vdots & \ddots & \vdots & \vdots & \vdots & \ddots & \vdots & \vdots \\ 0 & 0 & \cdots & 1 & 0 & 0 & \cdots & 0 & 0 \\ Z_t^2 & 0 & \cdots & 0 & 0 & 0 & \cdots & 0 & 0 \\ 0 & 0 & \cdots & 0 & 0 & 1 & \cdots & 0 & 0 \\ \vdots & \vdots & \ddots & \vdots & \vdots & \vdots & \ddots & \vdots & \vdots \\ 0 & 0 & \cdots & 0 & 0 & 0 & \cdots & 1 & 0 \end{bmatrix},$$

is negative (Bougerol and Picard, 1992). That is, $\{X_t\}$ is stationary and ergodic if and only if $\gamma_L := \inf_{1 \leq t < \infty} [(t+1)^{-1} \mathbb{E}\{\ln \|\mathbf{A}_{00} \ldots \mathbf{A}_{0t}\|_{op}\}] < 0$, where, for a matrix $\mathbf{A}$, $\|\mathbf{A}\|_{op}$ represents the matrix operator norm $\sup_{\mathbf{x} \neq \mathbf{0}} (\|\mathbf{A}\mathbf{x}\| / \|\mathbf{x}\|)$ and $\|\cdot\|$ is the euclidean norm. Note that γ_L can also be used to assess the stationarity/ergodicity of a GARCH process for which $p < 2$ or $q < 2$ by setting $\alpha_{0i} = 0$ for $p < i \leq 2$ and $\beta_{0j} = 0$ for $q < j \leq 2$. Additionally, if

https://doi.org/10.1017/S0266466612000047 Published online by Cambridge University Press

$\sum_{i=1}^{p} \alpha_{0i} + \sum_{j=1}^{q} \beta_{0j} < 1$, it can be shown that the process $\{X_t\}$ is not only stationary and ergodic but also has finite variance $\alpha_{00}/(1 - \sum_{i=1}^{p} \alpha_{0i} - \sum_{j=1}^{q} \beta_{0j})$ (Bollerslev, 1986). We make the following assumption throughout.

A1. $\gamma_L < 0$.

Therefore $\{X_t\}$ is stationary and ergodic, but we do not assume that the variance of $\{X_t\}$ is necessarily finite. Following Straumann (2005, p. 76), so that the GARCH parameter values $\alpha_{00}, \alpha_{01}, \ldots, \alpha_{0p}, \beta_{01}, \ldots, \beta_{0q}$ are unique, we also make the following assumption.

A2. There exists at least one $i > 0$ for which $\alpha_{0i} > 0$, $\alpha_{0p} + \beta_{0q} \neq 0$, and the polynomials $\alpha_0(x) := \sum_{i=1}^{p} \alpha_{0i} x^i$ and $\beta_0(x) := 1 - \sum_{j=1}^{q} \beta_{0j} x^j$ have no common roots.

If we assume

A3. $P(Z_t \neq 0) = 1$,

then it follows from (2.1)–(2.2) that $P(\cap_{t=-\infty}^{\infty} \{X_t^2 > 0\}) = 1$. Therefore, we have $\ln(X_t^2) = \ln(\sigma_t^2/\alpha_{00}) + \ln(\alpha_{00} Z_t^2)$ or $\ln(\alpha_{00} Z_t^2) = \ln(X_t^2) - \ln(\sigma_t^2/\alpha_{00})$, with $\sigma_t^2/\alpha_{00} = 1 + \sum_{i=1}^{p} (\alpha_{0i}/\alpha_{00}) X_{t-i}^2 + \sum_{j=1}^{q} \beta_{0j} (\sigma_{t-j}^2/\alpha_{00})$, and for arbitrary values $\alpha_0, \alpha_1, \ldots, \alpha_p, \beta_1, \ldots, \beta_q$, we define

$$\epsilon_t(\boldsymbol{\theta}) = \ln\left(X_t^2\right) - \ln\left(\tilde{\sigma}_t^2(\boldsymbol{\theta})\right), \qquad t = 1, \ldots, n, \tag{2.3}$$

with

$$\tilde{\sigma}_t^2(\boldsymbol{\theta}) := \begin{cases} 1, & t = \min\{1,\ p - q + 1\}, \ldots, p, \\ 1 + \sum_{i=1}^{p} \theta_i X_{t-i}^2 + \sum_{j=1}^{q} \theta_{p+j} \tilde{\sigma}_{t-j}^2(\boldsymbol{\theta}), & t = p + 1, \ldots, n, \end{cases} \tag{2.4}$$

and $\boldsymbol{\theta} := (\alpha_1/\alpha_0, \ldots, \alpha_p/\alpha_0, \beta_1, \ldots, \beta_q)'$. Let $\boldsymbol{\theta}_0 = (\alpha_{01}/\alpha_{00}, \ldots, \alpha_{0p}/\alpha_{00}, \beta_{01}, \ldots, \beta_{0q})'$ and $\epsilon_t = \ln(\alpha_{00} Z_t^2)$ and note that $\{\epsilon_t(\boldsymbol{\theta}_0)\}_{t=p+1}^{n}$ closely approximates $\{\epsilon_t\}_{t=p+1}^{n}$; the error is due to the initialization with ones in (2.4). Because A1 implies $\sum_{j=1}^{q} \beta_{0j} < 1$ (Bougerol and Picard, 1992, Cor. 2.3), $\boldsymbol{\theta}_0$ is in the parameter space $\boldsymbol{\Theta} := [0, \infty)^p \times \{(x_1, \ldots, x_q)' \in [0, 1)^q : x_1 + \cdots + x_q < 1\}$. Now suppose that the following assumption holds.

A4. λ is a nonconstant and nondecreasing function from $(0, 1)$ to $\mathbb{R}$.

For $\boldsymbol{\theta} \in \boldsymbol{\Theta}$, we introduce the R-function

$$D_n(\boldsymbol{\theta}) = \sum_{t=p+1}^{n} \lambda\left(\frac{R_t(\boldsymbol{\theta})}{n - p + 1}\right) \left[\epsilon_t(\boldsymbol{\theta}) - \overline{\epsilon(\boldsymbol{\theta})}\right], \tag{2.5}$$

https://doi.org/10.1017/S0266466612000047 Published online by Cambridge University Press

where $\{R_t(\boldsymbol{\theta})\}_{t=p+1}^n$ contains the ranks of the residuals $\{\epsilon_t(\boldsymbol{\theta})\}_{t=p+1}^n$ and $\overline{\epsilon(\boldsymbol{\theta})} := (n-p)^{-1}\sum_{t=p+1}^n \epsilon_t(\boldsymbol{\theta})$. The function D_n is similar to the R-function introduced in Jaeckel (1972) for estimating linear regression parameters. We, however, consider a weighted sum of the mean-corrected $\{\epsilon_t(\boldsymbol{\theta})\}_{t=p+1}^n$, instead of a weighted sum of the residuals (as in Jaeckel, 1972), to avoid assuming $\sum_{t=p+1}^n \lambda((t-p)/(n-p+1)) = 0$, which is required in Jaeckel (1972). Note that if $\{\epsilon_{(t)}(\boldsymbol{\theta})\}_{t=p+1}^n$ is the series $\{\epsilon_t(\boldsymbol{\theta})\}_{t=p+1}^n$ ordered from smallest to largest, (2.5) can also be written as $D_n(\boldsymbol{\theta}) = \sum_{t=p+1}^n \lambda((t-p)/(n-p+1))[\epsilon_{(t)}(\boldsymbol{\theta}) - \overline{\epsilon(\boldsymbol{\theta})}]$ and, if $\overline{\lambda} := (n-p)^{-1}\sum_{t=p+1}^n \lambda((t-p)/(n-p+1))$, $D_n(\boldsymbol{\theta}) = \sum_{t=p+1}^n [\lambda((t-p)/(n-p+1)) - \overline{\lambda}][\epsilon_{(t)}(\boldsymbol{\theta}) - \overline{\epsilon(\boldsymbol{\theta})}]$. Because it tends to be near zero when the elements of $\{\epsilon_t(\boldsymbol{\theta})\}_{t=p+1}^n$ are similar and gets larger as the values of $\{|\epsilon_{(t)}(\boldsymbol{\theta}) - \overline{\epsilon(\boldsymbol{\theta})}|\}_{t=p+1}^n$ increase, D_n is a measure of the dispersion of the residuals $\{\epsilon_t(\boldsymbol{\theta})\}_{t=p+1}^n$. Given a realization of length n from (2.1), $\{X_t\}_{t=1}^n$, we plan to estimate $\boldsymbol{\theta}_0$ by minimizing D_n. Our motivation for using the residuals $\{\epsilon_t(\boldsymbol{\theta})\}_{t=p+1}^n$ is that, given an appropriately chosen loss function, M-estimation with respect to $\{\epsilon_t(\boldsymbol{\theta})\}_{t=p+1}^n$ can be equivalent to ML or QML estimation for GARCH model parameters (Muler and Yohai, 2008). Additionally, it is not possible to estimate $\boldsymbol{\theta}_0$ by minimizing the dispersion of the residuals $\{X_t/\sqrt{\tilde{\sigma}_t^2(\boldsymbol{\theta})}\}_{t=p+1}^n$, because the values of $\{X_t/\sqrt{\tilde{\sigma}_t^2(\boldsymbol{\theta})}\}_{t=p+1}^n$ become more clustered about zero as the elements of $\boldsymbol{\theta}$ increase. By Theorem 2.1, D_n is a nonnegative, continuous function on $\boldsymbol{\Theta}$. Choices for the weight function λ are discussed in Section 3.1.

THEOREM 2.1. *Assume A1–A4 hold. If, for* $\boldsymbol{\theta} \in \boldsymbol{\Theta}$,

$$\{P_1(\boldsymbol{\theta}), \ldots, P_{(n-p)!}(\boldsymbol{\theta})\} = \{\{\epsilon_{1,p+1}(\boldsymbol{\theta}), \ldots, \epsilon_{1,n}(\boldsymbol{\theta})\}, \ldots, \{\epsilon_{(n-p)!,p+1}(\boldsymbol{\theta}), \ldots, \epsilon_{(n-p)!,n}(\boldsymbol{\theta})\}\}$$

contains the $(n-p)!$ *permutations of the sequence* $\{\epsilon_t(\boldsymbol{\theta})\}_{t=p+1}^n$ *(so, for* $j \in \{1, \ldots, (n-p)!\}$, $t \in \{p+1, \ldots, n\}$, $\epsilon_{j,t}(\boldsymbol{\theta})$ *represents the* $(t-p)$*th element of permutation* $P_j(\boldsymbol{\theta})$*), then*

$$D_n(\boldsymbol{\theta}) = \sup_{j \in \{1,\ldots,(n-p)!\}} \sum_{t=p+1}^n \lambda\left(\frac{t-p}{n-p+1}\right)\left[\epsilon_{j,t}(\boldsymbol{\theta}) - \overline{\epsilon(\boldsymbol{\theta})}\right].$$

In addition, D_n *is a nonnegative, continuous function on* $\boldsymbol{\Theta}$.

Proof. Recall that $D_n(\boldsymbol{\theta}) = \sum_{t=p+1}^n [\lambda((t-p)/(n-p+1)) - \overline{\lambda}][\epsilon_{(t)}(\boldsymbol{\theta}) - \overline{\epsilon(\boldsymbol{\theta})}]$ and let $a_n(t) = \lambda((t-p)/(n-p+1)) - \overline{\lambda}$ and $z_t(\boldsymbol{\theta}) = \epsilon_t(\boldsymbol{\theta}) - \overline{\epsilon(\boldsymbol{\theta})}$. The results of this theorem follow from the proof of Theorem 1 in Jaeckel (1972), where properties are given for $\sum_{t=p+1}^n a_n(t) z_{(t)}(\boldsymbol{\theta})$. ■

We are therefore estimating $\boldsymbol{\theta}_0 = (\alpha_{01}/\alpha_{00}, \ldots, \alpha_{0p}/\alpha_{00}, \beta_{01}, \ldots, \beta_{0q})'$ by minimizing the rank-based residual dispersion function (2.5). Other robust GARCH estimation techniques considered in the literature, M-estimation

https://doi.org/10.1017/S0266466612000047 Published online by Cambridge University Press

(Mukherjee, 2008; Muler and Yohai, 2008), least absolute deviations (Peng and Yao, 2003), and QML corresponding to noise distributions other than Gaussian (Berkes and Horváth, 2004), are also not used to directly estimate $\boldsymbol{\eta}_0 := (\alpha_{00}, \alpha_{01}, \ldots, \alpha_{0p}, \beta_{01}, \ldots, \beta_{0q})'$. Those methods are used instead to estimate $(\alpha_{00}/c, \alpha_{01}/c, \ldots, \alpha_{0p}/c, \beta_{01}, \ldots, \beta_{0q})'$, where $c > 0$ is unknown when the noise distribution is unknown. Gaussian QML (Berkes, Horváth, and Kokoszka, 2003; Francq and Zakoïan, 2004) and ML (Berkes and Horváth, 2004) can be used to directly estimate $\boldsymbol{\eta}_0$, but, because Gaussian QML estimators (QMLEs) have a rate of convergence slower than $n^{1/2}$ when $\mathbb{E}\{Z_t^4\} = \infty$ (Hall and Yao, 2003) and the noise distribution is unknown in practice, they are not robust techniques. Therefore, for estimating $\boldsymbol{\eta}_0$, we recommend that R-estimation be used as a preliminary technique. Once a R-estimate $\hat{\boldsymbol{\theta}}_R$ of $\boldsymbol{\theta}_0$ has been found, an appropriate noise distribution can be identified from the residuals $\{X_t/\sqrt{\tilde{\sigma}_t^2(\hat{\boldsymbol{\theta}}_R)}\}$, which resemble $\{\sqrt{\alpha_{00}}Z_t\}$ when $\hat{\boldsymbol{\theta}}_R$ is near $\boldsymbol{\theta}_0$, and $\boldsymbol{\eta}_0$ can be estimated via ML.

3. ASYMPTOTIC RESULTS

3.1. Limiting Distribution for R-Estimators

Let f and F denote the density and distribution functions for $\ln(Z_t^2)$. To obtain the limiting distribution for R-estimators of $\boldsymbol{\theta}_0$, we make the following additional assumptions.

A5. F is strictly increasing and differentiable on $\mathbb{R}$.

A6. f is uniformly continuous on $\mathbb{R}$.

A7. The weight function λ is bounded and left-continuous on $(0, 1)$.

We also consider the following conditions.

A8. $\theta_{0i} > 0$ for all $i \in \{1, \ldots, p+q\}$.

A9. The set $\{j : \beta_{0j} > 0\} \neq \emptyset$, and $\prod_{i=1}^{j_0} \alpha_{0i} > 0$ for $j_0 := \min\{j : \beta_{0j} > 0\}$.

A10. $\mathbb{E}X_t^4 < \infty$.

Finally, for values of $\boldsymbol{\theta} = (\theta_1, \ldots, \theta_{p+q})' \in \boldsymbol{\Theta}$, we define the series $\{\epsilon_t^*(\boldsymbol{\theta})\}$ so that, for all t, $\epsilon_t^*(\boldsymbol{\theta}) = \ln(X_t^2) - \ln(\breve{\sigma}_t^2(\boldsymbol{\theta}))$ with

$$\breve{\sigma}_t^2(\boldsymbol{\theta}) = 1 + \sum_{i=1}^{p} \theta_i X_{t-i}^2 + \sum_{j=1}^{q} \theta_{p+j} \breve{\sigma}_{t-j}^2(\boldsymbol{\theta}).$$

The sequence $\{\epsilon_t^*(\boldsymbol{\theta})\}_{t=p+1}^n$ is a stationary (stationarity follows from Berkes et al., 2003, Sect. 2) approximation of the residuals $\{\epsilon_t(\boldsymbol{\theta})\}_{t=p+1}^n$. Note that $\{\breve{\sigma}_t^2(\boldsymbol{\theta}_0)\} = \{\sigma_t^2/\alpha_{00}\}$ and $\{\epsilon_t^*(\boldsymbol{\theta}_0)\} = \{\epsilon_t\}$, with $\{\sigma_t^2\}$ defined in (2.2) and

https://doi.org/10.1017/S0266466612000047 Published online by Cambridge University Press

$\epsilon_t = \ln(\alpha_{00} Z_t^2)$. Since $\sum_{j=1}^q \theta_{p+j} < 1$ for $\boldsymbol{\theta} = (\theta_1, \ldots, \theta_{p+q})' \in \boldsymbol{\Theta}$, $1/(1 - \theta_{p+1}x - \cdots - \theta_{p+q}x^q)$ has a Laurent series expansion of the form $\sum_{j=0}^\infty c_j(\boldsymbol{\theta}) x^j$, where the coefficients $\{c_j(\boldsymbol{\theta})\}_{j=0}^\infty$ are geometrically decaying (Berkes et al., 2003, Sect. 2). Because

$$\frac{\partial \breve{\sigma}_t^2(\boldsymbol{\theta})}{\partial \theta_i} = \begin{cases} X_{t-i}^2 + \sum_{j=1}^q \theta_{p+j} \left(\partial \breve{\sigma}_{t-j}^2(\boldsymbol{\theta}) / \partial \theta_i \right), & i \in \{1, \ldots, p\}, \\ \breve{\sigma}_{t+p-i}^2(\boldsymbol{\theta}) + \sum_{j=1}^q \theta_{p+j} \left(\partial \breve{\sigma}_{t-j}^2(\boldsymbol{\theta}) / \partial \theta_i \right), & i \in \{p+1, \ldots, p+q\}, \end{cases}$$

it follows that the first partial derivatives of $\epsilon_t^*(\boldsymbol{\theta})$ are given by

$$\begin{aligned} \frac{\partial \epsilon_t^*(\boldsymbol{\theta})}{\partial \theta_i} &= \frac{-1}{\breve{\sigma}_t^2(\boldsymbol{\theta})} \frac{\partial \breve{\sigma}_t^2(\boldsymbol{\theta})}{\partial \theta_i} \\ &= \begin{cases} -\left(\sum_{j=0}^\infty c_j(\boldsymbol{\theta}) X_{t-i-j}^2 \right) / \breve{\sigma}_t^2(\boldsymbol{\theta}), & i \in \{1, \ldots, p\}, \\ -\left(\sum_{j=0}^\infty c_j(\boldsymbol{\theta}) \breve{\sigma}_{t+p-i-j}^2(\boldsymbol{\theta}) \right) / \breve{\sigma}_t^2(\boldsymbol{\theta}), & i \in \{p+1, \ldots, p+q\}. \end{cases} \end{aligned} \tag{3.1}$$

When $\boldsymbol{\theta} = \boldsymbol{\theta}_0$,

$$\frac{\partial \epsilon_t^*(\boldsymbol{\theta}_0)}{\partial \theta_i} = \begin{cases} -\alpha_{00} \left(\sum_{j=0}^\infty c_j(\boldsymbol{\theta}_0) X_{t-i-j}^2 \right) / \sigma_t^2, & i \in \{1, \ldots, p\}, \\ -\left(\sum_{j=0}^\infty c_j(\boldsymbol{\theta}_0) \sigma_{t+p-i-j}^2 \right) / \sigma_t^2, & i \in \{p+1, \ldots, p+q\} \end{cases}$$

(these are right derivatives when $\theta_{0i} = 0$), with $\sum_{j=0}^\infty c_j(\boldsymbol{\theta}_0) x^j = 1/(1 - \beta_{01}x - \cdots - \beta_{0q}x^q)$. We are now able to give the limiting distribution for R-estimators of $\boldsymbol{\theta}_0$.

THEOREM 3.1. *If A1–A7 and either A8, A9, or A10 hold, then there exists a sequence of minimizers $\hat{\boldsymbol{\theta}}_R \in \boldsymbol{\Theta}$ of $D_n(\cdot)$ in (2.5) such that*

$$n^{1/2} \left(\hat{\boldsymbol{\theta}}_R - \boldsymbol{\theta}_0 \right) \xrightarrow{d} \boldsymbol{\xi} := \operatorname{argmin}_{\mathbf{u} \in \Lambda} (\mathbf{u} - \mathbf{Y})' \tilde{K} \boldsymbol{\Gamma} (\mathbf{u} - \mathbf{Y}), \tag{3.2}$$

where $\Lambda = \Lambda_1 \times \cdots \times \Lambda_{p+q}$, with $\Lambda_i = \mathbb{R}$ if $\theta_{0i} > 0$ and $\Lambda_i = [0, \infty)$ if $\theta_{0i} = 0$, $\mathbf{Y} \sim N(\mathbf{0}, \boldsymbol{\Sigma})$, $\boldsymbol{\Sigma} := \tilde{J} \tilde{K}^{-2} \boldsymbol{\Gamma}^{-1}$, $\tilde{J} := \int_0^1 \lambda^2(x)\, dx - (\int_0^1 \lambda(x)\, dx)^2$, $\tilde{K} := \int_{-\infty}^\infty f(x)\, d\lambda(F(x))$, and

$$\boldsymbol{\Gamma} := \mathbb{E} \left\{ \left[\frac{\partial \epsilon_t^*(\boldsymbol{\theta}_0)}{\partial \boldsymbol{\theta}} - \mathbb{E} \left\{ \frac{\partial \epsilon_t^*(\boldsymbol{\theta}_0)}{\partial \boldsymbol{\theta}} \right\} \right] \left[\frac{\partial \epsilon_t^*(\boldsymbol{\theta}_0)}{\partial \boldsymbol{\theta}} - \mathbb{E} \left\{ \frac{\partial \epsilon_t^*(\boldsymbol{\theta}_0)}{\partial \boldsymbol{\theta}} \right\} \right]' \right\}.$$

The limiting random vector $\boldsymbol{\xi}$ is a unique, finite value almost surely.

Proof. By Lemma A.4 in the Appendix, $S_n(\mathbf{u}) = D_n(\boldsymbol{\theta}_0 + n^{-1/2}\mathbf{u}) - D_n(\boldsymbol{\theta}_0) \xrightarrow{d} S(\mathbf{u}) = \mathbf{u}'\mathbf{N} + \tilde{K}\mathbf{u}'\boldsymbol{\Gamma}\mathbf{u}/2$ on $C(\Lambda)$, where $\mathbf{N} \sim N(\mathbf{0}, \tilde{J}\boldsymbol{\Gamma})$ and $C(\Lambda)$ is the space of continuous functions on Λ where convergence is equivalent to uniform convergence on every compact subset. The derivatives $\partial \epsilon_t^*(\boldsymbol{\theta}_0)/\partial \theta_i$,

https://doi.org/10.1017/S0266466612000047 Published online by Cambridge University Press

$i \in \{1, \ldots, p+q\}$, have finite second moments (in the case of A8 this follows from the proof of Theorem 2.2(i) in Francq and Zakoïan, 2004, and in the case of A9 or A10 this follows from the proof of Lemma 8 in Francq and Zakoïan, 2007) so the matrix $\mathbf{\Gamma}$ exists, and, following the proof of Theorem 2.2(ii) in Francq and Zakoïan (2004), $\mathbf{\Gamma}$ is positive definite. Consequently, because $\tilde{K} > 0$ and Λ is a convex space, $S(\mathbf{u})$ has a unique minimum on Λ almost surely. Because $n^{1/2}(\hat{\boldsymbol{\theta}}_R - \boldsymbol{\theta}_0)$ minimizes $S_n(\mathbf{u})$, it follows from the proof of Lemma 2.2 and Remark 1 in Davis, Knight, and Liu (1992) that $n^{1/2}(\hat{\boldsymbol{\theta}}_R - \boldsymbol{\theta}_0) \xrightarrow{d} \operatorname{argmin}_{\mathbf{u}\in\Lambda} S(\mathbf{u})$. The result of this theorem holds because $\mathbf{N} \stackrel{d}{=} -\tilde{K}\mathbf{\Gamma}\mathbf{Y}$, which implies that $\operatorname{argmin}_{\mathbf{u}\in\Lambda} S(\mathbf{u}) \stackrel{d}{=} \operatorname{argmin}_{\mathbf{u}\in\Lambda}(\mathbf{Y}'\tilde{K}\mathbf{\Gamma}\mathbf{Y} - 2\mathbf{u}'\tilde{K}\mathbf{\Gamma}\mathbf{Y} + \mathbf{u}'\tilde{K}\mathbf{\Gamma}\mathbf{u}) = \operatorname{argmin}_{\mathbf{u}\in\Lambda}(\mathbf{u} - \mathbf{Y})'\tilde{K}\mathbf{\Gamma}(\mathbf{u} - \mathbf{Y})$. ■

If all parameter values are positive (i.e., A8 holds), it follows from Theorem 3.1 that $\hat{\boldsymbol{\theta}}_R$ is asymptotically normal.

COROLLARY 3.1. *If A1–A8 hold, then*

$$n^{1/2}\left(\hat{\boldsymbol{\theta}}_R - \boldsymbol{\theta}_0\right) \xrightarrow{d} \mathbf{Y} \sim N(\mathbf{0}, \mathbf{\Sigma}). \tag{3.3}$$

Proof. This result follows from (3.2), because $\tilde{K}\mathbf{\Gamma}$ is positive definite and, under A8, $\Lambda = \mathbf{R}^{p+q}$. ■

Remark 1. Assumptions A5 and A6 are mild. They hold, for example, if Z_t has a Laplace, logistic, $N(0,1)$, or rescaled student's t (rescaled to have unit variance) distribution.

Remark 2. As discussed in the proof of Theorem 3.1, given A8, A9, or A10, the derivatives $\partial\epsilon_t^*(\boldsymbol{\theta}_0)/\partial\theta_i$, $i \in \{1, \ldots, p+q\}$, have finite second moments. Because A8 is required for GARCH parameter estimates to be asymptotically unbiased and normal, it is a standard assumption in the literature. Conditions A9 and A10 are similar to assumptions introduced by Francq and Zakoïan (2007), who derived the limiting distribution for Gaussian QMLEs when some GARCH parameter values may be zero. In the absence of A8 and A9, $\mathbb{E}X_t^6 < \infty$, instead of A10 ($\mathbb{E}X_t^4 < \infty$), is required for the $n^{1/2}$-consistency of the Gaussian QMLEs, however (Francq and Zakoïan, 2007). Note that $\alpha_{01} > 0$ and $\beta_{01} > 0$ are sufficient for A9.

Remark 3. When a parameter vector may lie on a boundary of its parameter space, it is standard for estimators to converge in distribution to the minimizer of a random quadratic function on a convex space (D.W.K. Andrews, 1999). Francq and Zakoïan (2007) show this is true for Gaussian QMLEs of GARCH model parameters. The form of the limiting distribution for $\hat{\boldsymbol{\theta}}_R$ in (3.2) is therefore to be expected. Let the vector $(Y_1, \ldots, Y_{p+q})'$ contain the Gaussian random elements of $\mathbf{Y} \sim N(\mathbf{0}, \mathbf{\Sigma})$. If only the jth element of $\boldsymbol{\theta}_0$, θ_{0j}, equals zero and all other elements of $\boldsymbol{\theta}_0$ are positive, it follows from (3.2) that $n^{1/2}(\hat{\theta}_{j,R} - \theta_{0j}) = n^{1/2}\hat{\theta}_{j,R} \xrightarrow{d}$

https://doi.org/10.1017/S0266466612000047 Published online by Cambridge University Press

$Y_j I\{Y_j \geq 0\}$ ($I\{\cdot\}$ represents the indicator function), and so $\hat{\theta}_{j,R}$ is asymptotically half-normal in this case.

Remark 4. If f, the density function for $\ln(Z_t^2)$, is almost everywhere differentiable on $\mathbb{R}$, using integration by parts, it can be shown that $\tilde{K} = -\int_{-\infty}^{\infty} f'(x)\lambda(F(x))\,dx = -\int_0^1 [f'(F^{-1}(x))/f(F^{-1}(x))]\lambda(x)\,dx$. In practice, these integrals can be easier to evaluate than $\int_{-\infty}^{\infty} f(x)\,d\lambda(F(x))$.

Remark 5. Let f_Z represent the density function for the i.i.d. noise process $\{Z_t\}$ and, for $\boldsymbol{\eta} = (\eta_0, \eta_1, \ldots, \eta_{p+q})' \in (0,\infty) \times \boldsymbol{\Theta}$, let $\epsilon_t^{\dagger}(\boldsymbol{\eta}) = \ln(X_t^2) - \ln(\sigma_t^2(\boldsymbol{\eta}))$ with $\sigma_t^2(\boldsymbol{\eta}) = \eta_0 + \sum_{i=1}^{p} \eta_i X_{t-i}^2 + \sum_{j=1}^{q} \eta_{p+j}\sigma_{t-j}^2(\boldsymbol{\eta})$ $\forall t$. From Berkes and Horváth (2004), under general conditions that include A8, ML estimators of $\boldsymbol{\eta}_0 = (\alpha_{00}, \alpha_{01}, \ldots, \alpha_{0p}, \beta_{01}, \ldots, \beta_{0q})'$ are asymptotically normal with mean $\boldsymbol{\eta}_0$ and covariance matrix $4n^{-1}\tau^2\mathbf{A}^{-1}$, where $\tau^2 = (\mathbb{E}\{[1 + Z_t f_Z'(Z_t)/f_Z(Z_t)]^2\})^{-1}$ and $\mathbf{A} = \mathbb{E}\{[\partial\epsilon_t^{\dagger}(\boldsymbol{\eta}_0)/\partial\boldsymbol{\eta}][\partial\epsilon_t^{\dagger}(\boldsymbol{\eta}_0)/\partial\boldsymbol{\eta}]'\}$. Using the delta method (see, e.g., Rao, 1973, pp. 387–388), the corresponding estimators of $\boldsymbol{\theta}_0$, $\hat{\boldsymbol{\theta}}_{ML} := (\hat{\alpha}_{1,ML}/\hat{\alpha}_{0,ML}, \ldots, \hat{\alpha}_{p,ML}/\hat{\alpha}_{0,ML}, \hat{\beta}_{1,ML}, \ldots, \hat{\beta}_{q,ML})'$, are asymptotically Gaussian with mean $\boldsymbol{\theta}_0$ and covariance matrix $4n^{-1}\tau^2\mathbf{B}\mathbf{A}^{-1}\mathbf{B}'$, where $\mathbf{B}$ is the $(p+q) \times (p+q+1)$ Jacobian matrix

$$\mathbf{B} = \begin{bmatrix} -\alpha_{01}/\alpha_{00}^2 & 1/\alpha_{00} & 0 & \cdots & 0 & 0\cdots 0 \\ -\alpha_{02}/\alpha_{00}^2 & 0 & 1/\alpha_{00} & \cdots & 0 & 0\cdots 0 \\ \vdots & \vdots & \vdots & \ddots & \vdots & \vdots\ddots\vdots \\ -\alpha_{0p}/\alpha_{00}^2 & 0 & 0 & \cdots & 1/\alpha_{00} & 0\cdots 0 \\ 0 & 0 & 0 & \cdots & 0 & 1\cdots 0 \\ \vdots & \vdots & \vdots & \ddots & \vdots & \vdots\ddots\vdots \\ 0 & 0 & 0 & \cdots & 0 & 0\cdots 1 \end{bmatrix}.$$

For any p and q, matrix algebra can be used to show that $\mathbf{B}\mathbf{A}^{-1}\mathbf{B}' = \boldsymbol{\Gamma}^{-1}$ and therefore under Assumptions A1–A8, the asymptotic relative efficiency (ARE) for R-estimation with respect to ML is $4\tau^2\tilde{J}^{-1}\tilde{K}^2$.

If the weight function λ is proportional to $-f'(F^{-1}(x))/f(F^{-1}(x))$, then $\tilde{J}^{-1}\tilde{K}^2 = \mathbb{E}\{[f'(\ln(Z_t^2))/f(\ln(Z_t^2))]^2\}$ (Jurečková and Sen, 1996, Sect. 3.4). In addition, because $f(x) = \sqrt{e^x} f_Z(\sqrt{e^x})$ when the distribution for Z_t is symmetric about zero, it can be shown that $f'(\ln(Z_t^2))/f(\ln(Z_t^2)) = [1 + Z_t f_Z'(Z_t)/f_Z(Z_t)]/2$ in the symmetric case. Consequently, when the distribution for Z_t is symmetric about zero and a weight function

$$\lambda_f(x) \propto -f'(F^{-1}(x))/f(F^{-1}(x))$$

is used, $\tilde{J}^{-1}\tilde{K}^2 = \mathbb{E}\{[1 + Z_t f_Z'(Z_t)/f_Z(Z_t)]^2\}/4 = (4\tau^2)^{-1}$ and, if A1–A8 hold, R-estimation has the same asymptotic efficiency as ML.

https://doi.org/10.1017/S0266466612000047 Published online by Cambridge University Press

Remark 6. When $Z_t \sim N(0,1)$, the optimal weight function $\lambda_f(x) \propto [\Phi^{-1}((x+1)/2)]^2 - 1$, where Φ represents the standard normal distribution function. Hence, $\tilde{J}^{-1}\tilde{K}^2 = (4\tau^2)^{-1}$ when $Z_t \sim N(0,1)$ and $\lambda(x) = [\Phi^{-1}((x+1)/2)]^2 - 1$. However, the function $\lambda_N(x) := [\Phi^{-1}((x+1)/2)]^2 - 1$ does not satisfy Assumption A7, because it is unbounded as $x \to 1$, and so limiting results (3.2) and (3.3) do not necessarily hold for R-estimates obtained using λ_N. Bounded weight functions closely approximating λ_N that do satisfy the assumptions can, however, be found. For example, let $\lambda_{m,N}(x) = \lambda_N(x) I\{0 < x \leq 1 - 1/m\} + \lambda_N(1-1/m) I\{x > 1 - 1/m\}$, with $m \geq 2$. This weight function satisfies Assumptions A4 and A7 and, as $m \to \infty$, $\lambda_{m,N}$ converges pointwise to λ_N on $(0,1)$. In addition, $\tilde{J}_m := \int_0^1 \lambda_{m,N}^2(x)\,dx - (\int_0^1 \lambda_{m,N}(x)\,dx)^2 \to \int_0^1 \lambda_N^2(x)\,dx - (\int_0^1 \lambda_N(x)\,dx)^2$ and $\tilde{K}_m := \int_{-\infty}^{\infty} f(x)\,d\lambda_{m,N}(F(x)) \to \int_{-\infty}^{\infty} f(x)\,d\lambda_N(F(x))$. Therefore, in the case of Gaussian noise, large $m \geq 2$ can be chosen so that R-estimation with weight function $\lambda_{m,N}$ has essentially the same asymptotic efficiency as ML.

Gaussian ML estimators of GARCH model parameters are also consistent when the noise distribution is non-Gaussian (Francq and Zakoïan, 2004), and when all parameter values are positive and $\mathbb{E}\{Z_t^4\} < \infty$, Gaussian QMLEs of $\boldsymbol{\eta}_0$ are asymptotically normal with mean $\boldsymbol{\eta}_0$ and covariance matrix $n^{-1}(\mathbb{E}\{Z_t^4\} - 1)\mathbf{A}^{-1}$ (Berkes and Horváth, 2004; Francq and Zakoïan, 2004). It follows that the corresponding estimators of $\boldsymbol{\theta}_0$, $\hat{\boldsymbol{\theta}}_{GML} := (\hat{\alpha}_{1,GML}/\hat{\alpha}_{0,GML}, \ldots, \hat{\alpha}_{p,GML}/\hat{\alpha}_{0,GML}, \hat{\beta}_{1,GML}, \ldots, \hat{\beta}_{q,GML})'$, have a limiting normal distribution with mean $\boldsymbol{\theta}_0$ and covariance matrix $n^{-1}(\mathbb{E}\{Z_t^4\}-1)\mathbf{B}\mathbf{A}^{-1}\mathbf{B}' = n^{-1}(\mathbb{E}\{Z_t^4\}-1)\boldsymbol{\Gamma}^{-1}$, and so the ARE for R-estimation with respect to Gaussian QML is $(\mathbb{E}\{Z_t^4\} - 1)\tilde{J}^{-1}\tilde{K}^2$. The rate of convergence for Gaussian QMLEs is slower than $n^{1/2}$ when $\mathbb{E}\{Z_t^4\} = \infty$, however (Hall and Yao, 2003). Because $\mathbb{E}\{Z_t^4\} < \infty$ is not required for the $n^{1/2}$-consistency of $\hat{\boldsymbol{\theta}}_R$, R-estimation is more robust than traditional Gaussian QML.

In Table 3.1 we give the values of ARE $(\mathbb{E}\{Z_t^4\} - 1)\tilde{J}^{-1}\tilde{K}^2$ (rounded to the nearest three decimal places) when the weight function is $\lambda_{m,N}$ with m large (i.e., $\lambda(x) \approx \lambda_N(x)$) and the noise distribution is Laplace, logistic, $N(0,1)$, and rescaled t with various degrees of freedom. Note that, in the case of rescaled t_3 noise, we have an ARE of ∞, because in this case $\mathbb{E}\{Z_t^4\} = \infty$ and so Gaussian QMLEs have a rate of convergence slower than $n^{1/2}$. Because all AREs in Table 3.1 are greater than or equal to one, with equality only when the noise distribution is Gaussian, R-estimation is not only more robust than Gaussian QML but also tends to be more efficient. In the case of R-estimation for linear model parameters, R-estimation with the weight function that is optimal for Gaussian noise is always at least as asymptotically efficient as Gaussian QML (Chernoff and Savage, 1958; Gastwirth and Wolff, 1968; for discussion see Hettmansperger and McKean, 1998). As can be seen in Table 3.1, this is true in the case of GARCH parameter estimation for commonly used noise distributions, but it is possible for $(\mathbb{E}\{Z_t^4\} - 1)\tilde{J}^{-1}\tilde{K}^2$ to be much less than one

https://doi.org/10.1017/S0266466612000047 Published online by Cambridge University Press

TABLE 3.1. AREs for R-estimation with $\lambda(x) \approx \lambda_N(x)$ with respect to Gaussian QML

Distribution for noise $\{Z_t\}$	ARE
Laplace	1.221
Logistic	1.118
$N(0, 1)$	1.000
t_3	∞
t_5	2.321
t_7	1.341
t_9	1.167
t_{12}	1.082
t_{15}	1.049
t_{20}	1.026
t_{30}	1.011

when $\lambda(x) \approx [\Phi^{-1}((x+1)/2)]^2 - 1$, however (e.g., when $\ln(Z_t^2) \sim N(\mu = -0.125, \sigma^2 = 0.25)$, $(\mathbb{E}\{Z_t^4\} - 1)\tilde{J}^{-1}\tilde{K}^2 = 0.787$).

Remark 7. R-estimation is also relatively efficient when compared to more robust GARCH estimation techniques. To demonstrate, in Table 3.2, we give AREs for R-estimation with weight function $\lambda_{t_7}(x) := [7\{F_{t_7}^{-1}((x+1)/2)\}^2 - 5]/[\{F_{t_7}^{-1}((x+1)/2)\}^2 + 5]$, where F_{t_7} represents the distribution function for

TABLE 3.2. AREs for R-estimation with weight function λ_{t_7} with respect to (A) Laplace QML, (B) rescaled student's t_7 QML, (C) least absolute deviations, and (D) Gaussian QML

	ARE			
Distribution for noise $\{Z_t\}$	(A)	(B)	(C)	(D)
Laplace	0.986	1.014	2.053	1.233
Logistic	1.013	1.002	2.097	1.139
$N(0, 1)$	1.094	1.041	2.607	0.958
t_3	1.411	1.052	1.551	∞
t_5	1.058	1.005	1.867	2.489
t_7	1.029	1.000	2.040	1.400
t_9	1.028	1.002	2.147	1.198
t_{12}	1.036	1.007	2.250	1.093
t_{15}	1.043	1.011	2.315	1.050
t_{20}	1.053	1.017	2.383	1.016
t_{30}	1.064	1.024	2.454	0.991

https://doi.org/10.1017/S0266466612000047 Published online by Cambridge University Press

rescaled t_7 noise, with respect to (A) Laplace QML (Berkes and Horváth, 2004; Mukherjee, 2008), (B) rescaled student's t_7 QML (Muler and Yohai, 2008), and (C) the least absolute deviations estimation technique proposed by Peng and Yao (2003) involving log-transformed GARCH residuals. The weight function λ_{t_7} is optimal when the noise distribution is rescaled student's t_7 and satisfies Assumptions A4 and A7. Under general conditions, which include A8, estimators (A)–(C) are asymptotically normal and $n^{1/2}$-consistent for $(\alpha_{00}/c, \alpha_{01}/c, \ldots, \alpha_{0p}/c, \beta_{01}, \ldots, \beta_{0q})'$, for different values of $c > 0$. These techniques can therefore also be used to obtain $n^{1/2}$-consistent and asymptotically normal estimators of $\boldsymbol{\theta}_0$. For the noise distributions considered in Table 3.2, R-estimation with weight function λ_{t_7} is more efficient than Laplace QML for all distributions except Laplace, and R-estimation uniformly dominates rescaled student's t_7 QML and least absolute deviations with respect to asymptotic efficiency. AREs for R-estimation with weight function λ_{t_7} with respect to Gaussian QML are also given in Table 3.2 (column (D)). Because, when compared to other techniques, R-estimation with weight function λ_{t_7} performs well for light, medium, and heavier-tailed noise distributions, we recommend that it or R-estimation with a similar weight function be used in practice when the noise distribution is unknown. The weight function λ_{t_7} is plotted in Figure 3.1, along with λ_N. Note that $\lambda_{t_7}(x)$ and $\lambda_N(x)$ are fairly similar except near $x = 1$.

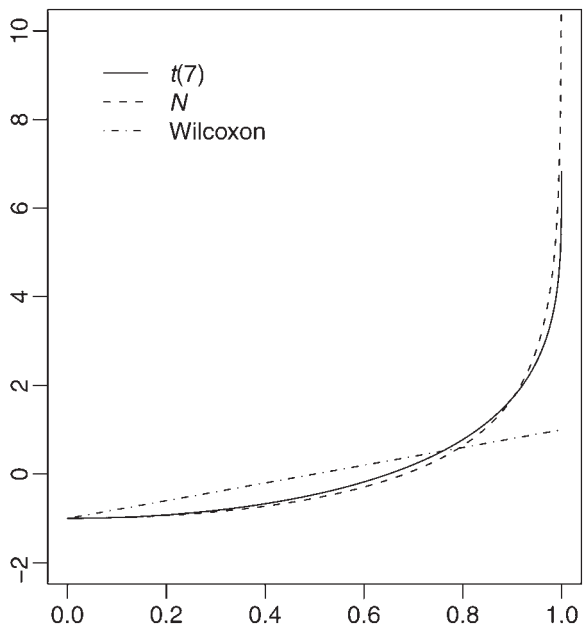


FIGURE 3.1. The weight functions λ_{t_7}, λ_N, and Wilcoxon weight function λ_W.

https://doi.org/10.1017/S0266466612000047 Published online by Cambridge University Press

TABLE 3.3. AREs for R-estimation with the Wilcoxon function with respect to R-estimation with weight function λ_{t_7}

Distribution for noise $\{Z_t\}$	ARE
Laplace	0.760
Logistic	0.736
$N(0, 1)$	0.634
t_3	0.911
t_5	0.801
t_7	0.753
t_9	0.727
t_{12}	0.703
t_{15}	0.690
t_{20}	0.676
t_{30}	0.662

Remark 8. The Wilcoxon weight function $\lambda_W(x) = 2x - 1$ is also shown in Figure 3.1. When estimating linear regression or linear time series model parameters via R-estimation, it is generally the recommended weight function when the noise distribution is unknown, because the corresponding R-estimators tend to be relatively efficient (see, e.g., Hettmansperger and McKean, 1998; B. Andrews, Davis, and Breidt, 2007; B. Andrews, 2008). In the case of linear model estimation, λ_W is the optimal weight function when the noise distribution is logistic, and for R-estimation of GARCH model parameters, λ_W is optimal when $\ln(Z_t^2)$ is logistic. However, R-estimation with the Wilcoxon function is not relatively efficient for GARCH models with commonly used noise distributions. To demonstrate, in Table 3.3 we give AREs for R-estimation with the Wilcoxon function with respect to R-estimation with weight function λ_{t_7}. All AREs are less than one.

Remark 9. For estimating $\boldsymbol{\eta}_0 = (\alpha_{00}, \alpha_{01}, \ldots, \alpha_{0p}, \beta_{01}, \ldots, \beta_{0q})'$, an alternative to using Gaussian QML or ML with a noise distribution resembling the empirical distribution for R-estimation residuals is to estimate α_{00} via $\hat{\alpha}_0 := n^{-1} \sum_{t=p+1}^{n} \exp(\epsilon_t(\hat{\boldsymbol{\theta}}_R))$ and then let $\hat{\boldsymbol{\eta}} = (\hat{\alpha}_0, \hat{\alpha}_0\hat{\theta}_{1,R}, \ldots, \hat{\alpha}_0\hat{\theta}_{p,R}, \hat{\theta}_{p+1,R}, \ldots, \hat{\theta}_{p+q,R})'$. Because

$$\hat{\alpha}_0 = n^{-1} \sum_{t=p+1}^{n} \exp(\epsilon_t(\hat{\boldsymbol{\theta}}_R)) = n^{-1} \sum_{t=p+1}^{n} \frac{X_t^2}{\tilde{\sigma}_t^2(\hat{\boldsymbol{\theta}}_R)} = n^{-1} \sum_{t=p+1}^{n} \alpha_{00} Z_t^2 \frac{\sigma_t^2/\alpha_{00}}{\tilde{\sigma}_t^2(\hat{\boldsymbol{\theta}}_R)} \tag{3.4}$$

and, using the proof of Lemma 5.8 in Berkes et al. (2003), it can be shown that the right-hand side of (3.4) equals $n^{-1} \sum_{t=p+1}^{n} \alpha_{00} Z_t^2 + o_p(1)$, $\hat{\alpha}_0 \xrightarrow{P} \alpha_{00}$. By

https://doi.org/10.1017/S0266466612000047 Published online by Cambridge University Press

Theorem 3.1, $\hat{\boldsymbol{\theta}}_R \xrightarrow{P} \boldsymbol{\theta}_0 = (\alpha_{01}/\alpha_{00}, \ldots, \alpha_{0p}/\alpha_{00}, \beta_{01}, \ldots, \beta_{0q})'$, and so $\hat{\boldsymbol{\eta}}$ is consistent for $\boldsymbol{\eta}_0$. From Theorem 3.1 and Corollary 3.1, we have the limiting distributions for $\hat{\theta}_{i,R}$, $i = p+1, \ldots, p+q$, and when $\mathbb{E}\{Z_t^4\} < \infty$, it can be shown that $n^{1/2}(\hat{\alpha}_0 - \alpha_{00}) \xrightarrow{d} \alpha_{00}(\boldsymbol{\xi}'\mathbb{E}\{\partial\epsilon_t^*(\boldsymbol{\theta}_0)/\partial\boldsymbol{\theta}\} + \tilde{Y})$ and, for $i \in \{1, \ldots, p\}$, $n^{1/2}(\hat{\alpha}_0\hat{\theta}_{i,R} - \alpha_{0i}) \xrightarrow{d} \alpha_{00}\xi_i + \alpha_{0i}(\boldsymbol{\xi}'\mathbb{E}\{\partial\epsilon_t^*(\boldsymbol{\theta}_0)/\partial\boldsymbol{\theta}\} + \tilde{Y})$, with the random vector $\boldsymbol{\xi}$ as defined in equation (3.2), $\tilde{Y} \sim N(0, \mathrm{Var}\{Z_t^2\})$ independent of $\boldsymbol{\xi}$, and ξ_i denotes the ith element of $\boldsymbol{\xi}$. It follows that, if A1–A8 hold and $\mathbb{E}\{Z_t^4\} < \infty$, then $n^{1/2}(\hat{\alpha}_0 - \alpha_{00}) \xrightarrow{d} N(0, \alpha_{00}^2[(\mathbb{E}\{\partial\epsilon_t^*(\boldsymbol{\theta}_0)/\partial\boldsymbol{\theta}\})'\,\boldsymbol{\Sigma}\,\mathbb{E}\{\partial\epsilon_t^*(\boldsymbol{\theta}_0)/\partial\boldsymbol{\theta}\} + \mathrm{Var}\{Z_t^2\}])$ and, for $i \in \{1, \ldots, p\}$, $n^{1/2}(\hat{\alpha}_0\hat{\theta}_{i,R} - \alpha_{0i}) \xrightarrow{d} N(0, \mathbf{C}_i'\boldsymbol{\Sigma}\,\mathbf{C}_i + \alpha_{0i}^2\mathrm{Var}\{Z_t^2\})$, where

$$\mathbf{C}_i := (\underbrace{0, \ldots, 0}_{i-1 \text{ times}}, \alpha_{00}, \underbrace{0, \ldots, 0}_{p+q-i \text{ times}})' + \alpha_{0i}\,\mathbb{E}\left\{\frac{\partial\epsilon_t^*(\boldsymbol{\theta}_0)}{\partial\boldsymbol{\theta}}\right\}.$$

However, as is the case for Gaussian QMLEs of $\boldsymbol{\eta}_0$, when $\mathbb{E}\{Z_t^4\} = \infty$, the estimators $\hat{\alpha}_0$ and $\hat{\alpha}_0\hat{\theta}_{i,R}$, $i \in \{1, \ldots, p\}$, can be shown to have a rate of convergence slower than $n^{1/2}$, and so this is not a robust technique.

3.2. Order Selection and Interval Estimation

To use Theorem 3.1 to make inferences about $\boldsymbol{\theta}_0$, estimates of $\tilde{K}$ and $\boldsymbol{\Gamma}$ are needed. Because λ is determined by the practitioner, $\tilde{J}$ is known. For estimating $\boldsymbol{\Gamma}$, we propose using

$$\hat{\boldsymbol{\Gamma}}_n := \frac{1}{n}\sum_{t=p+1}^{n}\left[\frac{\partial\epsilon_t(\hat{\boldsymbol{\theta}}_R)}{\partial\boldsymbol{\theta}} - \left(\frac{1}{n}\sum_{s=p+1}^{n}\frac{\partial\epsilon_s(\hat{\boldsymbol{\theta}}_R)}{\partial\boldsymbol{\theta}}\right)\right] \times \left[\frac{\partial\epsilon_t(\hat{\boldsymbol{\theta}}_R)}{\partial\boldsymbol{\theta}} - \left(\frac{1}{n}\sum_{s=p+1}^{n}\frac{\partial\epsilon_s(\hat{\boldsymbol{\theta}}_R)}{\partial\boldsymbol{\theta}}\right)\right]',$$

with $\{\epsilon_t(\boldsymbol{\theta})\}$ as defined in (2.3). Because $\hat{\boldsymbol{\theta}}_R \xrightarrow{P} \boldsymbol{\theta}_0$, by the proofs of Lemmas 11 and 12 in Francq and Zakoïan (2007), $\hat{\boldsymbol{\Gamma}}_n \xrightarrow{P} \boldsymbol{\Gamma}$. A consistent estimator for $\tilde{K}$ is given in the following theorem.

THEOREM 3.2. *Consider the empirical distribution function* $\hat{F}_n(x) := (n-p)^{-1}\sum_{t=p+1}^{n} I\{\epsilon_t(\hat{\boldsymbol{\theta}}_R) \le x\}$ *and the kernel density estimator*

$$\hat{f}_n(x) := \frac{1}{b_n n}\sum_{t=p+1}^{n}\kappa\left(\frac{\epsilon_t(\hat{\boldsymbol{\theta}}_R) - x}{b_n}\right), \tag{3.5}$$

https://doi.org/10.1017/S0266466612000047 Published online by Cambridge University Press

where κ is a uniformly continuous, differentiable kernel density function on $\mathbb{R}$ such that κ' is uniformly continuous on $\mathbb{R}$ and $\int |x \ln |x||^{1/2} |\kappa'(x)|\, dx < \infty$, and the bandwidth sequence $\{b_n\}$ is chosen so that $b_n \xrightarrow{P} 0$ and $b_n^2 \sqrt{n} \xrightarrow{P} \infty$ as $n \to \infty$. If A1–A7 and either A8, A9, or A10 hold, then

$$\hat{K}_n := \int_{-\infty}^{\infty} \hat{f}_n(x)\, d\lambda(\hat{F}_n(x)) = \sum_{t=p+1}^{n} \hat{f}_n\left(\epsilon_{(t)}(\hat{\boldsymbol{\theta}}_R)\right) \left[\lambda\left(\frac{t-p}{n-p}\right) - \lambda\left(\frac{t-p-1}{n-p}\right)\right] \xrightarrow{P} \tilde{K}.$$

Proof. Let f_ϵ and F_ϵ denote the density and distribution functions for $\epsilon_t = \ln(\alpha_{00} Z_t^2)$. By Lemma A.5 in the Appendix, $\sup_{x \in \mathbb{R}} |\hat{f}_n(x) - f_\epsilon(x)| \xrightarrow{P} 0$, and using the Glivenko–Cantelli theorem, $\sup_{x \in \mathbb{R}} |\hat{F}_n(x) - F_\epsilon(x)| \xrightarrow{P} 0$. As a result, $\hat{K}_n \xrightarrow{P} \int_{-\infty}^{\infty} f_\epsilon(x)\, d\lambda(F_\epsilon(x))$ by the proof of Theorem 4.5.3 in Koul (2002). Because $f_\epsilon(x) = f(x - \ln(\alpha_{00}))$ and $F_\epsilon(x) = F(x - \ln(\alpha_{00}))$, $\tilde{K} = \int_{-\infty}^{\infty} f_\epsilon(x)\, d\lambda(F_\epsilon(x))$, and so the proof is complete. ■

It follows that $\hat{\boldsymbol{\Sigma}}_n := \tilde{J} \hat{K}_n^{-2} \hat{\boldsymbol{\Gamma}}_n^{-1}$ is consistent for $\boldsymbol{\Sigma} = \tilde{J} \tilde{K}^{-2} \boldsymbol{\Gamma}^{-1}$. Note that the Gaussian and student's t densities satisfy the conditions for the kernel density function κ in Theorem 3.2.

For GARCH order selection, we can test null hypotheses of the form

$$\mathrm{H}_0 : \theta_{0i_1} = \cdots = \theta_{0i_m} = 0, \tag{3.6}$$

with $1 \le i_1 < \cdots < i_m \le p + q$. Following Corollary 3.1, a corresponding Wald test statistic is given by

$$W_n := n(\mathbf{W}\hat{\boldsymbol{\theta}}_R)'(\mathbf{W}\hat{\boldsymbol{\Sigma}}_n \mathbf{W}')^{-1} \mathbf{W}\hat{\boldsymbol{\theta}}_R, \tag{3.7}$$

where $\mathbf{W} = [w_{j,k}]$ is the $m \times (p+q)$ matrix with $w_{\ell, i_\ell} = 1$ for $\ell \in \{1, \ldots, m\}$ and $w_{j,k} = 0$ otherwise. The limiting distribution for W_n under H_0 is given in the following theorem. This limiting distribution has a simple form when we are testing the nullity of just one parameter.

THEOREM 3.3. *Under (3.6), if A1–A7 and either A9 or A10 hold, then $W_n \xrightarrow{d} \tilde{W} := \boldsymbol{\xi}' \mathbf{W}' (\mathbf{W}\boldsymbol{\Sigma}\mathbf{W}')^{-1} \mathbf{W}\boldsymbol{\xi}$ as $n \to \infty$, with $\boldsymbol{\xi}$ as defined in (3.2). If $m = 1$ and $\theta_{0j} > 0$ for $j \ne i_1$, then $\tilde{W} \stackrel{d}{=} V^2 I\{V \ge 0\}$ with $V \sim N(0,1)$.*

Proof. By Theorem 3.1, under these conditions, $n^{1/2} \mathbf{W}\hat{\boldsymbol{\theta}}_R \xrightarrow{d} \mathbf{W}\boldsymbol{\xi}$. Because $\hat{\boldsymbol{\Sigma}}_n \xrightarrow{P} \boldsymbol{\Sigma}$, it follows that $W_n \xrightarrow{d} \tilde{W}$. If $m = 1$ and $\theta_{0j} > 0$ for $j \ne i_1$, following Remark 3, $\mathbf{W}\boldsymbol{\xi} = \xi_{i_1} = Y_{i_1} I\{Y_{i_1} \ge 0\}$ with $Y_{i_1} \sim N(0, \Sigma_{i_1,i_1})$ (Σ_{i_1,i_1} represents element i_1, i_1 of $\boldsymbol{\Sigma}$), and so $\tilde{W} = \Sigma_{i_1,i_1}^{-1} Y_{i_1}^2 I\{Y_{i_1} \ge 0\} \stackrel{d}{=} V^2 I\{V \ge 0\}$. ■

Therefore, assuming $\theta_{0j} > 0$ for $j \ne i_1$, we can reject the null hypothesis $\mathrm{H}_0 : \theta_{0i_1} = 0$ at level of significance $\alpha < \frac{1}{2}$ if $W_n > \chi^2_{1,1-2\alpha}$, where $\chi^2_{1,1-2\alpha}$

https://doi.org/10.1017/S0266466612000047 Published online by Cambridge University Press

represents the $1 - 2\alpha$ quantile of the chi-squared distribution with one degree of freedom. Finding a critical value for testing the nullity of multiple parameters $\theta_{0i_1}, \ldots, \theta_{0i_m}$ is not as simple because, in general, the distribution for the limiting random variable $\tilde{W}$ under H_0 depends on the values of the other model parameters and the distribution for the i.i.d. noise $\{Z_t\}$. However, one can examine the distribution for W_n under H_0 in practice by simulating GARCH series of length n with parameter vector $\boldsymbol{\theta}^* = (\theta_1^*, \ldots, \theta_{p+q}^*)'$, with $\theta_j^* = 0$ for $j \in \{i_1, \ldots, i_m\}$ and $\theta_j^* = \hat{\theta}_{j,R}$ otherwise, and i.i.d. noise following the empirical distribution for R-estimation residuals, and then obtaining the corresponding values for W_n.

Test statistics for testing the nullity of GARCH model parameters using Gaussian QMLEs are considered in Francq and Zakoïan (2009). Limiting results in Francq and Zakoïan (2009) require $\mathbb{E}\{Z_t^4\} < \infty$, however. In contrast, when A9 holds, higher order moment conditions are not required for the limiting result in Theorem 3.3.

Although parameter estimation for the ARCH(p) model with $\theta_{01} = \cdots = \theta_{0p} = 0$ is not considered in this paper because A2 is not satisfied, our results can be used to show that, under $H_0 : \theta_{01} = \cdots = \theta_{0p} = 0$, if $\mathbb{E}\{Z_t^4\} < \infty$, then $W_n \xrightarrow{d} \tilde{W} = \boldsymbol{\xi}' \boldsymbol{\Sigma}^{-1} \boldsymbol{\xi}$, where in this case $\boldsymbol{\xi} = (Y_1 I\{Y_1 \geq 0\}, \ldots, Y_p I\{Y_p \geq 0\})'$, $\mathbf{Y} = (Y_1, \ldots, Y_p)' \sim N(\mathbf{0}, \boldsymbol{\Sigma})$, and $\boldsymbol{\Sigma} = \tilde{J}\tilde{K}^{-2}(\mathrm{Var}\{Z_t^2\})^{-1}\mathbf{I}$ ($\mathbf{I}$ represents the identity matrix). It follows that, under these conditions, $W_n \xrightarrow{d} V_1^2 I\{V_1 \geq 0\} + \cdots + V_p^2 I\{V_p \geq 0\}$, where $V_1, \ldots, V_p$ are i.i.d. $N(0,1)$. Consequently, R-estimation and the Wald test statistic (3.7) can be used to identify ARCH-type conditional heteroskedasticity in an observed time series. However, because $\mathbb{E}\{Z_t^4\} < \infty$ is required, this technique is just as robust as more traditional Gaussian likelihood-based techniques (see Francq and Zakoïan, 2009). Therefore, in practice, the rank-based Wald test statistic is most useful for choosing between GARCH models of different orders that are under consideration for an observed conditionally heteroskedastic series. Once appropriate GARCH model orders p and q have been identified, confidence intervals for the elements of $\boldsymbol{\theta}_0$ can be obtained using the limiting normal result in Corollary 3.1 and the consistent estimate $\hat{\boldsymbol{\Sigma}}_n$ of $\boldsymbol{\Sigma}$.

4. NUMERICAL RESULTS

4.1. Simulation Study

In this section, we give the results of a simulation study to assess the quality of the asymptotic approximations for finite samples. First, for each of 1,000 replicates, we simulated a GARCH(1,1) series with parameters $\alpha_{00} = 0.01$, $\alpha_{01} = 0.5$, and $\beta_{01} = 0.4$ and found the R-estimate $\hat{\boldsymbol{\theta}}_R$ of $\boldsymbol{\theta}_0 = (0.5/0.01, 0.4)' = (50, 0.4)'$ by minimizing D_n in (2.5). To reduce the possibility of the optimizer getting trapped at local minima, we used 100 starting values for each replicate. Starting

https://doi.org/10.1017/S0266466612000047 Published online by Cambridge University Press

values for α_1 and $\beta_1 = \theta_2$, with $\alpha_1 + \beta_1 < 1$, were randomly chosen, and then, because $\alpha_1 + \beta_1 < 1$ implies $\mathrm{Var}\{X_t\} = \alpha_0/(1 - \alpha_1 - \beta_1)$, we used the values of $\alpha_1/[s_X^2(1 - \alpha_1 - \beta_1)]$, where s_X^2 represents sample variance for $\{X_t\}_{t=1}^n$, for the starting values of $\theta_1 = \alpha_1/\alpha_0$. We evaluated D_n at each of the 100 candidate values and then reduced the collection of initial values to the three with the smallest values of D_n. Using these three initial values as starting points, we found optimized values by implementing the Nelder–Mead algorithm (Nelder and Mead, 1965). The optimized value for which D_n was smallest was chosen to be $\hat{\boldsymbol{\theta}}_R$. By Corollary 3.1, in this GARCH(1,1) case, $\hat{\boldsymbol{\theta}}_R$ is asymptotically normal with mean $\boldsymbol{\theta}_0$ and covariance matrix $n^{-1}\boldsymbol{\Sigma} = n^{-1}\tilde{J}\tilde{K}^{-2}\boldsymbol{\Gamma}^{-1}$, with

$$\boldsymbol{\Gamma} = \mathrm{Var}\left\{\frac{\partial \epsilon_t^*(\boldsymbol{\theta}_0)}{\partial \boldsymbol{\theta}}\right\} = \mathrm{Var}\left\{\begin{bmatrix} \alpha_{00}(\sum_{j=0}^{\infty}\beta_{01}^j X_{t-1-j}^2)/\sigma_t^2 \\ (\sum_{j=0}^{\infty}\beta_{01}^j \sigma_{t-1-j}^2)/\sigma_t^2 \end{bmatrix}\right\}.$$

Confidence intervals for the elements of $\boldsymbol{\theta}_0$ were therefore constructed using the asymptotic normality and $\hat{\boldsymbol{\Sigma}}_n$, the consistent estimator of $\boldsymbol{\Sigma}$. For the kernel density estimator (3.5), we used the standard Gaussian kernel density function and, because of its recommendation in Silverman (1986, p. 48), we used bandwidth $b_n = 0.9n^{-1/5}\min\{s_\epsilon, IQR_\epsilon/1.34\}$, where s_ϵ and IQR_ϵ represent sample standard deviation and interquartile range for $\{\epsilon_t(\hat{\boldsymbol{\theta}}_R)\}_{t=p+1}^n$.

Results of these simulations for $N(0,1)$ and rescaled t_3 noise, and weight functions $\lambda_{t_7}(x) = [7\{F_{t_7}^{-1}((x+1)/2)\}^2 - 5]/[\{F_{t_7}^{-1}((x+1)/2)\}^2 + 5]$ and $\lambda_W(x) = 2x - 1$, are given in Table 4.1. We show the empirical means, standard deviations, and percentage coverages of nominal 95% confidence intervals for the R-estimates of $\boldsymbol{\theta}_0$. The asymptotic means and standard deviations are also given in Table 4.1. We see that the R-estimates appear nearly unbiased and the asymptotic standard deviations fairly accurately reflect the true variability of the estimates, so R-estimation with weight function λ_{t_7} is more precise than R-estimation with the Wilcoxon weight function. We also see that the confidence interval coverages are close to the nominal 95% level, especially when n =2,000. Normal probability plots show that the R-estimates are approximately normal, particularly for large n.

We also ran simulations to assess the accuracy of the asymptotic approximations in the case of one null parameter value by fitting GARCH(2,1) and GARCH(1,2) models to GARCH(1,1) series with $\boldsymbol{\theta}_0 = (50, 0.4)'$. When $\theta_{0j} = 0$, we obtained the Wald statistic (3.7) for testing $\mathrm{H}_0 : \theta_{0j} = 0$ and compared it to $\chi^2_{1,0.9} = 2.706$, the level 0.05 critical value for testing the nullity of one parameter assuming all other parameters are positive. Empirical means and standard deviations for the estimates, and the percentages of Wald statistics above 2.706 are given in Table 4.2. The R-estimators appear consistent in both the GARCH(2,1) and GARCH(1,2) cases, and the test sizes are close to the nominal 0.05 level, particularly when n is large.

https://doi.org/10.1017/S0266466612000047 Published online by Cambridge University Press

TABLE 4.1. Empirical means, standard deviations, and percent coverages of nominal 95% confidence intervals for R-estimates of GARCH model parameters. The $N(0,1)$ and rescaled t_3 noise distributions and weight functions λ_{t_7} and λ_W were used.

		Asymptotic		Empirical		
n	Mean	Std. dev. ($N(0,1)$, t_3)		Mean ($N(0,1)$, t_3)	Std. dev. ($N(0,1)$, t_3)	% coverage ($N(0,1)$, t_3)
A. Weight function $\lambda_{t_7}(x) = [7\{F_{t_7}^{-1}((x+1)/2)\}^2 - 5]/[\{F_{t_7}^{-1}((x+1)/2)\}^2 + 5]$						
250	$\theta_1 = \alpha_1/\alpha_0 = 50$	19.6, 25.3		48.3, 52.0	20.2, 28.6	85.4, 89.0
	$\theta_2 = \beta_1 = 0.4$	0.104, 0.158		0.378, 0.369	0.124, 0.178	92.6, 91.0
500	$\theta_1 = \alpha_1/\alpha_0 = 50$	13.9, 17.9		50.3, 49.8	14.4, 17.8	91.7, 91.0
	$\theta_2 = \beta_1 = 0.4$	0.073, 0.112		0.389, 0.385	0.079, 0.122	94.8, 95.1
2,000	$\theta_1 = \alpha_1/\alpha_0 = 50$	6.93, 8.94		49.8, 50.4	7.00, 8.83	94.1, 95.6
	$\theta_2 = \beta_1 = 0.4$	0.037, 0.056		0.398, 0.399	0.038, 0.057	96.6, 97.6
B. Wilcoxon weight function $\lambda_W(x) = 2x - 1$						
250	$\theta_1 = \alpha_1/\alpha_0 = 50$	24.6, 26.5		52.0, 52.9	32.9, 30.6	87.5, 88.5
	$\theta_2 = \beta_1 = 0.4$	0.130, 0.166		0.375, 0.377	0.148, 0.185	93.2, 92.1
500	$\theta_1 = \alpha_1/\alpha_0 = 50$	17.4, 18.7		51.2, 51.9	19.5, 21.3	91.1, 92.9
	$\theta_2 = \beta_1 = 0.4$	0.092, 0.117		0.389, 0.395	0.102, 0.120	94.2, 96.9
2,000	$\theta_1 = \alpha_1/\alpha_0 = 50$	8.70, 9.37		50.0, 50.7	8.79, 9.57	94.4, 95.5
	$\theta_2 = \beta_1 = 0.4$	0.046, 0.059		0.397, 0.395	0.046, 0.060	97.6, 96.7

4.2. GARCH Modeling

In Figure 4.1a, we give the daily log-returns for the Japanese yen to U.S. dollar exchange rate for 4 January 1993–31 December 2002. Sample autocorrelation functions for these data, $\{X_t\}_{t=1}^{2,514}$, and their absolute values and squares are given in Figure 4.1b–d with the bounds $\pm 1.96/\sqrt{2,514}$, approximate 95% confidence bounds for the sample correlations assuming the observations are i.i.d. (see, e.g., Brockwell and Davis, 1991, Sect. 7.2). Because the log-returns $\{X_t\}$ exhibit volatility clustering and appear uncorrelated but dependent, a GARCH model appears appropriate for this series. To find a suitable fitted model, we replaced the 17 values of $X_t = 0$ with $X_t = 10^{-8}$ (a very slight alteration of the series that allows us to assume A3 and compute $\{\ln(X_t^2)\}$) and fit low-order ARCH/GARCH models to the data via R-estimation with weight function $\lambda_{t_7}(x) = [7\{F_{t_7}^{-1}((x+1)/2)\}^2 - 5]/[\{F_{t_7}^{-1}((x+1)/2)\}^2 + 5]$. ARCH(1) and ARCH(2) models were not appropriate because those residuals were dependent, but, as is often the case with log-returns series, GARCH(1,1) residuals appeared independent. To demonstrate, in Figure 4.2 we give sample autocorrelation functions for the absolute values and squares of the residuals $\{X_t/\sqrt{\tilde{\sigma}_t^2(\hat{\boldsymbol{\theta}}_R)}\}$

https://doi.org/10.1017/S0266466612000047 Published online by Cambridge University Press

TABLE 4.2. Empirical means, standard deviations, and the percentages of Wald statistics above 2.706 for R-estimates of GARCH model parameters. The $N(0,1)$ and rescaled t_3 noise distributions and weight functions λ_{t_7} and λ_W were used.

		Empirical		
n		Mean ($N(0,1)$, t_3)	Std. dev. ($N(0,1)$, t_3)	% Wald stats. > 2.706 ($N(0,1)$, t_3)
A. Weight function $\lambda_{t_7}(x) = [7\{F_{t_7}^{-1}((x+1)/2)\}^2 - 5]/[\{F_{t_7}^{-1}((x+1)/2)\}^2 + 5]$				
250	$\theta_1 = \alpha_1/\alpha_0 = 50$	41.7, 42.3	21.2, 26.3	
	$\theta_2 = \alpha_2/\alpha_0 = 0$	7.15, 10.6	8.97, 14.5	16.2, 7.0
	$\theta_3 = \beta_1 = 0.4$	0.296, 0.273	0.171, 0.207	
500	$\theta_1 = \alpha_1/\alpha_0 = 50$	43.9, 43.8	15.5, 19.0	
	$\theta_2 = \alpha_2/\alpha_0 = 0$	4.87, 7.62	6.16, 9.58	13.7, 11.4
	$\theta_3 = \beta_1 = 0.4$	0.326, 0.307	0.124, 0.164	
2,000	$\theta_1 = \alpha_1/\alpha_0 = 50$	46.9, 46.6	7.72, 10.0	
	$\theta_2 = \alpha_2/\alpha_0 = 0$	2.66, 3.79	3.35, 5.00	9.1, 9.4
	$\theta_3 = \beta_1 = 0.4$	0.369, 0.359	0.053, 0.082	
250	$\theta_1 = \alpha_1/\alpha_0 = 50$	51.2, 52.0	22.5, 28.9	
	$\theta_2 = \beta_1 = 0.4$	0.297, 0.276	0.148, 0.188	
	$\theta_3 = \beta_2 = 0$	0.060, 0.080	0.095, 0.127	9.9, 7.1
500	$\theta_1 = \alpha_1/\alpha_0 = 50$	49.9, 51.6	14.5, 18.8	
	$\theta_2 = \beta_1 = 0.4$	0.337, 0.312	0.109, 0.144	
	$\theta_3 = \beta_2 = 0$	0.039, 0.059	0.065, 0.094	7.9, 8.3
2,000	$\theta_1 = \alpha_1/\alpha_0 = 50$	50.1, 50.4	7.11, 9.13	
	$\theta_2 = \beta_1 = 0.4$	0.368, 0.356	0.054, 0.081	
	$\theta_3 = \beta_2 = 0$	0.020, 0.030	0.030, 0.044	5.9, 6.5
B. Wilcoxon weight function $\lambda_W(x) = 2x - 1$				
250	$\theta_1 = \alpha_1/\alpha_0 = 50$	41.6, 43.7	26.2, 29.9	
	$\theta_2 = \alpha_2/\alpha_0 = 0$	8.48, 10.8	10.9, 14.0	7.9, 3.4
	$\theta_3 = \beta_1 = 0.4$	0.279, 0.286	0.191, 0.219	
500	$\theta_1 = \alpha_1/\alpha_0 = 50$	43.6, 45.2	18.7, 20.9	
	$\theta_2 = \alpha_2/\alpha_0 = 0$	6.44, 7.82	8.17, 10.2	13.2, 9.1
	$\theta_3 = \beta_1 = 0.4$	0.314, 0.308	0.146, 0.169	
2,000	$\theta_1 = \alpha_1/\alpha_0 = 50$	46.8, 46.5	9.61, 10.2	
	$\theta_2 = \alpha_2/\alpha_0 = 0$	2.91, 3.95	4.00, 5.48	7.9, 8.4
	$\theta_3 = \beta_1 = 0.4$	0.365, 0.360	0.071, 0.083	
250	$\theta_1 = \alpha_1/\alpha_0 = 50$	54.6, 56.0	42.7, 33.7	
	$\theta_2 = \beta_1 = 0.4$	0.288, 0.271	0.167, 0.201	
	$\theta_3 = \beta_2 = 0$	0.078, 0.090	0.124, 0.141	7.6, 8.1
500	$\theta_1 = \alpha_1/\alpha_0 = 50$	51.3, 52.3	18.5, 19.9	
	$\theta_2 = \beta_1 = 0.4$	0.326, 0.318	0.128, 0.156	
	$\theta_3 = \beta_2 = 0$	0.050, 0.063	0.083, 0.100	8.0, 8.6
2,000	$\theta_1 = \alpha_1/\alpha_0 = 50$	50.5, 50.6	8.84, 9.44	
	$\theta_2 = \beta_1 = 0.4$	0.363, 0.351	0.070, 0.085	
	$\theta_3 = \beta_2 = 0$	0.026, 0.032	0.039, 0.047	6.0, 7.1

https://doi.org/10.1017/S0266466612000047 Published online by Cambridge University Press

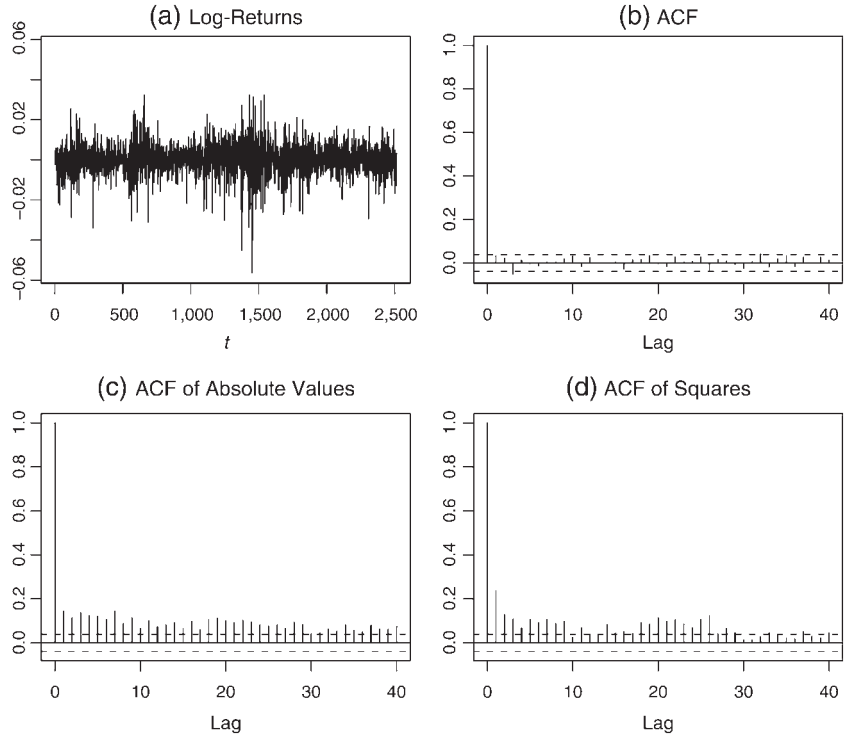


FIGURE 4.1. (a) The daily log-returns for the Japanese yen to U.S. dollar exchange rate for 4 January 1993–31 December 2002, with sample autocorrelation functions for (b) the log-returns, (c) their absolute values, and (d) their squares.

from the GARCH(1,1) fitted model; note that the values of $\{X_t/\sqrt{\tilde{\sigma}_t^2(\hat{\boldsymbol{\theta}}_R)}\}$ resemble $\{X_t/\sqrt{\sigma_t^2/\alpha_{00}}\} = \{\sqrt{\alpha_{00}}Z_t\}$ when $\hat{\boldsymbol{\theta}}_R$ is close to the true parameter vector $\boldsymbol{\theta}_0$. The corresponding R-estimates for the GARCH(1,1) parameter values are $\hat{\theta}_1 = \widehat{\alpha_1/\alpha_0} = 45{,}274$ and $\hat{\theta}_2 = \hat{\beta}_1 = 0.9388$, with approximate 95% confidence intervals $(20{,}793, 69{,}755)$ and $(0.9170, 0.9606)$. Higher order GARCH(2,1) and GARCH(1,2) models were also considered, but low values of the Wald test statistic led us to fail to reject the null hypotheses $H_0 : \alpha_{02}/\alpha_{00} = 0$ and $H_0 : \beta_{02} = 0$ at the 0.05 level of significance. This analysis therefore indicates that a GARCH(1,1) model is suitable for the exchange rate log-returns.

We then considered using ML to fit a GARCH(1,1) model to the data and obtain individual estimates for α_{00} and α_{01}. A kernel estimate of the density for the standardized GARCH(1,1) residuals, that is, $\{X_t/\sqrt{\tilde{\sigma}_t^2(\hat{\boldsymbol{\theta}}_R)}\}$ standardized to have variance one, is given in Figure 4.3a, along with the $N(0,1)$

https://doi.org/10.1017/S0266466612000047 Published online by Cambridge University Press

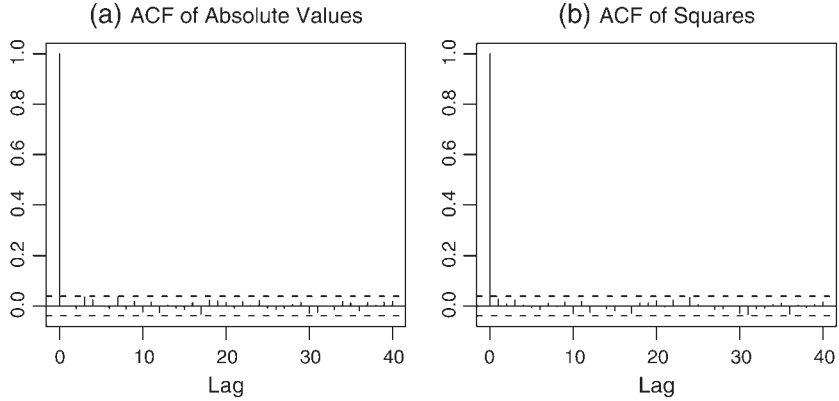


FIGURE 4.2. Sample autocorrelation functions for the (a) absolute values and (b) squares of the GARCH(1,1) residuals obtained via R-estimation.

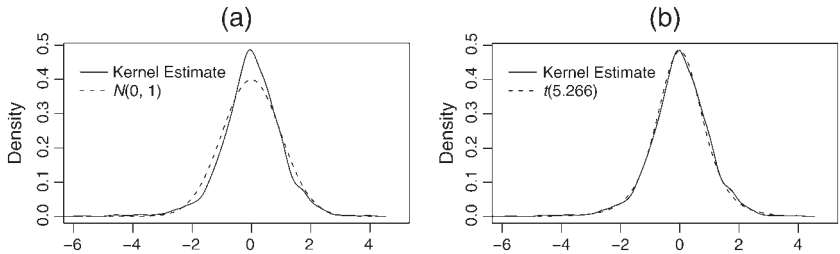


FIGURE 4.3. Kernel estimate of the density for the standardized GARCH(1,1) residuals with (a) the $N(0,1)$ density function and (b) the rescaled $t_{5.266}$ density function.

density. Because the distribution for the residuals appears roughly symmetric, but more peaked and heavier-tailed than Gaussian, we considered modeling the GARCH(1,1) noise distribution as rescaled student's t. The ML estimate of degrees of freedom is 5.266, and in Figure 4.3b it can be seen that the kernel density estimate for the standardized residuals and the $t_{5.266}$ density are close, so it appears reasonable to model the log-returns series as GARCH(1,1) with i.i.d. rescaled $t_{5.266}$ noise. Corresponding ML estimates of the model parameters are $\hat{\alpha}_0 = 9.632 \times 10^{-7}$, $\hat{\alpha}_1 = 0.04356$, and $\hat{\beta}_1 = 0.9393$, and, using the theory of Berkes and Horváth (2004), approximate 95% confidence intervals for the parameter values are $(4.310 \times 10^{-7}, 14.953 \times 10^{-7})$, $(0.02663, 0.06050)$, and $(0.9176, 0.9610)$. Note that the ML estimate for β_{01} and the corresponding confidence interval are nearly the same as those obtained via R-estimation, even though no specific distributional information was used for R-estimation. Also, $\hat{\alpha}_1/\hat{\alpha}_0 = 45{,}231$, which is quite close to the R-estimate of $\theta_{01} = \alpha_{01}/\alpha_{00}$.

https://doi.org/10.1017/S0266466612000047 Published online by Cambridge University Press

Finally, to verify that the rescaled $t_{5.266}$ distribution is suitable for the fitted GARCH(1,1) noise process, we used the ML residuals and the Kolmogorov–Smirnov test described in Koul and Ling (2006) to test this null hypothesis. The test statistic K_n equals 1.194, and via simulation we found a corresponding p-value of 0.409. We therefore failed to reject H_0 (the noise distribution is $t_{5.266}$) at the 0.05 level of significance, and so this test result indicates that the rescaled $t_{5.266}$ distribution is appropriate.

REFERENCES

Abramson, A. & I. Cohen (2008) Single-sensor audio source separation using classification and estimation approach and GARCH modeling. *IEEE Transactions on Audio, Speech and Language Processing* 16, 1528–1540.

Andrews, B. (2008) Rank-based estimation for autoregressive moving average time series models. *Journal of Time Series Analysis* 29, 51–73.

Andrews, B., R.A. Davis, & F.J. Breidt (2007) Rank-based estimation for all-pass time series models. *Annals of Statistics* 35, 844–869.

Andrews, D.W.K. (1999) Estimation when a parameter is on a boundary. *Econometrica* 67, 1341–1383.

Andrews, M.E. (2003) *Parameter estimation for all-pass time series models*. Ph.D. Dissertation, Colorado State University.

Berkes, I. & L. Horváth (2004) The efficiency of the estimators of the parameters in GARCH processes. *Annals of Statistics* 32, 633–655.

Berkes, I., L. Horváth, & P. Kokoszka (2003) GARCH processes: Structure and estimation. *Bernoulli* 9, 201–227.

Billingsley, P. (1961) The Lindeberg-Lévy theorem for martingales. *Proceedings of the American Mathematical Society* 12, 788–792.

Billingsley, P. (1999) *Convergence of Probability Measures*, 2nd ed. Wiley.

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

Bollerslev, T. (1987) A conditionally heteroskedastic time series model for speculative prices and rates of return. *Review of Economics and Statistics* 69, 542–547.

Bougerol, P. & N. Picard (1992) Stationarity of GARCH processes and of some nonnegative time series. *Journal of Econometrics* 52, 115–127.

Brockwell, P.J. & R.A. Davis (1991) *Time Series: Theory and Methods*, 2nd ed. Springer-Verlag.

Campbell, S.D. & F.X. Diebold (2005) Weather forecasting for weather derivatives. *Journal of the American Statistical Association* 100, 6–16.

Chernoff, H. & I.R. Savage (1958) Asymptotic normality and efficiency of certain nonparametric test statistics. *Annals of Mathematical Statistics* 29, 972–994.

Davis, R.A., K. Knight, & J. Liu (1992) M-estimation for autoregressions with infinite variance. *Stochastic Processes and Their Applications* 40, 145–180.

Engle, R.F. (1982) Autoregressive conditional heteroscedasticity with estimates of the variance of United Kingdom inflation. *Econometrica* 50, 987–1008.

Engle, R.F. & G. González-Rivera (1991) Semiparametric ARCH models. *Journal of Business & Economic Statistics* 9, 345–359.

Ewing, B.T., J.B. Kruse, & M.A. Thompson (2008) Analysis of time-varying turbulence in geographically-dispersed wind energy markets. *Energy Sources, Part B: Economics, Planning, and Policy* 3, 340–347.

Fan, J. & Q. Yao (2003) *Nonlinear Time Series: Nonparametric and Parametric Methods*. Springer-Verlag.

https://doi.org/10.1017/S0266466612000047 Published online by Cambridge University Press

Francq, C. & J.M. Zakoïan (2004) Maximum likelihood estimation of pure GARCH and ARMA-GARCH processes. *Bernoulli* 10, 605–637.

Francq, C. & J.M. Zakoïan (2007) Quasi-maximum likelihood estimation in GARCH processes when some coefficients are equal to zero. *Stochastic Processes and Their Applications* 117, 1265–1284.

Francq, C. & J.M. Zakoïan (2009) Testing the nullity of GARCH coefficients: Correction of the standard test and relative efficiency comparisons. *Journal of the American Statistical Association* 104, 313–324.

Gastwirth, J.L. & S.S. Wolff (1968) An elementary method for obtaining lower bounds on the asymptotic power of rank tests. *Annals of Mathematical Statistics* 39, 2128–2130.

Hall, P. & Q. Yao (2003) Inference in ARCH and GARCH models with heavy-tailed errors. *Econometrica* 71, 285–317.

Hettmansperger, T.P. & J.W. McKean (1998) *Robust Nonparametric Statistical Methods*. Arnold.

Hoti, S., M. McAleer, & F. Chan (2005) Modelling the spillover effects in the volatility of atmospheric carbon dioxide concentrations. *Mathematics and Computers in Simulation* 69, 46–56.

Huang, H.-H., Y.-M. Shiu, & P.-S. Lin (2008) HDD and CDD option pricing with market price of weather risk for Taiwan. *Journal of Futures Markets* 28, 790–814.

Jaeckel, L.A. (1972) Estimating regression coefficients by minimizing the dispersion of the residuals. *Annals of Mathematical Statistics* 43, 1449–1458.

Jurečková, J. & P.K. Sen (1996) *Robust Statistical Procedures: Asymptotics and Interrelations*. Wiley.

Koul, H.L. (2002) *Weighted Empirical Processes in Dynamic Nonlinear Models*, 2nd ed. Springer-Verlag.

Koul, H.L. & S. Ling (2006) Fitting an error distribution in some heteroscedastic time series models. *Annals of Statistics* 34, 994–1012.

Mukherjee, K. (2007) Generalized R-estimators under conditional heteroscedasticity. *Journal of Econometrics* 141, 383–415.

Mukherjee, K. (2008) M-estimation in GARCH models. *Econometric Theory* 24, 1530–1553.

Muler, N. & V.J. Yohai (2008) Robust estimates for GARCH models. *Journal of Statistical Planning and Inference* 138, 2918–2940.

Nelder, J.A. & R. Mead (1965) A simplex method for function minimization. *Computer Journal* 7, 308–313.

Nelson, D. (1990) Stationarity and persistence in the GARCH(1,1) model. *Econometric Theory* 6, 318–334.

Peng, L. & Q. Yao (2003) Least absolute deviations estimation for ARCH and GARCH models. *Biometrika* 90, 967–975.

Rao, C.R. (1973) *Linear Statistical Inference and Its Applications*, 2nd ed. Wiley.

Romilly, P. (2005) Time series modelling of global mean temperature for managerial decision-making. *Journal of Environmental Management* 76, 61–70.

Shephard, N. (1996) Statistical aspects of ARCH and stochastic volatility. In D.R. Cox, D.V. Hinkley, & O.E. Barndorff-Nielsen (eds.), *Time Series Models In Econometrics, Finance, and Other Fields*, pp. 1–67. Chapman & Hall.

Silverman, B.W. (1978) Weak and strong uniform consistency of the kernel estimate of a density and its derivatives. *Annals of Statistics* 6, 177–184.

Silverman, B.W. (1986) *Density Estimation for Statistics and Data Analysis*. Chapman & Hall.

Straumann, D. (2005) *Estimation in Conditionally Heteroscedastic Time Series Models*. Springer-Verlag.

APPENDIX

This Appendix contains proofs of the lemmas used to establish Theorems 3.1 and 3.2. Assume conditions A1–A7 and either A8, A9, or A10 hold throughout. To begin, note that

https://doi.org/10.1017/S0266466612000047 Published online by Cambridge University Press

following (2.3)–(2.4), the first partial derivatives for the residuals $\{\epsilon_t(\boldsymbol{\theta})\}_{t=1}^n$ are given by $\partial\epsilon_t(\boldsymbol{\theta})/\partial\theta_i = -[\partial\tilde{\sigma}_t^2(\boldsymbol{\theta})/\partial\theta_i]/\tilde{\sigma}_t^2(\boldsymbol{\theta})$, $i \in \{1, \ldots, p+q\}$, with

$$\frac{\partial\tilde{\sigma}_t^2(\boldsymbol{\theta})}{\partial\theta_i} = \begin{cases} 0, & t \le p, \\ X_{t-i}^2 + \sum_{j=1}^q \theta_{p+j}\left(\partial\tilde{\sigma}_{t-j}^2(\boldsymbol{\theta})/\partial\theta_i\right), & t \in \{p+1, \ldots, n\},\ i \in \{1, \ldots, p\}, \\ \tilde{\sigma}_{t+p-i}^2(\boldsymbol{\theta}) + \sum_{j=1}^q \theta_{p+j}\left(\partial\tilde{\sigma}_{t-j}^2(\boldsymbol{\theta})/\partial\theta_i\right), & t \in \{p+1, \ldots, n\},\ i \in \{p+1, \ldots, p+q\}. \end{cases}$$

And we let

$$\overline{\frac{\partial\epsilon(\boldsymbol{\theta})}{\partial\boldsymbol{\theta}}} = \frac{1}{n-p}\sum_{t=p+1}^n \frac{\partial\epsilon_t(\boldsymbol{\theta})}{\partial\boldsymbol{\theta}} \quad \text{and} \quad \overline{\frac{\partial\epsilon^*(\boldsymbol{\theta})}{\partial\boldsymbol{\theta}}} = \frac{1}{n-p}\sum_{t=p+1}^n \frac{\partial\epsilon_t^*(\boldsymbol{\theta})}{\partial\boldsymbol{\theta}}$$

(recall that the partial derivatives $\partial\epsilon_t^*(\boldsymbol{\theta})/\partial\theta_i$, $i \in \{1, \ldots, p+q\}$, are given in (3.1)).

LEMMA A.1. *As* $n \to \infty$,

$$\frac{1}{\sqrt{n}}\sum_{t=p+1}^n \lambda(F_\epsilon(\epsilon_t))\left[\frac{\partial\epsilon_t(\boldsymbol{\theta}_0)}{\partial\boldsymbol{\theta}} - \overline{\frac{\partial\epsilon(\boldsymbol{\theta}_0)}{\partial\boldsymbol{\theta}}}\right] \xrightarrow{d} \mathbf{N} \sim N(\mathbf{0}, \tilde{J}\tilde{\boldsymbol{\Gamma}}),$$

where $\epsilon_t = \ln\left(\alpha_{00}Z_t^2\right)$ *and* F_ϵ *represents the distribution function for* ϵ_t.

Proof. By the proof of Lemma 12 in Francq and Zakoïan (2007), $n^{-1/2} \times \sum_{t=p+1}^n |\partial\epsilon_t(\boldsymbol{\theta}_0)/\partial\theta_i - \partial\epsilon_t^*(\boldsymbol{\theta}_0)/\partial\theta_i| \xrightarrow{P} 0$ for any $i \in \{1, \ldots, p+q\}$. It therefore suffices to show that

$$\begin{aligned} &\frac{1}{\sqrt{n}}\sum_{t=p+1}^n \lambda(F_\epsilon(\epsilon_t))\left[\frac{\partial\epsilon_t^*(\boldsymbol{\theta}_0)}{\partial\boldsymbol{\theta}} - \overline{\frac{\partial\epsilon^*(\boldsymbol{\theta}_0)}{\partial\boldsymbol{\theta}}}\right] \\ &\quad= \frac{1}{\sqrt{n}}\sum_{t=p+1}^n [\lambda(F_\epsilon(\epsilon_t)) - \mathbb{E}\{\lambda(F_\epsilon(\epsilon_t))\}]\left[\frac{\partial\epsilon_t^*(\boldsymbol{\theta}_0)}{\partial\boldsymbol{\theta}} - \overline{\frac{\partial\epsilon^*(\boldsymbol{\theta}_0)}{\partial\boldsymbol{\theta}}}\right] \\ &\quad= \frac{1}{\sqrt{n}}\sum_{t=p+1}^n [\lambda(F_\epsilon(\epsilon_t)) - \mathbb{E}\{\lambda(F_\epsilon(\epsilon_t))\}]\left[\frac{\partial\epsilon_t^*(\boldsymbol{\theta}_0)}{\partial\boldsymbol{\theta}} - \mathbb{E}\left\{\frac{\partial\epsilon_t^*(\boldsymbol{\theta}_0)}{\partial\boldsymbol{\theta}}\right\}\right] \\ &\qquad- \left(\frac{1}{\sqrt{n}}\sum_{t=p+1}^n [\lambda(F_\epsilon(\epsilon_t)) - \mathbb{E}\{\lambda(F_\epsilon(\epsilon_t))\}]\right)\left(\overline{\frac{\partial\epsilon^*(\boldsymbol{\theta}_0)}{\partial\boldsymbol{\theta}}} - \mathbb{E}\left\{\frac{\partial\epsilon_t^*(\boldsymbol{\theta}_0)}{\partial\boldsymbol{\theta}}\right\}\right) \end{aligned} \tag{A.1}$$

converges in distribution to $\mathbf{N}$. By the central limit theorem, $n^{-1/2}\sum_{t=p+1}^n [\lambda(F_\epsilon(\epsilon_t)) - \mathbb{E}\{\lambda(F_\epsilon(\epsilon_t))\}] \xrightarrow{d} N(0, \tilde{J})$ and, because $\{\partial\epsilon_t^*(\boldsymbol{\theta}_0)/\partial\boldsymbol{\theta}\}$ is stationary ergodic with $\mathbb{E}|\partial\epsilon_t^*(\boldsymbol{\theta}_0)/\partial\theta_i| < \infty\ \forall i$ (Francq and Zakoïan, 2007), $\overline{\partial\epsilon^*(\boldsymbol{\theta}_0)/\partial\boldsymbol{\theta}} \xrightarrow{P} \mathbb{E}\{\partial\epsilon_t^*(\boldsymbol{\theta}_0)/\partial\boldsymbol{\theta}\}$. Hence, (A.1) equals

$$\frac{1}{\sqrt{n}}\sum_{t=p+1}^n [\lambda(F_\epsilon(\epsilon_t)) - \mathbb{E}\{\lambda(F_\epsilon(\epsilon_t))\}]\left[\frac{\partial\epsilon_t^*(\boldsymbol{\theta}_0)}{\partial\boldsymbol{\theta}} - \mathbb{E}\left\{\frac{\partial\epsilon_t^*(\boldsymbol{\theta}_0)}{\partial\boldsymbol{\theta}}\right\}\right] + o_p(1). \tag{A.2}$$

https://doi.org/10.1017/S0266466612000047 Published online by Cambridge University Press

Equation (A.2) converges in distribution to $\mathbf{N}$ by the central limit theorem for martingale differences (Billingsley, 1961). ■

LEMMA A.2. *For any $T \in (0, \infty)$, as $n \to \infty$,*

$$\sup_{\mathbf{u}\in\Lambda,\, \|\mathbf{u}\|\le T} \left| \frac{1}{\sqrt{n}} \sum_{t=p+1}^{n} \mathbf{u}' \left[\lambda \left(\frac{R_t\left(\boldsymbol{\theta}_0 + n^{-1/2}\mathbf{u}\right)}{n-p+1} \right) - \lambda(F_\epsilon(\epsilon_t)) \right] \right.$$

$$\left. \times \left[\frac{\partial \epsilon_t(\boldsymbol{\theta}_0)}{\partial \boldsymbol{\theta}} - \overline{\frac{\partial \epsilon(\boldsymbol{\theta}_0)}{\partial \boldsymbol{\theta}}} \right] - \tilde{K}\mathbf{u}'\boldsymbol{\Gamma}\mathbf{u} \right| \tag{A.3}$$

is $o_p(1)$. (Recall from Theorem 3.1 that $\Lambda = \Lambda_1 \times \cdots \times \Lambda_{p+q}$, with $\Lambda_i = \mathbb{R}$ if $\theta_{0i} > 0$ and $\Lambda_i = [0, \infty)$ if $\theta_{0i} = 0$.)

Proof. If $R_{t,n}(\mathbf{u}) := R_t(\boldsymbol{\theta}_0 + n^{-1/2}\mathbf{u})$ and $\mathbf{V}^*_{t,n} := \partial\epsilon_t(\boldsymbol{\theta}_0)/\partial\boldsymbol{\theta} - \overline{\partial\epsilon(\boldsymbol{\theta}_0)/\partial\boldsymbol{\theta}}$, then (A.3) can be expressed as

$$\sup_{\mathbf{u}\in\Lambda,\, \|\mathbf{u}\|\le T} \left| \frac{1}{\sqrt{n}} \sum_{t=p+1}^{n} \mathbf{u}' \left[\lambda \left(\frac{R_{t,n}(\mathbf{u})}{n-p+1} \right) - \lambda(F_\epsilon(\epsilon_t)) \right] \mathbf{V}^*_{t,n} - \tilde{K}\mathbf{u}'\boldsymbol{\Gamma}\mathbf{u} \right|.$$

Because the weight function λ is left-continuous and $\tilde{K} = \int_{-\infty}^{\infty} f_\epsilon(x)\, d\lambda(F_\epsilon(x)) = \int_0^1 f_\epsilon(F_\epsilon^{-1}(y))\, d\lambda(y)$, (A.3) equals

$$\sup_{\mathbf{u}\in\Lambda,\, \|\mathbf{u}\|\le T} \left| \int_0^1 \left(\frac{1}{\sqrt{n}} \sum_{t=p+1}^{n} \mathbf{u}' \left[I\left\{ \frac{R_{t,n}(\mathbf{u})}{n-p+1} \le y \right\} - I\{F_\epsilon(\epsilon_t) \le y\} \right] \right.$$

$$\left. \times \mathbf{V}^*_{t,n} + f_\epsilon(F_\epsilon^{-1}(y))\mathbf{u}'\boldsymbol{\Gamma}\mathbf{u} \right) d\lambda(y) \Bigg|,$$

which is bounded above by

$$\sup_{\mathbf{u}\in\Lambda,\, \|\mathbf{u}\|\le T,\, y\in(0,1)} \left| \frac{1}{\sqrt{n}} \sum_{t=p+1}^{n} \mathbf{u}' \left[I\left\{ \frac{R_{t,n}(\mathbf{u})}{n-p+1} \le y \right\} - I\{F_\epsilon(\epsilon_t) \le y\} \right] \right.$$

$$\left. \times \mathbf{V}^*_{t,n} + f_\epsilon(F_\epsilon^{-1}(y))\mathbf{u}'\boldsymbol{\Gamma}\mathbf{u} \right| \tag{A.4}$$

$$\times [\lambda(1) - \lambda(0)]. \tag{A.5}$$

Equation (A.4) is $o_p(1)$; a related result is obtained in B. Andrews (2008, proof of Lem. A.5) in the case of rank-based estimation for autoregressive moving average models and a similar proof can be used here, so we omit the details. Because λ is bounded, the proof of this lemma is complete. ■

Next consider the mixed partial derivatives of $\{\epsilon_t(\boldsymbol{\theta})\}_{t=1}^n$:

$$\frac{\partial^2 \epsilon_t(\boldsymbol{\theta})}{\partial\theta_i \partial\theta_j} = \frac{1}{\tilde{\sigma}_t^4(\boldsymbol{\theta})} \left(\frac{\partial\tilde{\sigma}_t^2(\boldsymbol{\theta})}{\partial\theta_i} \frac{\partial\tilde{\sigma}_t^2(\boldsymbol{\theta})}{\partial\theta_j} - \tilde{\sigma}_t^2(\boldsymbol{\theta}) \frac{\partial^2\tilde{\sigma}_t^2(\boldsymbol{\theta})}{\partial\theta_i \partial\theta_j} \right), \qquad i, j \in \{1, \ldots, p+q\},$$

https://doi.org/10.1017/S0266466612000047 Published online by Cambridge University Press

with

$$\frac{\partial^2 \tilde{\sigma}_t^2(\boldsymbol{\theta})}{\partial\theta_i \partial\theta_j}$$

$$= \begin{cases} 0, & i, j \in \{1, \ldots, p\}, \\ \frac{\partial \tilde{\sigma}_{t+p-j}^2(\boldsymbol{\theta})}{\partial \theta_i} + \sum_{k=1}^{q} \theta_{p+k} \frac{\partial^2 \tilde{\sigma}_{t-k}^2(\boldsymbol{\theta})}{\partial\theta_i \partial\theta_j}, & i \in \{1, \ldots, p\},\ j \in \{p+1, \ldots, p+q\}, \\ \frac{\partial \tilde{\sigma}_{t+p-j}^2(\boldsymbol{\theta})}{\partial \theta_i} + \frac{\partial \tilde{\sigma}_{t+p-i}^2(\boldsymbol{\theta})}{\partial \theta_j} + \sum_{k=1}^{q} \theta_{p+k} \frac{\partial^2 \tilde{\sigma}_{t-k}^2(\boldsymbol{\theta})}{\partial\theta_i \partial\theta_j}, & i, j \in \{p+1, \ldots, p+q\}, \end{cases}$$

for $t \in \{p+1, \ldots, n\}$, and $\partial^2 \tilde{\sigma}_t^2(\boldsymbol{\theta})/(\partial\theta_i \partial\theta_j) = 0$ for $t \le p$. Let

$$\overline{\frac{\partial^2 \epsilon(\boldsymbol{\theta})}{\partial\boldsymbol{\theta}\partial\boldsymbol{\theta}'}} = \frac{1}{n-p} \sum_{t=p+1}^{n} \frac{\partial^2 \epsilon_t(\boldsymbol{\theta})}{\partial\boldsymbol{\theta}\partial\boldsymbol{\theta}'}.$$

LEMMA A.3. *For any $T \in (0, \infty)$, as $n \to \infty$,*

$$\sup_{\mathbf{u},\mathbf{v}\in\Lambda,\, \|\mathbf{u}\|,\|\mathbf{v}\|\le T} \left| \frac{1}{n} \sum_{t=p+1}^{n} \mathbf{u}' \lambda \left(\frac{R_t\left(\boldsymbol{\theta}_0 + n^{-1/2}\mathbf{u}\right)}{n-p+1} \right) \times \left[\frac{\partial^2 \epsilon_t(\boldsymbol{\theta}_0 + n^{-1/2}\mathbf{v})}{\partial\boldsymbol{\theta}\partial\boldsymbol{\theta}'} - \overline{\frac{\partial^2 \epsilon(\boldsymbol{\theta}_0 + n^{-1/2}\mathbf{v})}{\partial\boldsymbol{\theta}\partial\boldsymbol{\theta}'}} \right] \mathbf{u} \right| \xrightarrow{P} 0.$$

Proof. For any $i, j \in \{1, \ldots, p+q\}$, the sequence $\{\partial^2 \epsilon_t^*(\boldsymbol{\theta}_0)/(\partial\theta_i \partial\theta_j)\}$ of mixed partial derivatives of $\{\epsilon_t^*(\boldsymbol{\theta})\}$ at $\boldsymbol{\theta} = \boldsymbol{\theta}_0$ is stationary, ergodic with $\mathbb{E}|\partial^2 \epsilon_t^*(\boldsymbol{\theta}_0)/(\partial\theta_i \partial\theta_j)| < \infty$ and $n^{-1} \sum_{t=p+1}^{n} |\partial^2 \epsilon_t(\boldsymbol{\theta}_0)/(\partial\theta_i \partial\theta_j) - \partial^2 \epsilon_t^*(\boldsymbol{\theta}_0)/(\partial\theta_i \partial\theta_j)| \xrightarrow{P} 0$ (Francq and Zakoïan, 2007). Following Lemma A.1, it can therefore be shown that

$$\sup_{\mathbf{u}\in\Lambda,\, \|\mathbf{u}\|\le T} \left| \frac{1}{n} \sum_{t=p+1}^{n} \mathbf{u}' \lambda(F_\epsilon(\epsilon_t)) \left[\frac{\partial^2 \epsilon_t(\boldsymbol{\theta}_0)}{\partial\boldsymbol{\theta}\partial\boldsymbol{\theta}'} - \overline{\frac{\partial^2 \epsilon(\boldsymbol{\theta}_0)}{\partial\boldsymbol{\theta}\partial\boldsymbol{\theta}'}} \right] \mathbf{u} \right| \xrightarrow{P} 0,$$

and following Lemma A.2,

$$\sup_{\mathbf{u}\in\Lambda,\, \|\mathbf{u}\|\le T} \left| \frac{1}{n} \sum_{t=p+1}^{n} \mathbf{u}' \left[\lambda \left(\frac{R_t\left(\boldsymbol{\theta}_0 + n^{-1/2}\mathbf{u}\right)}{n-p+1} \right) - \lambda(F_\epsilon(\epsilon_t)) \right] \times \left[\frac{\partial^2 \epsilon_t(\boldsymbol{\theta}_0)}{\partial\boldsymbol{\theta}\partial\boldsymbol{\theta}'} - \overline{\frac{\partial^2 \epsilon(\boldsymbol{\theta}_0)}{\partial\boldsymbol{\theta}\partial\boldsymbol{\theta}'}} \right] \mathbf{u} \right| \xrightarrow{P} 0.$$

Because

$$\sup_{\mathbf{v}\in\Lambda,\, \|\mathbf{v}\|\le T} \frac{1}{n} \sum_{t=p+1}^{n} \left| \frac{\partial^2 \epsilon_t(\boldsymbol{\theta}_0 + n^{-1/2}\mathbf{v})}{\partial\theta_i \partial\theta_j} - \frac{\partial^2 \epsilon_t(\boldsymbol{\theta}_0)}{\partial\theta_i \partial\theta_j} \right| \xrightarrow{P} 0 \quad \forall i, j \in \{1, \ldots, p+q\}$$

(see Francq and Zakoïan, 2007, proof of Lem. 11), the proof is complete. ■

https://doi.org/10.1017/S0266466612000047 Published online by Cambridge University Press

Now, for $\mathbf{u} \in \Lambda$ and $\delta_1, \delta_2 \in [0,1]$, let

$$\tilde{U}_n(\mathbf{u},\delta_1,\delta_2) = \sum_{t=p+1}^{n} \lambda\left(\frac{R_t\left(\boldsymbol{\theta}_0 + n^{-1/2}\delta_1\mathbf{u}\right)}{n-p+1}\right)\left[\epsilon_t\left(\boldsymbol{\theta}_0 + \frac{\delta_2\mathbf{u}}{\sqrt{n}}\right) - \overline{\epsilon\left(\boldsymbol{\theta}_0 + \frac{\delta_2\mathbf{u}}{\sqrt{n}}\right)}\right]$$
$$- \sum_{t=p+1}^{n} \lambda\left(\frac{R_t\left(\boldsymbol{\theta}_0 + n^{-1/2}\delta_1\mathbf{u}\right)}{n-p+1}\right)\left[\epsilon_t\left(\boldsymbol{\theta}_0 + \frac{\delta_1\mathbf{u}}{\sqrt{n}}\right) - \overline{\epsilon\left(\boldsymbol{\theta}_0 + \frac{\delta_1\mathbf{u}}{\sqrt{n}}\right)}\right]$$

and

$$\tilde{V}_n(\mathbf{u},\delta_1,\delta_2) = \sum_{t=p+1}^{n} \lambda\left(\frac{R_t\left(\boldsymbol{\theta}_0 + n^{-1/2}\delta_2\mathbf{u}\right)}{n-p+1}\right)\left[\epsilon_t\left(\boldsymbol{\theta}_0 + \frac{\delta_2\mathbf{u}}{\sqrt{n}}\right) - \overline{\epsilon\left(\boldsymbol{\theta}_0 + \frac{\delta_2\mathbf{u}}{\sqrt{n}}\right)}\right]$$
$$- \sum_{t=p+1}^{n} \lambda\left(\frac{R_t\left(\boldsymbol{\theta}_0 + n^{-1/2}\delta_2\mathbf{u}\right)}{n-p+1}\right)\left[\epsilon_t\left(\boldsymbol{\theta}_0 + \frac{\delta_1\mathbf{u}}{\sqrt{n}}\right) - \overline{\epsilon\left(\boldsymbol{\theta}_0 + \frac{\delta_1\mathbf{u}}{\sqrt{n}}\right)}\right].$$

Using Taylor series expansions,

$$\begin{aligned}
&\tilde{U}_n(\mathbf{u},\delta_1,\delta_2)\\
&\quad= \sum_{t=p+1}^{n} \lambda\left(\frac{R_t\left(\boldsymbol{\theta}_0 + n^{-1/2}\delta_1\mathbf{u}\right)}{n-p+1}\right)\left[\epsilon_t\left(\boldsymbol{\theta}_0 + \frac{\delta_2\mathbf{u}}{\sqrt{n}}\right) - \overline{\epsilon\left(\boldsymbol{\theta}_0 + \frac{\delta_2\mathbf{u}}{\sqrt{n}}\right)} - \epsilon_t(\boldsymbol{\theta}_0) + \overline{\epsilon(\boldsymbol{\theta}_0)}\right]\\
&\qquad- \sum_{t=p+1}^{n} \lambda\left(\frac{R_t\left(\boldsymbol{\theta}_0 + n^{-1/2}\delta_1\mathbf{u}\right)}{n-p+1}\right)\left[\epsilon_t\left(\boldsymbol{\theta}_0 + \frac{\delta_1\mathbf{u}}{\sqrt{n}}\right) - \overline{\epsilon\left(\boldsymbol{\theta}_0 + \frac{\delta_1\mathbf{u}}{\sqrt{n}}\right)} - \epsilon_t(\boldsymbol{\theta}_0) + \overline{\epsilon(\boldsymbol{\theta}_0)}\right]\\
&\quad= \frac{\delta_2-\delta_1}{\sqrt{n}} \sum_{t=p+1}^{n} \mathbf{u}'\lambda\left(\frac{R_t\left(\boldsymbol{\theta}_0 + n^{-1/2}\delta_1\mathbf{u}\right)}{n-p+1}\right)\left[\frac{\partial\epsilon_t(\boldsymbol{\theta}_0)}{\partial\boldsymbol{\theta}} - \overline{\frac{\partial\epsilon(\boldsymbol{\theta}_0)}{\partial\boldsymbol{\theta}}}\right]\\
&\qquad+ \frac{\delta_2^2}{2n} \sum_{t=p+1}^{n} \mathbf{u}'\lambda\left(\frac{R_t\left(\boldsymbol{\theta}_0 + n^{-1/2}\delta_1\mathbf{u}\right)}{n-p+1}\right)\left[\frac{\partial^2\epsilon_t(\boldsymbol{\theta}_n^*(\mathbf{u},\delta_1,\delta_2))}{\partial\boldsymbol{\theta}\partial\boldsymbol{\theta}'} - \overline{\frac{\partial^2\epsilon(\boldsymbol{\theta}_n^*(\mathbf{u},\delta_1,\delta_2))}{\partial\boldsymbol{\theta}\partial\boldsymbol{\theta}'}}\right]\mathbf{u}\\
&\qquad- \frac{\delta_1^2}{2n} \sum_{t=p+1}^{n} \mathbf{u}'\lambda\left(\frac{R_t\left(\boldsymbol{\theta}_0 + n^{-1/2}\delta_1\mathbf{u}\right)}{n-p+1}\right)\left[\frac{\partial^2\epsilon_t(\boldsymbol{\theta}_n^*(\mathbf{u},\delta_1,\delta_1))}{\partial\boldsymbol{\theta}\partial\boldsymbol{\theta}'} - \overline{\frac{\partial^2\epsilon(\boldsymbol{\theta}_n^*(\mathbf{u},\delta_1,\delta_1))}{\partial\boldsymbol{\theta}\partial\boldsymbol{\theta}'}}\right]\mathbf{u},
\end{aligned} \tag{A.6}$$

and, similarly,

$$\begin{aligned}
&\tilde{V}_n(\mathbf{u},\delta_1,\delta_2)\\
&\quad= \frac{\delta_2-\delta_1}{\sqrt{n}} \sum_{t=p+1}^{n} \mathbf{u}'\lambda\left(\frac{R_t\left(\boldsymbol{\theta}_0 + n^{-1/2}\delta_2\mathbf{u}\right)}{n-p+1}\right)\left[\frac{\partial\epsilon_t(\boldsymbol{\theta}_0)}{\partial\boldsymbol{\theta}} - \overline{\frac{\partial\epsilon(\boldsymbol{\theta}_0)}{\partial\boldsymbol{\theta}}}\right]\\
&\qquad+ \frac{\delta_2^2}{2n} \sum_{t=p+1}^{n} \mathbf{u}'\lambda\left(\frac{R_t\left(\boldsymbol{\theta}_0 + n^{-1/2}\delta_2\mathbf{u}\right)}{n-p+1}\right)\left[\frac{\partial^2\epsilon_t(\boldsymbol{\theta}_n^*(\mathbf{u},\delta_2,\delta_2))}{\partial\boldsymbol{\theta}\partial\boldsymbol{\theta}'} - \overline{\frac{\partial^2\epsilon(\boldsymbol{\theta}_n^*(\mathbf{u},\delta_2,\delta_2))}{\partial\boldsymbol{\theta}\partial\boldsymbol{\theta}'}}\right]\mathbf{u}\\
&\qquad- \frac{\delta_1^2}{2n} \sum_{t=p+1}^{n} \mathbf{u}'\lambda\left(\frac{R_t\left(\boldsymbol{\theta}_0 + n^{-1/2}\delta_2\mathbf{u}\right)}{n-p+1}\right)\left[\frac{\partial^2\epsilon_t(\boldsymbol{\theta}_n^*(\mathbf{u},\delta_2,\delta_1))}{\partial\boldsymbol{\theta}\partial\boldsymbol{\theta}'} - \overline{\frac{\partial^2\epsilon(\boldsymbol{\theta}_n^*(\mathbf{u},\delta_2,\delta_1))}{\partial\boldsymbol{\theta}\partial\boldsymbol{\theta}'}}\right]\mathbf{u},
\end{aligned} \tag{A.7}$$

where the values of $\boldsymbol{\theta}_n^*(\mathbf{u},\cdot,\cdot)$ lie between $\boldsymbol{\theta}_0$ and $\boldsymbol{\theta}_0 + n^{-1/2}\mathbf{u}$. The functions $\tilde{U}_n$ and $\tilde{V}_n$ are used in the proof of Lemma A.4.

https://doi.org/10.1017/S0266466612000047 Published online by Cambridge University Press

LEMMA A.4. *For* $\mathbf{u} \in \Lambda$, *let* $S_n(\mathbf{u}) = D_n(\boldsymbol{\theta}_0 + n^{-1/2}\mathbf{u}) - D_n(\boldsymbol{\theta}_0)$ *and* $S(\mathbf{u}) = \mathbf{u}'\mathbf{N} + \tilde{K}\mathbf{u}'\boldsymbol{\Gamma}\mathbf{u}/2$, *where* $\mathbf{N} \sim N(\mathbf{0}, \tilde{J}\boldsymbol{\Gamma})$. *Then* $S_n(\cdot) \xrightarrow{d} S(\cdot)$ *on* $C(\Lambda)$, *the space of continuous functions on* Λ *where convergence is equivalent to uniform convergence on every compact subset.*

Proof. Let $\mathbf{u} \in \Lambda$ and suppose m is any positive integer. Because

$$S_n(\mathbf{u}) = D_n\left(\boldsymbol{\theta}_0 + n^{-1/2}\mathbf{u}\right) - D_n(\boldsymbol{\theta}_0) = \sum_{k=1}^{m}\left[D_n\left(\boldsymbol{\theta}_0 + \frac{k\mathbf{u}}{m\sqrt{n}}\right) - D_n\left(\boldsymbol{\theta}_0 + \frac{(k-1)\mathbf{u}}{m\sqrt{n}}\right)\right],$$

we have

$$\sum_{k=1}^{m}\tilde{U}_n\left(\mathbf{u}, \frac{k-1}{m}, \frac{k}{m}\right) \le S_n(\mathbf{u}) \le \sum_{k=1}^{m}\tilde{V}_n\left(\mathbf{u}, \frac{k-1}{m}, \frac{k}{m}\right) \tag{A.8}$$

by Theorem 2.1. Using (A.6), (A.7), and Lemmas A.1, A.2, and A.3,

$$\begin{bmatrix} \tilde{U}_n\left(\mathbf{u}, 0, \frac{1}{m}\right) \\ \vdots \\ \tilde{U}_n\left(\mathbf{u}, \frac{m-1}{m}, 1\right) \\ \tilde{V}_n\left(\mathbf{u}, 0, \frac{1}{m}\right) \\ \vdots \\ \tilde{V}_n\left(\mathbf{u}, \frac{m-1}{m}, 1\right) \end{bmatrix} \xrightarrow{d} \begin{bmatrix} \frac{1}{m}\mathbf{u}'\mathbf{N} + \frac{0}{m^2}\tilde{K}\mathbf{u}'\boldsymbol{\Gamma}\mathbf{u} \\ \vdots \\ \frac{1}{m}\mathbf{u}'\mathbf{N} + \frac{m-1}{m^2}\tilde{K}\mathbf{u}'\boldsymbol{\Gamma}\mathbf{u} \\ \frac{1}{m}\mathbf{u}'\mathbf{N} + \frac{1}{m^2}\tilde{K}\mathbf{u}'\boldsymbol{\Gamma}\mathbf{u} \\ \vdots \\ \frac{1}{m}\mathbf{u}'\mathbf{N} + \frac{m}{m^2}\tilde{K}\mathbf{u}'\boldsymbol{\Gamma}\mathbf{u} \end{bmatrix}$$

on $\mathbb{R}^{2m}$. Hence,

$$\begin{bmatrix} \sum_{k=1}^{m}\tilde{U}_n\left(\mathbf{u}, \frac{k-1}{m}, \frac{k}{m}\right) \\ \sum_{k=1}^{m}\tilde{V}_n\left(\mathbf{u}, \frac{k-1}{m}, \frac{k}{m}\right) \end{bmatrix} \xrightarrow{d} \begin{bmatrix} \mathbf{u}'\mathbf{N} + \frac{m-1}{2m}\tilde{K}\mathbf{u}'\boldsymbol{\Gamma}\mathbf{u} \\ \mathbf{u}'\mathbf{N} + \frac{m+1}{2m}\tilde{K}\mathbf{u}'\boldsymbol{\Gamma}\mathbf{u} \end{bmatrix}$$

on $\mathbb{R}^2$. For any $\varepsilon > 0$, there exists an integer m large enough so that $\mathbf{u}'\mathbf{N} + (m-1)/(2m)\tilde{K}\mathbf{u}'\boldsymbol{\Gamma}\mathbf{u}$ and $\mathbf{u}'\mathbf{N} + (m+1)/(2m)\tilde{K}\mathbf{u}'\boldsymbol{\Gamma}\mathbf{u}$ are both in an ε-neighborhood of $S(\mathbf{u}) = \mathbf{u}'\mathbf{N} + \tilde{K}\mathbf{u}'\boldsymbol{\Gamma}\mathbf{u}/2$. Thus, for any $\mathbf{u} \in \Lambda$, $S_n(\mathbf{u}) \xrightarrow{d} S(\mathbf{u})$. It can be shown similarly that all finite-dimensional distributions of $S_n(\cdot)$ converge to those of $S(\cdot)$.

Using (A.8) and an argument in M.E. Andrews (2003, pp. 84–86), it can be shown that $\lim_{\delta\to 0^+}\limsup_{n\to\infty} P(\sup_{\mathbf{u},\mathbf{v}\in K,\, \|\mathbf{u}-\mathbf{v}\|\le\delta} |S_n(\mathbf{u}) - S_n(\mathbf{v})| > \eta) = 0$ for any $\eta > 0$ and any compact subset $K \subset \Lambda$. Therefore, $S_n(\cdot)$ is tight on $C(K)$ for any compact set $K \subset \Lambda$. It follows that $S_n(\cdot) \xrightarrow{d} S(\cdot)$ on $C(\Lambda)$ by Theorem 7.1 in Billingsley (1999). ■

LEMMA A.5. *Under the conditions of Theorem 3.2,* $\sup_{x\in\mathbb{R}} |\hat{f}_n(x) - f_\epsilon(x)| \xrightarrow{P} 0$ *as* $n \to \infty$.

Proof. If, for $x \in \mathbb{R}$, $f_n(x) := (b_n n)^{-1}\sum_{t=p+1}^{n}\kappa([\epsilon_t - x]/b_n)$, then $\sup_{x\in\mathbb{R}} |f_n(x) - f_\epsilon(x)| \xrightarrow{P} 0$ by Theorem A in Silverman (1978). Using a proof similar to that of Lemma 16 on page 88 of M.E. Andrews (2003), it can be shown that $\sup_{x\in\mathbb{R}} |\hat{f}_n(x) - f_n(x)| \xrightarrow{P} 0$, and so this lemma holds. ■

https://doi.org/10.1017/S0266466612000047 Published online by Cambridge University Press