Econometric Theory, **24**, 2008, 1530–1553. Printed in the United States of America.
doi:10.1017/S0266466608080602

M-ESTIMATION IN GARCH MODELS

KANCHAN MUKHERJEE
Lancaster University

This paper derives asymptotic normality of a class of *M*-estimators in the generalized autoregressive conditional heteroskedastic (GARCH) model. The class of estimators includes least absolute deviation and Huber's estimator in addition to the well-known quasi maximum likelihood estimator. For some estimators, the asymptotic normality results are obtained only under the existence of fractional unconditional moment assumption on the error distribution and some mild smoothness and moment assumptions on the score function.

1. INTRODUCTION

Volatility or the instantaneous variability of a financial time series is an important concept in many econometric models. In his seminal work, Engle (1982) modeled volatility of a series $\{X_t; 1 \leq t \leq n\}$ as a linear function of the squares of past observations and called it the autoregressive conditional heteroskedastic (ARCH) model. Later Bollerslev (1986) proposed a useful extension of the ARCH model called the generalized ARCH (GARCH) model where the volatility is a linear function of both the squares of past observations and past volatility. In the GARCH(p,q) model, where $p, q \geq 1$ are known integers, the following representation of the series $\{X_t; t \in \mathcal{Z}\}$ is assumed:

$$X_t = \sigma_t \epsilon_t, \tag{1.1}$$

where $\{\epsilon_t; t \in \mathcal{Z}\}$ are unobservable independent and identically distributed (i.i.d.) errors symmetric about zero and

$$\sigma_t^2 = \omega_0 + \sum_{i=1}^{p} \alpha_{0i} X_{t-i}^2 + \sum_{j=1}^{q} \beta_{0j} \sigma_{t-j}^2, \qquad t \in \mathcal{Z}, \tag{1.2}$$

with $\omega_0, \alpha_{0i}, \beta_{0j} > 0, \forall i, j$. In this paper, we are concerned with the problem of robust *M*-estimation of some function of the model parameter vector

$$\boldsymbol{\theta}_0 = [\omega_0, \alpha_{01}, \ldots, \alpha_{0p}, \beta_{01}, \ldots, \beta_{0q}]' \tag{1.3}$$

I thank Professor Peter Robinson for his valuable comments on an earlier version of this paper. I also thank Professor Bruce Hansen and the two anonymous referees for making many useful comments that have improved the paper to a great extent. I am grateful to Mr. Farhat Iqbal and Mr. Jack Ndebu for their help in doing extensive simulation study and computation of *M*-estimates. Address correspondence to Kanchan Mukherjee, Department of Mathematics and Statistics, Lancaster University, Lancaster LA1 4YF, United Kingdom; e-mail: k.mukherjee@lancaster.ac.uk.

 © 2008 Cambridge University Press 0266-4666/08 $15.00

https://doi.org/10.1017/S0266466608080602 Published online by Cambridge University Press

based on the observations $\{X_t; 1 \leq t \leq n\}$ and with deriving its asymptotic properties.

Gaussian likelihood is routinely used to estimate parameters of the GARCH model. Here the estimator is obtained as a maximizer of the logarithm of a standard Gaussian likelihood function of the errors, and the resulting estimator is called the quasi maximum likelihood estimator (QMLE). The asymptotic normality of the QMLE was established by Weiss (1986) for the ARCH model and by Lee and Hansen (1994) and Lumsdaine (1996) for the GARCH (1,1) model using its special structure and also the assumption on unit variance and the existence of unconditional moments of order at least 4 for the error distribution. Asymptotics of the QMLE were discussed by Jeantheau (1998) and Comte and Lieberman (2003) for the multivariate GARCH(p,q) model under strong assumptions on the error distribution and by Hall and Yao (2003) for the GARCH(p,q) model under heavy-tailed error distribution. Berkes, Horvath, and Kokoszka (2003) (hereafter BHK) derived many nice technical results on the GARCH model (1.1) and (1.2) and used them to derive the asymptotic normality of the QMLE. Straumann and Mikosch (2006) studied the QMLE under general conditional heteroskedastic models based on stochastic recurrence equations, and Robinson and Zaffaroni (2006) derived the asymptotic normality of the QMLE for the ARCH(∞) model, which includes GARCH as a very special case.

Several studies on financial data have suggested that the existence of the fourth moment needed for the asymptotic normality of the QMLE is not tenable quite often in practice, and the Gaussian likelihood is often not appropriate because of the large number of outliers in the variable. Engle and Gonzalez-Rivera (1991) advocated the need for alternative estimators that can improve on QMLE. Consequently, Newey and Steigerwald (1997) considered a parametric family of densities for the errors and used this family of densities to construct pseudo maximum likelihood estimators (PMLEs). Li and Turtle (2000) considered an estimating function approach to estimate the parameters of the ARCH model based on some conditional moments assumptions as an attempt to improve on the QMLE for non-Gaussian errors. Peng and Yao (2003) considered least absolute deviation (LAD)–type estimators of three different varieties, and Berkes and Horvath (2004) considered the PMLE for the GARCH(p,q) model where $p, q \geq 1$. Berkes and Horvath (2004) derived asymptotic normality of some of the PMLEs under the existence of a fractional unconditional moment of the error distribution when the score function is three times differentiable over $(0,\infty)$. Their class of estimators includes both LAD and QMLE and also some other important score functions. However, the identifiability condition of the parameters to be estimated stipulates that the value for the unconditional error (or function of error) moment such as $\mathrm{E}(\epsilon^2) = 1$ or $\mathrm{E}(|\epsilon|) = 1$ or $\mathrm{E}\{|\epsilon|/(1 + |\epsilon|)\}$ is known; see the general condition (1.16) and displays (2.1)–(2.3) of specific examples in Berkes and Horvath (2004). Clearly, such conditions are impossible to verify and hence

https://doi.org/10.1017/S0266466608080602 Published online by Cambridge University Press

are very undesirable. This motivates us to derive the asymptotics of the PMLE, or, more generally, of the M-estimators, without making such assumptions.

In both Peng and Yao (2003) and Berkes and Horvath (2004), it was implicit that a reparametrization of the underlying parameters is necessary to derive the asymptotics. In this paper we show that an M-estimator based on a score function H consistently estimates

$$\boldsymbol{\theta}_{0H} = [c_H \omega_0, c_H \alpha_{01}, \ldots, c_H \alpha_{0p}, \beta_{01}, \ldots, \beta_{0q}]', \tag{1.4}$$

where c_H is a constant that depends on the score function H through the error distribution. In particular, an M-estimator can estimate $\boldsymbol{\theta}_0$ if and only if $c_H = 1$. Hence, using the QMLE, we can estimate $\boldsymbol{\theta}_0$ if and only if the error variance is unity, which is a standard assumption in the literature.

Our formulation allows us to consider an M-estimator based on signed score or LAD and Huber's k-score, among others. For the score functions, we do not assume strong conditions such as monotonicity or continuity; however, we impose mild differentiability conditions on them outside a finite number of points and also skew-symmetry around zero. In this way, we can include wide range of score functions such as LAD, QMLE, MLE, and so on, and also nonmonotone score function. Our method of proof is also different than the usual method of showing uniform convergence of M-scores and their derivatives to their corresponding expectation. It should be mentioned here that the results of this paper and also those of Berkes and Horvath (2004) cannot be applied to ARCH models readily because here it is assumed that $q \geq 1$ and all parameters are nonzero.

In Section 2, we define the class of M-estimators and examples. Section 3 contains the assumptions and main results concerning its asymptotic distribution. Section 4 contains detailed discussions of the examples. In Section 5, we report the results of a simulation study and compute M-estimates based on three different score functions in the context of a real data set containing monthly log returns of the IBM stock from 1926 to 1999 for which a GARCH(1,1) model is assumed. Section 6 states the main conclusions. The details of proofs are given in the Appendix.

2. *M*-ESTIMATORS

Throughout this paper, for a function g, $\dot{g}$ and $\ddot{g}$ will denote the first and second derivatives, respectively, whenever they exist. Also $\log^+(x) := I(x > 1) \log(x)$ and $\text{sign}(x) := I(x > 0) - I(x < 0)$. For a vector or a matrix $\boldsymbol{A}$, $|\boldsymbol{A}|$ will denote the largest absolute value of the entries of $\boldsymbol{A}$, and $\boldsymbol{A}'$ will denote its transpose. Moreover, ϵ will denote a random variable having same distribution as $\{\epsilon_t, \in \mathcal{Z}\}$.

Let $\psi : \mathbb{R} \to \mathbb{R}$ be a skew-symmetric function (i.e., $\psi(-x) = -\psi(x)$, $\forall x \in \mathbb{R} - \{0\}$) that is differentiable in all but a finite number of points. Let $\mathcal{D} \subset \mathbb{R}$

https://doi.org/10.1017/S0266466608080602 Published online by Cambridge University Press

denote the set of points where ψ is differentiable and let $\bar{\mathcal{D}}$ denote its complement. Let $H(x) := x\psi(x)$, $x \in \mathbb{R}$. Note that $H(-x) = H(x)$, $\forall x$. The function H will be called the "score function" for the M-estimation in the scale model. Examples are as follows.

Example 1

LAD score. Let $\psi(x) = \text{sign}(x)$. Then $\bar{\mathcal{D}} = \{0\}$ and $H(x) = |x|$.

Example 2

Huber's k-score. Let $\psi(x) = xI(|x| \le k) + k\,\text{sign}(x)I(|x| > k)$, where $k > 0$ is known. Then $\bar{\mathcal{D}} = \{-k, k\}$ and $H(x) = x^2 I(|x| \le k) + k|x|I(|x| > k)$.

Example 3

QMLE. Let $\psi(x) = x$. Then $H(x) = x^2$.

Example 4

Score function for the maximum likelihood estimation (MLE). Let $\psi(x) = -\dot{f}_0(x)/f_0(x)$, where f_0 is the true density of ϵ, assumed to be known. Then $H(x) = x\{-\dot{f}_0(x)/f_0(x)\}$.

Example 5

Score function for the exponential PMLE. Let $\psi(x) = a|x|^{b-1}\,\text{sign}(x)$, where $a > 0$ and $1 < b \le 2$ are known constants. Such a score can be motivated from the class of densities considered by Nelson (1991) and Robinson and Zaffaroni (2006) to model the innovations of the exponential GARCH model. Here $\bar{\mathcal{D}} = \{0\}$ and $H(x) = a|x|^b$.

Example 6

Score function for the Cauchy PMLE. Let $\psi(x) = \lambda\,\text{sign}(x)/(1 + |x|)$, where $\lambda > 1$ is known. Then $\bar{\mathcal{D}} = \{0\}$ and $H(x) = \lambda|x|/(1 + |x|)$. Note that here ψ is a decreasing function on $(0,\infty)$.

Next we define M-estimators. Recall that in the location model, an M-estimator is defined as the solution to a certain system of equations involving residual functions. To follow the same approach, we first discuss the concept of variance function as follows. First assume that for some $\kappa > 0$,

$$\mathrm{E}[|\epsilon|^\kappa] < \infty. \tag{2.1}$$

Then from Lemma 2.3 and Theorem 2.1 of BHK, σ_t^2 of (1.2) has the following unique almost sure representation:

https://doi.org/10.1017/S0266466608080602 Published online by Cambridge University Press

$$\sigma_t^2 = c_0 + \sum_{j=1}^{\infty} c_j X_{t-j}^2, \qquad t \in \mathcal{Z}, \tag{2.2}$$

where $\{c_j; j \geq 0\}$ are defined in (2.7)–(2.9) of BHK and in (2.5), which follows.

Let Θ be a compact subset of $(0,\infty)^{1+p} \times (0,1)^q$. A typical element in Θ is denoted by $\boldsymbol{\theta} = [\omega, \alpha_1, \ldots, \alpha_p, \beta_1, \ldots, \beta_q]'$. Define the variance function on Θ by

$$v_t(\boldsymbol{\theta}) = c_0(\boldsymbol{\theta}) + \sum_{j=1}^{\infty} c_j(\boldsymbol{\theta}) X_{t-j}^2, \qquad \boldsymbol{\theta} \in \Theta, \qquad t \in \mathcal{Z}, \tag{2.3}$$

where the coefficients $\{c_j(\boldsymbol{\theta}); j \geq 0\}$ are given in BHK (Sect. 3 and display (3.1)) with the property

$$c_j(\boldsymbol{\theta}_0) = c_j, \quad \forall j \geq 0. \tag{2.4}$$

Hence the variance functions satisfy $v_t(\boldsymbol{\theta}_0) = \sigma_t^2$, $t \in \mathcal{Z}$. As an example of this, for the GARCH(1,1) model, with $\boldsymbol{\theta} = (\omega, \alpha, \beta)'$,

$$c_0(\omega, \alpha, \beta) = \omega/(1-\beta), \qquad c_j(\omega, \alpha, \beta) = \alpha\beta^{j-1}, \qquad j \geq 1. \tag{2.5}$$

Using (2.4), (1.1) can be rewritten as

$$X_t = \{v_t(\boldsymbol{\theta}_0)\}^{1/2} \epsilon_t, \qquad 1 \leq t \leq n. \tag{2.6}$$

In (2.6), if f denotes the error density, then the conditional density of X_t given information available up to time $t-1$ will be $v_t^{-1/2}(\boldsymbol{\theta}_0) f\{v_t^{-1/2}(\boldsymbol{\theta}_0) X_t\}$, $1 \leq t \leq n$. Hence, motivated by the maximum likelihood estimator, one can define a random quantity as a minimizer of the negative log likelihood function $(1/n)\sum_{t=1}^n [(\frac{1}{2}) \log v_t(\boldsymbol{\theta}) - \log f\{X_t / v_t^{1/2}(\boldsymbol{\theta})\}]$, $\boldsymbol{\theta} \in \Theta$, or as a solution of its derivative function

$$\sum_{t=1}^{n} \left(\frac{1}{2}\right) [1 - H^*\{X_t / v_t^{1/2}(\boldsymbol{\theta})\}] \{\dot{v}_t(\boldsymbol{\theta}) / v_t(\boldsymbol{\theta})\} = \mathbf{0},$$

where $H^*(x) := x\{-\dot{f}(x)/f(x)\}$.

More generally, with a score function H, we can then define $\boldsymbol{\theta}_n$ in the model (2.3) and (2.6) as a solution of the equation

$$\sum_{t=1}^{n} \left(\frac{1}{2}\right) \{1 - H\{X_t / v_t^{1/2}(\boldsymbol{\theta})\}\} \{\dot{v}_t(\boldsymbol{\theta}) / v_t(\boldsymbol{\theta})\} = \mathbf{0}. \tag{2.7}$$

Note however that $\boldsymbol{\theta}_n$'s are noncomputable because $v_t(\boldsymbol{\theta})$'s are nonobservable.

https://doi.org/10.1017/S0266466608080602 Published online by Cambridge University Press

Next consider an observable approximation $\{\hat{v}_t(\boldsymbol{\theta})\}$ of the process $\{v_t(\boldsymbol{\theta})\}$ defined by

$$\hat{v}_t(\boldsymbol{\theta}) = c_0(\boldsymbol{\theta}) + I(2 \le t) \sum_{j=1}^{t-1} c_j(\boldsymbol{\theta}) X_{t-j}^2, \qquad \boldsymbol{\theta} \in \Theta \qquad 1 \le t \le n.$$

Then an *M*-estimator $\hat{\boldsymbol{\theta}}_n$ is defined as a solution of

$$\sum_{t=1}^{n} \left(\frac{1}{2} \right) \{1 - H\{X_t / \hat{v}_t^{1/2}(\boldsymbol{\theta})\}\} \{\dot{\hat{v}}_t(\boldsymbol{\theta}) / \hat{v}_t(\boldsymbol{\theta})\} = \mathbf{0}. \tag{2.8}$$

For $H(x) = x^2$ of Example 3, $\hat{\boldsymbol{\theta}}_n$ is the celebrated QMLE as discussed by Engle (1982) and Weiss (1986). For $H(x) = |x|$ of Example 1, $\hat{\boldsymbol{\theta}}_n$ can be called the LAD estimator.

3. ASYMPTOTIC DISTRIBUTION OF $\hat{\boldsymbol{\theta}}_n$

The results of this paper are derived under the following assumptions.

Model assumptions. The parameter space Θ is a compact set, and its interior Θ_0 contains both $\boldsymbol{\theta}_0$ and $\boldsymbol{\theta}_{0H}$ of (1.3) and (1.4), respectively. Moreover, (2.1), (2.3), and (2.6) hold, and $\{X_t\}$ are stationary and ergodic.

Conditions on the score function

Identifiability condition. Corresponding to the score function H, there exists a unique number $c_H > 0$ satisfying

$$\mathrm{E}[H(\epsilon / c_H^{1/2})] = 1. \tag{3.1}$$

Moment conditions

$$\mathrm{E}[H(\epsilon / c_H^{1/2})]^2 < \infty \quad \text{and} \quad 0 < \mathrm{E}\{(\epsilon / c_H^{1/2}) \dot{H}(\epsilon / c_H^{1/2})\} < \infty. \tag{3.2}$$

Smoothness conditions

Condition SM1. There exists a function L satisfying

$$|H(es) - H(e)| \le L(e) |s^2 - 1|, \qquad e \in \mathbb{R}^1, \qquad s > 0, \tag{3.3}$$

where

$$\mathrm{E} \log^+ \{L(\epsilon / c_H^{1/2})\} < \infty. \tag{3.4}$$

Condition SM2. There exists a function Λ such that for $e \in \mathbb{R}^1$, $s > 0$, $es, e \in \mathcal{D}$,

$$|\dot{H}(es) - \dot{H}(e)| \le \Lambda(e) |s - 1|, \tag{3.5}$$

where

$$\mathrm{E}\{|\epsilon / c_H^{1/2}| \Lambda(\epsilon / c_H^{1/2})\} < \infty. \tag{3.6}$$

https://doi.org/10.1017/S0266466608080602 Published online by Cambridge University Press

Condition SM3. There exists a function Λ^* satisfying

$$|\Lambda(e + es) - \Lambda(e)| \leq \Lambda^*(e)s, \qquad e \in \mathbb{R}^1, \qquad s > 0, \tag{3.7}$$

where

$$\mathrm{E}\log^+\{\Lambda^*(\epsilon/c_H^{1/2})\} < \infty. \tag{3.8}$$

Remark 3.1. Recall that if ϕ is a nondecreasing and odd score function satisfying some smoothness conditions, then there exists a point θ_ϕ such that $\mathrm{E}[\phi(\epsilon - \theta_\phi)] = 0$. Because a location estimation problem, say, $x_t = \theta + \epsilon_t$ can be rewritten as $x_t = (\theta + \theta_\phi) + (\epsilon_t - \theta_\phi)$, a location-invariant M-estimator based on ϕ estimates $\theta + \theta_\phi$. Condition (3.1) is a natural counterpart of the location model in the scale estimation problem. To illustrate, suppose that the score function H in a scale estimation problem is nondecreasing on $[0,\infty)$ with $\lim_{c\to 0+} H(\epsilon/c^{1/2}) = H(\infty)$ and $\lim_{c\to\infty} H(\epsilon/c^{1/2}) = H(0)$ and suppose that the expectation function $g(c) = \mathrm{E}\{H(\epsilon/c^{1/2})\}$ satisfying $H(0) \leq g(c) \leq H(\infty)$ has a range containing 1. Then, there exists a point $c_H > 0$ satisfying $\mathrm{E}[H(\epsilon/c_H^{1/2})] = 1$ that is assumed unique in (3.1). Consequently, the scale-invariant M-estimator based on H actually estimates $\boldsymbol{\theta}_{0H}$.

In other words, the assumption $\mathrm{E}[H(\epsilon/c_H^{1/2})] = 1$ is essentially an identifiability condition and determines what the M-estimator can consistently estimate. It is used for showing that $n^{-1}\dot{\boldsymbol{M}}_n(\boldsymbol{\theta}_{0H}) \to \mathbf{0}$ almost surely (a.s.) in (A.21) in the Appendix. It corresponds to the identification condition that the quasi-log-likelihood have a unique maximum at the true conditional mean and relative scale parameters as considered by Newey and Steigerwald (1997) and Berkes and Horvath (2004). Li and Turtle (2000, p. 177) imposed a similar condition on the conditional fourth moment of X_t's.

Bougerol and Picard (1992) discussed necessary and sufficient conditions for the existence of a stationary solution to (1.1) and (1.2). Giraitis, Kokoszka, and Leipus (2000) discussed necessary and sufficient conditions for a stationary solution under $\mathrm{E}(X_1^2) < \infty$.

To state our main results on $\hat{\boldsymbol{\theta}}_n$ of (2.8), define the score function factor

$$\sigma^2(H) := 4\,\mathrm{var}\{H(\epsilon/c_H^{1/2})\}/[\mathrm{E}\{(\epsilon/c_H^{1/2})\dot{H}(\epsilon/c_H^{1/2})\}]^2,$$

where $\mathrm{var}\{H(\epsilon/c_H^{1/2})\}$ is assumed to be positive in the moment condition. Also, define

$$\boldsymbol{G} := \mathrm{E}\{\dot{v}_1(\boldsymbol{\theta}_{0H})\dot{v}_1'(\boldsymbol{\theta}_{0H})/v_1^2(\boldsymbol{\theta}_{0H})\}.$$

THEOREM 3.1. *Suppose that the model assumptions, the identifiability condition, the moment conditions, and the smoothness conditions SM1–SM3 hold. Then*

$$n^{1/2}(\hat{\boldsymbol{\theta}}_n - \boldsymbol{\theta}_{0H}) \to N[0, \sigma^2(H)\boldsymbol{G}^{-1}]. \tag{3.9}$$

https://doi.org/10.1017/S0266466608080602 Published online by Cambridge University Press

Remark 3.2. The preceding result states that using the score function H we can consistently estimate $\boldsymbol{\theta}_{0H}$. With $H(x) = x^2$, $c_H = \mathrm{E}(\epsilon^2)$, and hence, using the QMLE, we can consistently estimate $[\mathrm{E}(\epsilon^2)\omega_0, \mathrm{E}(\epsilon^2)\alpha_{01}, \ldots, \mathrm{E}(\epsilon^2)\alpha_{0p}, \beta_{01}, \ldots, \beta_{0q}]'$. Note that $\mathrm{E}(\epsilon^2) = 1$ is a standard assumption in the literature, except in Berkes and Horvath (2004), where known value of c_H is assumed for different H. Hence, when the error variance is unity, we can estimate $\boldsymbol{\theta}$ using the QMLE. We can estimate $\boldsymbol{\theta}$ using any other score function H whenever $c_H = 1$ for the corresponding error distribution.

Note that Theorem 3.1 is derived under weak moment assumptions on the error distribution. We imposed conditions on the score function H that, in most examples discussed in Section 4, are translated to very mild moment assumptions on the error distribution. In Example 6, existence of Condition (2.1) is enough to have asymptotic normality of the estimators, where κ can be a fraction and need not even be of known value.

Consistent estimator of $\sigma^2(H)\boldsymbol{G}^{-1}$ can be obtained as follows. From (2.6) and (A.2) of the Appendix, $\epsilon_t/c_H^{1/2} = X_t/v_t^{1/2}(\boldsymbol{\theta}_{0H})$. Because

$$\sigma^2(H) = 4[\mathrm{E}\{H(\epsilon/c_H^{1/2})\}^2 - 1][\mathrm{E}\{(\epsilon/c_H^{1/2})\dot{H}(\epsilon/c_H^{1/2})\}]^{-2},$$

it can be estimated based on residuals by

$$\widehat{\sigma^2(H)} = 4\left[(1/n)\sum_{t=1}^{n}\{H(X_t/\hat{v}_t^{1/2}(\hat{\boldsymbol{\theta}}_n))\}^2 - \left[(1/n)\sum_{t=1}^{n}H(X_t/\hat{v}_t^{1/2}(\hat{\boldsymbol{\theta}}_n))\right]^2\right] \times \left[(1/n)\sum_{t=1}^{n}\{(X_t/\hat{v}_t^{1/2}(\hat{\boldsymbol{\theta}}_n))\dot{H}(X_t/\hat{v}_t^{1/2}(\hat{\boldsymbol{\theta}}_n))\}\right]^{-2}.$$

Also, $\boldsymbol{G}^{-1}$ can be estimated by $(\hat{\boldsymbol{G}})^{-1}$ where

$$\hat{\boldsymbol{G}} = (1/n)\sum_{t=1}^{n}\{\dot{\hat{v}}_t(\hat{\boldsymbol{\theta}}_n)\dot{\hat{v}}_t'(\hat{\boldsymbol{\theta}}_n)/\hat{v}_t^2(\hat{\boldsymbol{\theta}}_n)\}.$$

The following result is obtained by strengthening (3.4) to

$$\mathrm{E}\{L(\epsilon/c_H^{1/2})\} < \infty \tag{3.10}$$

and also under the condition

$$\mathrm{E}(X_0^2) < \infty. \tag{3.11}$$

PROPOSITION 3.1. *In addition to the assumptions and conditions of Theorem 3.1, assume that (3.10) and (2.11) hold. Then* $\widehat{\sigma^2(H)}(\hat{\boldsymbol{G}})^{-1}$ *converges in probability to* $\sigma^2(H)\boldsymbol{G}^{-1}$.

Remark 3.3. Note that (3.10) does not impose any extra condition on the error distributions for the six examples discussed in the next section. Condition

https://doi.org/10.1017/S0266466608080602 Published online by Cambridge University Press

(3.11) is standard in the literature for proving the asymptotic normality of estimators; see, for example, Giraitis et al. (2000), Hall and Yao (2003), and Peng and Yao (2003). However, we assumed this condition only to estimate the asymptotic standard errors (SEs) of the M-estimators. It is stronger than (2.1) when, for example, κ is a fraction.

4. DISCUSSION ON EXAMPLES

We now discuss the results of Section 3 with the following choices of ψ.

Example 1

LAD. Assume that

$$\mathrm{E}(\epsilon_1^2) < \infty. \tag{4.1}$$

Here (3.1) holds with $c_H^{1/2} = \mathrm{E}(|\epsilon|)$. Under (4.1), (3.2) holds. Because

$$|H(es) - H(e)| = \alpha|e|\,|(s-1)| \le \alpha|e|\,|(s-1)|(s+1) = \alpha|e|\,|s^2-1|,$$

(3.3) is satisfied with $L(e) = |e|$. Because $\mathrm{E}|\epsilon| < \infty$, (3.4) holds. Because $\dot{H}(es) - \dot{H}(e) = 0$, (3.5) is satisfied with, say, $\Lambda(e) = 1$. Consequently, conditions (3.6)–(3.8) are trivially satisfied.

Remark 4.1. Peng and Yao (2003) defined three different types of LAD estimators based on regression relationships after a reparametrization so that the errors or their appropriate functions have median zero. The LAD of this paper is defined without assuming such relationship because, in practice, the error distribution is unknown. Peng and Yao (2003, Remark 1) did mention that a LAD estimator of our Example 1 can be considered, but its asymptotic properties are complex and this was not pursued.

Example 2

Huber's estimator. Assume that

$$\mathrm{E}(\epsilon_1^2) < \infty. \tag{4.2}$$

To verify (3.1), define a function g on $(0,\infty)$ by

$$g(c) := \mathrm{E}[H(c\epsilon)] = c^2 \int_{-k/c}^{k/c} x^2 f(x)\,dx + kc \int_{k/c}^{\infty} x f(x)\,dx + kc \int_{-\infty}^{-k/c} (-x) f(x)\,dx.$$

https://doi.org/10.1017/S0266466608080602 Published online by Cambridge University Press

Note that under (4.2), $\lim_{c\to 0+} g(c) = 0$ and $\lim_{c\to\infty} g(c) = \infty$. Moreover, g is continuous and strictly increasing because

$$\begin{aligned}\dot{g}(c) &= 2c\int_{-k/c}^{k/c} x^2 f(x)\,dx + c^2\{(-k/c^2)(k/c)^2 f(k/c) - (k/c^2)(-k/c)^2 f(-k/c)\}\\ &\quad + k\int_{k/c}^{\infty} xf(x)\,dx - kc(-k/c^2)(k/c)f(k/c)\\ &\quad + k\int_{-\infty}^{-k/c} (-x)f(x)\,dx + kc(k/c^2)(k/c)f(-k/c)\\ &= 2c\int_{-k/c}^{k/c} x^2 f(x)\,dx + k\int_{k/c}^{\infty} xf(x)\,dx + k\int_{-\infty}^{-k/c} (-x)f(x)\,dx > 0.\end{aligned}$$

Hence c_H in (3.1) is unique. Under (4.2), (3.2) holds. Next we verify (3.3) with $L(e) = e^2$. Here

$$\begin{aligned}H(es) - H(e) &= e^2 s^2 I(|e| \le k/s) + k|es|I(|e| > k/s) - e^2 I(|e| \le k)\\ &\quad - k|e|I(|e| > k).\end{aligned}$$

Case 1. $|e| \le k/s$ and $|e| \le k$. Here $H(es) - H(e) = e^2(s^2 - 1)$.

Case 2. $|e| \le k/s$ and $|e| > k$. Here $s < 1$; $H(es) - H(e) = (es)^2 + k \times -|e| < (k)^2 + k \times -k < 0$. Hence $|H(es) - H(e)| = k \times |e| - e^2 s^2 \le |e| \times |e| - e^2 s^2 = e^2(1 - s^2)$.

Case 3. $|e| > k/s$ and $|e| \le k$. Here $s > 1$; $H(es) - H(e) = k \times |es| - e^2 > k \times k - k^2 \ge 0$. Hence $|H(es) - H(e)| = k \times |es| - e^2 \le |es| \times |es| - e^2 = e^2(s^2 - 1)$.

Case 4. $|e| > k/s$ and $|e| > k$. Here $H(es) - H(e) = k|e|(s - 1)$, and so

$$|H(es) - H(e)| = k \times |e|\,|(s-1)| \le |e| \times |e|\,|(s-1)|(s+1) = e^2|s^2 - 1|.$$

In a similar fashion one can verify (3.5) and (3.6) with $\Lambda(e) = C|e|$ where C is a large positive constant and (3.7) and (3.8) with $\Lambda^*(e) = C|e|$.

Example 3

QMLE. Assume that

$$\mathrm{E}(\epsilon^4) < \infty. \tag{4.3}$$

Here (3.1) holds with $c_H = \mathrm{E}(\epsilon^2)$. Under (4.3), (3.2) holds. Assumptions (3.3) and (3.4) hold with $L(e) = e^2$, and (3.5) and (3.6) hold with $\Lambda(e) = 2|e|$. Finally, (3.7) and (3.8) hold with $\Lambda^*(e) = 2|e|$.

https://doi.org/10.1017/S0266466608080602 Published online by Cambridge University Press

Example 4

MLE. Note that integration by parts yields that $\mathrm{E}[H(\epsilon)] = 1$ and so $c_H = 1$ under the uniqueness assumption. The validity of (3.2)–(3.8) needs to be verified when the explicit form of f_0 is given.

Example 5

Exponential PMLE. Assume that

$$\mathrm{E}|\epsilon|^{2b} < \infty. \tag{4.4}$$

Under (4.4), (3.1) and (3.2) hold with $c_H = \{a\mathrm{E}|\epsilon|^b\}^{2/b}$. Also (3.3) and (3.4) hold with $L(e) = a|e|^b$. Using sign (es) = sign (e) sign (s) and $s \geq 0$, (3.5) and (3.6) hold with $\Lambda(e) = ab|e|^{b-1}$. Finally, using the mean value theorem on the function $g(s) = (1 + s)^{b-1}$ at $s = 0$, (3.7) and (3.8) hold with $\Lambda^*(e) = ab(b - 1)|e|^{b-1}$.

The usefulness of the preceding estimators can be further demonstrated by considering a family of t-density with $\nu > 0$ degrees of freedom where the density of the error random variable ϵ is proportional to

$$(1 + x^2/\nu)^{-(\nu+1)/2}. \tag{4.5}$$

Note that $\mathrm{E}|\epsilon|^\mu < \infty$ for all $0 < \mu < \nu$, $\mathrm{E}(\epsilon) = 0$ for $\nu > 1$, and $\mathrm{Var}(\epsilon) = \nu/(\nu - 2)$ for $\nu > 2$.

When $(2 < \nu \leq 4)$, for any b such that $2 < 2b < \nu$, $\mathrm{E}|\epsilon|^{2b} < \infty$ but $\mathrm{E}\epsilon^4 = \infty$. Therefore, the PMLE satisfies (3.9), whereas the asymptotic normality of the QMLE does not hold.

Example 6

Cauchy PMLE. For (3.1), consider the function $g(c) = \mathrm{E}\{|\epsilon|/(c + |\epsilon|)\}$. Note that $g(0) = 1$ and by the bounded convergence theorem $\lim_{c\to\infty} g(c) = 0$. Moreover, assuming that the differentiation and expectation can be interchanged,

$$\dot{g}(c) = -\mathrm{E}\{|\epsilon|/(c + |\epsilon|)^2\} < 0.$$

Hence there is a unique c satisfying $g(c) = 1/\lambda$, and thus (3.1) is verified. Because

$$\begin{aligned} |H(es) - H(e)| &= |-\{\lambda|e|\,|s - 1|\}/\{(1 + |e|s)(1 + |e|)\}| \\ &\leq \lambda\{|e|/(1 + |e|)\}|s - 1| < \lambda|s^2 - 1|, \end{aligned}$$

(3.3) and (3.4) hold with $L(e) = \lambda$. Similarly, (3.5) and (3.6) hold with a constant function Λ, and so (3.7) and (3.8) hold trivially.

https://doi.org/10.1017/S0266466608080602 Published online by Cambridge University Press

The preceding class of estimators is useful for error distributions for which κ in (2.1) is possibly a fraction and even unknown. For illustration, suppose that the error density satisfies (4.5) for some unknown ν with $0 < \nu < 4$. Because (2.1) holds with $\kappa = \nu/2$, the estimator based on any known $\lambda > 1$ satisfies (3.9), whereas the asymptotic normality of the QMLE does not hold.

5. SIMULATION AND DATA ANALYSIS

Simulation study. We simulated n observations from a GARCH(1,1) model of (1.1) and (1.2) R times (replication) and computed the QMLE, LAD, and Huber's estimates with $k = 1.5$ for each replication. We use variants of the iteratively weighted least squares method to compute these three M-estimates. These methods are motivated by an algorithm proposed by Mak, Wong, and Li (1997) to compute the QMLE for a general nonlinear time series model with conditional heterogeneous variances; for details of our algorithms and the corresponding programs written using the software R, see Iqbal and Mukherjee (2007). We report here the simulation results when the true parameters are $\omega_0 = 1.5$, $\alpha_0 = 0.15$, and $\beta_0 = 0.55$ and (i) Scheme N where the error distribution is standard normal and (ii) Scheme T where the error distribution is a standardized t-distribution with 4 degrees of freedom; for simulation results under other parameter combinations and error distributions, see Iqbal and Mukherjee (2007).

In Table 1, we have reported mean squared errors (MSEs) based on R replications with score function H when (i) $H(x) = x^2$ (QMLE), (ii) $H(x) = |x|$ (LAD), and (iii) the Huber's score function with $k = 1.5$. From (3.9), (i) the QMLE estimates $(\omega_0, \alpha_0, \beta_0)$, (ii) the LAD estimates $(c_L\omega_0, c_L\alpha_0, \beta_0)$ where $c_L = 2/\pi$ under Scheme N and 0.55 under Scheme T, and (iii) the Huber's score estimates $(c_{HU}\omega_0, c_{HU}\alpha_0, \beta_0)$ where $c_{HU} = 0.825$ under Scheme N and $c_{HU} = 0.725$ under Scheme T. Note that in the last few cases, c_L and c_{HU} do not have closed form expressions and so their numerical values are obtained by solving equations that are a law of large numbers version of (3.1), namely, $(1/I)\sum_{i=1}^{I}\{H(\epsilon_i/c^{1/2})\} - 1 \approx 0$, using the simple bisection method for solving univariate equations. Finally, for example, the MSE of the Huber's estimator of ω is computed as $\sum_{r=1}^{R}(\hat{\omega}_r - c_{HU}\omega_0)^2/R$ where $\hat{\omega}_r$ is the Huber's estimate from the rth replication, $1 \le r \le R$.

TABLE 1. Estimated MSEs of the M-estimators in the GARCH(1,1) model

	Scheme N	$n = 1{,}000$	$R = 21$	Scheme T	$n = 1{,}000$	$R = 71$
Estimand	QMLE	LAD	Huber	QMLE	LAD	Huber
$c_H\omega$	0.3768809	0.1484265	0.2208393	0.9456990	0.0907649	0.1785953
$c_H\alpha$	0.0018090	0.0007984	0.0016491	0.0079990	0.0008005	0.0021937
β	0.0214904	0.0215348	0.0227471	0.0443205	0.0239880	0.0277554

https://doi.org/10.1017/S0266466608080602 Published online by Cambridge University Press

Discussion. From Table 1 and some other simulation results it turns out that even for normal error density and large n, the QMLE performs worse than the LAD with the exception that it performs slightly better while estimating β. Under Scheme T when the fourth moment of the error density is infinite, the performance of the QMLE is even worse, as expected, and in this case LAD outperforms the Huber's estimator also.

IBM data and estimates. Next we consider the monthly log returns of IBM stock from 1926 to 1999 (888 observations $\{r_t, 1 \le t \le n = 888\}$ with $r_1 = -1.0434$ and $r_n = 4.5633$). Tsay (2005, Ch. 3, Examp. 3.4) has analyzed this important data set and fitted various types of conditional heteroskedastic models to it. The data can be found at http://faculty.chicagogsb.edu/ruey.tsay/teaching/fts2/m-ibmsplnsu.dat. Tsay (2005) used an autoregressive model of order one with intercept and with GARCH(1,1) errors to model these data as $r_t = \mu + \rho r_{t-1} + X_t$ with $X_t = \sigma_t \epsilon_t$ and $\sigma_t^2 = \omega_0 + \alpha_0 X_{t-1}^2 + \beta_0 \sigma_{t-1}^2$, $1 \le t \le n$. Using the Gaussian likelihood for errors, Tsay (2005) obtained $\hat{\mu} = 1.23$, $\hat{\rho} = 0.099$, $\hat{\omega}_0 = 3.206$, $\hat{\alpha}_0 = 0.103$, and $\hat{\beta}_0 = 0.825$ with respective standard errors (SEs) 0.222, 0.037, 0.947, 0.021, and 0.037. To compare M-estimates with Tsay's estimates, we use $\{r_{t+1} - 1.23 - 0.099 r_t, 1 \le t \le (n-1)\}$ as our observations and compute (i) the QMLE, (ii) the LAD, and (iii) the Huber's estimate with $k = 1.5$ for the parameters of the GARCH(1,1) model. In Table 2 we report M-estimates and their estimated SEs.

Discussion. The QMLE estimates that we obtained are slightly different from those of Tsay (2005) for two main reasons. First, Tsay (2005) performed simultaneous estimation of intercept and other parameters of the autoregressive and GARCH(1,1) models, whereas we assumed that the intercept and slope parameters of the autoregressive models are known. Second, Tsay assumed $\mathrm{E}(\epsilon^2) = 1$, whereas we just assumed that $\mathrm{E}(\epsilon^2) < \infty$. From Remark 3.2, the QMLE actually estimates $(\mathrm{E}(\epsilon^2)\omega_0, \mathrm{E}(\epsilon^2)\alpha_0, \beta_0)$. Under $\mathrm{E}(\epsilon^2) = 1$, it estimates $(\omega_0, \alpha_0, \beta_0)$.

Notice that in a GARCH(1,1) model, *any* M-estimator should consistently estimate the third coordinate of $\boldsymbol{\theta}_{0H} = (c_H \omega_0, c_H \alpha_0, \beta_0)$ and this is reflected in the last row of Table 2 as the three different M-estimates are 0.84, 0.86, and 0.81 which are quite close to each other.

Table 2. QMLE, LAD, and Huber's estimates from the IBM data and their SEs

Parameter	QMLE	SE	LAD	SE	Huber's estimate	SE
$c_H\omega$	2.9606623	1.3854702	2.0682954	0.9445278	2.8448848	1.1873861
$c_H\alpha$	0.0974596	0.0309250	0.0912957	0.0251676	0.1236431	0.0323297
β	0.8357814	0.0529580	0.8598516	0.0391407	0.8109211	0.0493048

https://doi.org/10.1017/S0266466608080602 Published online by Cambridge University Press

6. CONCLUSIONS

In this paper we showed, among other things, that using the commonly used QMLE, we can consistently estimate the parameters of the GARCH models only up to a multiplicative constant when fourth moment exists for the error distributions. However, under the existence of some fractional error moment, it is still possible to estimate consistently scaler multiples of the GARCH parameters; what we estimate depends very much on the score function used for estimation, and this is different from what we observe in the *M*-estimation of parameters in the usual linear regression and autoregressive models.

REFERENCES

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

Berkes, I., L. Horvath, & P. Kokoszka (2003) GARCH processes: Structure and estimation. *Bernoulli* 9, 201–228.

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

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

Comte, F. & O. Lieberman (2003) Asymptotic theory for multivariate GARCH processes. *Journal of Multivariate Analysis* 84, 61–84.

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

Engle, R.F. & G. Gonzalez-Rivera (1991) Semiparametric ARCH models. *Journal of Business & Economic Statistics* 9, 345–349.

Giraitis, L., P. Kokoszka, & R. Leipus (2000) Stationary ARCH models: Dependence structure and central limit theorem. *Econometric Theory* 16, 3–22.

Hall, P. & C.C. Heyde (1980) *Martingale Limit Theory and Its Applications*. Academic Press.

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

Iqbal, F. & K. Mukherjee (2007) Computation of *M*-Estimates in GARCH Models. Technical Report, Department of Mathematical Sciences, University of Liverpool.

Jeantheau, T. (1998) Strong consistency of estimators for multivariate ARCH models. *Econometric Theory* 14, 70–86.

Klimko, L.A. & P.I. Nelson (1978) On conditional least squares estimation for stochastic processes. *Annals of Statistics* 6, 629–642.

Lee, S.W. & B.E. Hansen (1994) Asymptotic theory for the GARCH(1,1) quasi-maximum likelihood estimator. *Econometric Theory* 10, 29–52.

Li, D. & H. Turtle (2000) Semiparametric ARCH models: An estimating function approach. *Journal of Business & Economic Statistics* 18, 174–186.

Lumsdaine, R.L. (1996) Consistency and asymptotic normality of the quasi-maximum likelihood estimator in IGARCH(1,1) and covariance stationary GARCH(1,1) models. *Econometrica* 64, 575–596.

Mak, T., H. Wong, & W. Li (1997) Estimation of nonlinear time series with conditional heteroscedastic variances by iteratively weighted least squares. *Computational Statistics & Data Analysis* 24, 169–178.

Ndebu, J. (2006) Some estimation methods in ARCH and GARCH models. M.Sc. Dissertation, Department of Mathematical Sciences, University of Liverpool.

https://doi.org/10.1017/S0266466608080602 Published online by Cambridge University Press

Nelson, D. (1991) Conditional heteroscedasticity in asset returns: A new approach. *Econometrica* 59, 347–370.
Newey, W. & D. Steigerwald (1997) Asymptotic bias for quasi-maximum-likelihood estimators in conditional heteroscedasticity models. *Econometrica* 65, 587–599.
Peng, L. & Q. Yao (2003) Least absolute deviation estimation for ARCH and GARCH models. *Biometrika* 90, 967–975.
Robinson, P. & P. Zaffaroni (2006) Pseudo-maximum-likelihood estimation of ARCH(∞) models. *Annals of Statistics* 34, 1049–1074.
Straumann, D. & T. Mikosch (2006) Quasi-maximum-likelihood estimation in conditionally heteroscedastic time series: A stochastic recurrence equations approach. *Annals of Statistics* 34, 2449–2495.
Tsay, R.S. (2005) *Analysis of Financial Time Series*, 2nd ed. Wiley.
Weiss, A.A. (1986) Asymptotic Theory for ARCH models: Estimation and testing. *Econometric Theory* 2, 107–131.

APPENDIX

In this section, first we state several results from BHK that will be useful in deriving the theoretical results of this paper. Throughout, C, C_1, and so on, will denote generic positive constants whose values will possibly change from expression to expression.

Fact 1. Let $\{(A_t, B_t, C_t); t \geq 0\}$ be a sequence of identically distributed random variables. If $\mathrm{E}\log^+ A_0 + \mathrm{E}\log^+ B_0 + \mathrm{E}\log^+ C_0 < \infty$, then for any $|r| < 1$,

$$\sum_{t=0}^{\infty} (A_t + B_t C_t) r^t \quad \text{converges with probability 1.} \tag{A.1}$$

Consider the variance function $v_t(\boldsymbol{\theta}) = c_0(\boldsymbol{\theta}) + \sum_{j=1}^{\infty} c_j(\boldsymbol{\theta}) X_{t-j}^2$ defined in (2.3).

Fact 2. The coefficients $\{c_j(\boldsymbol{\theta})\}$ and $v_t(\boldsymbol{\theta})$ are differentiable in the interior Θ_0 of Θ for all $t \in \mathcal{Z}$. Moreover,

$$v_t(\boldsymbol{\theta}_{0H}) = c_H v_t(\boldsymbol{\theta}_0). \tag{A.2}$$

Fact 3. Let $0 < u_* < 1$ be defined by

$$u_* = \min\{\omega, \alpha_1, \ldots, \alpha_p, \beta_1, \ldots, \beta_q; \qquad \boldsymbol{\theta} = [\omega, \alpha_1, \ldots, \alpha_p, \beta_1, \ldots, \beta_q]' \in \Theta\}.$$

Then there is a number $0 < \rho < 1$ such that $\forall \boldsymbol{\theta} \in \Theta_0$,

$$C_1 (u_*)^j \leq c_j(\boldsymbol{\theta}) \leq C_2 (\rho)^j, \quad \forall j \geq 0, \tag{A.3}$$

$$|\dot{c}_j(\boldsymbol{\theta})| \leq C_3 (\rho)^j, \quad \forall j \geq 0, \tag{A.4}$$

and

$$|\ddot{c}_j(\boldsymbol{\theta})| \leq C_4 \rho^j, \quad \forall j \geq 0. \tag{A.5}$$

https://doi.org/10.1017/S0266466608080602 Published online by Cambridge University Press

Fact 4. There exist random variables Z_0 and Z_1, both independent of $\{\epsilon_t; t \geq 1\}$, such that

$$0 < v_t(\boldsymbol{\theta}) - \hat{v}_t(\boldsymbol{\theta}) \leq \rho^t Z_0 \tag{A.6}$$

and

$$|\dot{v}_t(\boldsymbol{\theta}) - \dot{\hat{v}}_t(\boldsymbol{\theta})| \leq \rho^t Z_1. \tag{A.7}$$

Fact 5. For any $\nu > 0$

$$\mathrm{E}[\sup\{|\dot{v}_1(\boldsymbol{\theta})/v_1(\boldsymbol{\theta})|^\nu \ \boldsymbol{\theta} \in \Theta_0\}] < \infty \tag{A.8}$$

and

$$\mathrm{E}[\sup\{|\ddot{v}_1(\boldsymbol{\theta})/v_1(\boldsymbol{\theta})|^\nu \boldsymbol{\theta} \in \Theta_0\}] < \infty. \tag{A.9}$$

Let $h(x) = (\frac{1}{2})\{1 - H(x)\}$ and $G(x) = x\dot{H}(x)$. Using

$$G(es) - G(e) = se\{\dot{H}(es) - \dot{H}(e)\} + e\dot{H}(e)(s - 1)$$

and

$$G(es_1) - G(es_2) = (s_1 - s_2)e\{\dot{H}(es_1) - \dot{H}(e) + \dot{H}(e)\}$$
$$+ es_2\{\dot{H}\{(s_1/s_2)(es_2)\} - \dot{H}(es_2)\}$$

and the smoothness conditions SM1–SM3, we obtain the following facts.

Fact 6.

$$|G(es) - G(e)| \leq |s - 1|\,|e|\{s\Lambda(e) + |\dot{H}(e)|\}, \tag{A.10}$$

and

$$|G(es_1) - G(es_2)| \leq |s_1 - s_2|\,|e|\{|s_1 - 1|\Lambda(e) + |\dot{H}(e)| + \Lambda(es_2)\}. \tag{A.11}$$

There are several ways to derive the asymptotic distribution of estimators obtained through the roots of certain equations. In the present case the criterion equations of (2.8) are smooth functions of $\boldsymbol{\theta}$. Hence to derive the asymptotic distribution of $\hat{\boldsymbol{\theta}}_n$ we use a modified result of Klimko and Nelson (1978, Thm. 2.1, Cor. 2.1 and Thm. 2.2) that discusses the asymptotic distribution of a sequence of solutions $\{\hat{\boldsymbol{\alpha}}_n\}$ to some estimating equations defined by a smooth criterion function $\{Q_n\}$. See also Hall and Heyde (1980).

Accordingly, let $Q_n; S \to \mathbb{R}$ be a criterion function where S is a compact subset of $\mathbb{R}^d$. Let Q_n be twice differentiable with respect to $\boldsymbol{\alpha}$ in a δ-neighborhood $\{\boldsymbol{\alpha}; \|\boldsymbol{\alpha} - \boldsymbol{\alpha}_0\| \leq \delta\} \subset S$ of $\boldsymbol{\alpha}_0 \in S$ with partial derivatives $\dot{\boldsymbol{Q}}_n$ and $\ddot{\boldsymbol{Q}}_n$ and having the following representation:

https://doi.org/10.1017/S0266466608080602 Published online by Cambridge University Press

$$Q_n(\boldsymbol{\alpha}) = Q_n(\boldsymbol{\alpha}_0) + (\boldsymbol{\alpha} - \boldsymbol{\alpha}_0)'\dot{\boldsymbol{Q}}_n(\boldsymbol{\alpha}_0) + \left(\frac{1}{2}\right)(\boldsymbol{\alpha} - \boldsymbol{\alpha}_0)'\ddot{\boldsymbol{Q}}_n(\boldsymbol{\alpha}_0)(\boldsymbol{\alpha} - \boldsymbol{\alpha}_0)$$

$$+ \left(\frac{1}{2}\right)(\boldsymbol{\alpha} - \boldsymbol{\alpha}_0)'\boldsymbol{T}_n(\boldsymbol{\alpha}^*)(\boldsymbol{\alpha} - \boldsymbol{\alpha}_0), \tag{A.12}$$

where $\boldsymbol{\alpha}^*$ is an intermediate point satisfying $\|\boldsymbol{\alpha}^* - \boldsymbol{\alpha}_0\| \leq \delta$ and $\boldsymbol{T}_n(\boldsymbol{\alpha}^*) = \ddot{\boldsymbol{Q}}_n(\boldsymbol{\alpha}^*) - \ddot{\boldsymbol{Q}}_n(\boldsymbol{\alpha}_0)$. Suppose that $\boldsymbol{\alpha}_0 \in S$ is the true parameter with the following properties:

$$n^{-1}\dot{\boldsymbol{Q}}_n(\boldsymbol{\alpha}_0) \to \mathbf{0} \quad \text{in probability,} \tag{A.13}$$

$$(2n)^{-1}\ddot{\boldsymbol{Q}}_n(\boldsymbol{\alpha}_0) \to \boldsymbol{V}_Q \quad \text{in probability for some positive definite matrix } \boldsymbol{V}_Q, \tag{A.14}$$

$$\lim_{n\to\infty}\lim_{\delta\to 0+} \sup\{(n\delta)^{-1}|\boldsymbol{T}_n(\boldsymbol{\alpha}^*)|; \|\boldsymbol{\alpha}^* - \boldsymbol{\alpha}_0\| \leq \delta\} < \infty \quad \text{a.s.} \tag{A.15}$$

Define a sequence of estimators $\{\hat{\boldsymbol{\alpha}}_n\} \in S$ as a solution of the set of equations

$$\dot{\boldsymbol{Q}}_n(\hat{\boldsymbol{\alpha}}_n) = \mathbf{0}. \tag{A.16}$$

THEOREM KN (Klimko and Nelson). *Suppose that (A.12)–(A.15) hold. Then the following conditions hold:*

(i) for every $\epsilon > 0$, there is an event $\mathcal{E}$ with $P(\mathcal{E}) > 1 - \epsilon$ and an n_0 such that for all $n \geq n_0$, $\{\hat{\boldsymbol{\alpha}}_n\}$ satisfies the equation (A.16) and Q_n attains a relative minimum at $\{\hat{\boldsymbol{\alpha}}_n\}$ on $\mathcal{E}$. Moreover, $\hat{\boldsymbol{\alpha}}_n \to \boldsymbol{\alpha}_0$ in probability.

(ii) In addition, if for some positive definite matrix $\boldsymbol{W}_Q$,

$$\left(\frac{1}{2}\right)n^{-1/2}\dot{\boldsymbol{Q}}_n(\boldsymbol{\alpha}_0) \Rightarrow N[\mathbf{0}, \boldsymbol{W}_Q], \tag{A.17}$$

then

$$n^{1/2}(\hat{\boldsymbol{\alpha}}_n - \boldsymbol{\alpha}_0) \Rightarrow N[\mathbf{0}, \boldsymbol{V}_Q^{-1}\boldsymbol{W}_Q\boldsymbol{V}_Q^{-1}]. \tag{A.18}$$

Idea of the Proof of Theorem 3.1. Define a function ρ by $\rho(x) = \int_0^x \psi(t)\,dt$ for $x \geq 0$ and $\rho(x) = \rho(-x)$ for $x < 0$. Define

$$m_t(\boldsymbol{\theta}) = \rho\{X_t/v_t^{1/2}(\boldsymbol{\theta})\} + \left(\frac{1}{2}\right)\log v_t(\boldsymbol{\theta}), \qquad M_n(\boldsymbol{\theta}) = \sum_{t=1}^{n} m_t(\boldsymbol{\theta}),$$

$$\hat{m}_t(\boldsymbol{\theta}) = \rho\{X_t/\hat{v}_t^{1/2}(\boldsymbol{\theta})\} + \left(\frac{1}{2}\right)\log \hat{v}_t(\boldsymbol{\theta}), \quad \text{and} \quad \hat{M}_n(\boldsymbol{\theta}) = \sum_{t=1}^{n} \hat{m}_t(\boldsymbol{\theta}).$$

Notice that

$$\dot{\boldsymbol{M}}_n(\boldsymbol{\theta}) = \sum_{t=1}^{n}\left(\frac{1}{2}\right)\{1 - H\{X_t/v_t^{1/2}(\boldsymbol{\theta})\}\}\{\dot{v}_t(\boldsymbol{\theta})/v_t(\boldsymbol{\theta})\}$$

https://doi.org/10.1017/S0266466608080602 Published online by Cambridge University Press

and

$$\dot{\hat{M}}_n(\boldsymbol{\theta}) = \sum_{t=1}^{n} \left(\frac{1}{2}\right)\{1 - H\{X_t/\hat{v}_t^{1/2}(\boldsymbol{\theta})\}\}\{\dot{\hat{v}}_t(\boldsymbol{\theta})/\hat{v}_t(\boldsymbol{\theta})\}.$$

Hence, in view of (A.16), $\boldsymbol{\theta}_n$ from (2.7) and $\hat{\boldsymbol{\theta}}_n$ from (2.8) can also be considered as (noncomputable or computable) "*M*-estimators," based on criterion functions $\boldsymbol{M}_n$ and $\hat{\boldsymbol{M}}_n$, respectively. Therefore, our proof is based on verifications of the conditions of Theorem KN for the criterion functions $\hat{\boldsymbol{M}}_n$. We do these in two steps. First we verify the conditions of Theorem KN for the criterion functions $\boldsymbol{M}_n$, and then we show that the difference between $\hat{\boldsymbol{M}}_n$ and $\boldsymbol{M}_n$ is small.

Condition (A.12) is verified for both $\boldsymbol{M}_n$ and $\hat{\boldsymbol{M}}_n$ using the differentiability of the score function H. To verify (A.13) for $\boldsymbol{M}_n$, we use the identifiability condition, whereas to verify the same for $\hat{\boldsymbol{M}}_n$, we use the smoothness properties of $\{c_j(\boldsymbol{\theta})\}$, Fact 4, and Fact 6. Model assumptions involving stationarity and ergodicity of $\{X_t\}$ and moment conditions are used in the verification of (A.14) for $\boldsymbol{M}_n$, and the smoothness conditions are used to control the difference between $\ddot{\hat{M}}_n(\boldsymbol{\theta}) - \ddot{M}_n(\boldsymbol{\theta})$. Finally, verification of (A.15) is long and is done directly for $\hat{\boldsymbol{M}}_n$; it uses Facts 1–3 to obtain different bounds on $\{v_t(\boldsymbol{\theta}) - \hat{v}_t(\boldsymbol{\theta})\}$ and their derivatives and also uses Fact 5.

Verification of (A.12). Note that

$$\dot{m}_t(\boldsymbol{\theta}) = h\{X_t/v_t^{1/2}(\boldsymbol{\theta})\}\{\dot{v}_t(\boldsymbol{\theta})/v_t(\boldsymbol{\theta})\}, \qquad \dot{\hat{m}}_t(\boldsymbol{\theta}) = h\{X_t/\hat{v}_t^{1/2}(\boldsymbol{\theta})\}\{\dot{\hat{v}}_t(\boldsymbol{\theta})/\hat{v}_t(\boldsymbol{\theta})\}, \quad \textbf{(A.19)}$$

$$\ddot{m}_t(\boldsymbol{\theta}) = \left(\frac{1}{4}\right) G\{X_t/v_t^{1/2}(\boldsymbol{\theta})\}\{\dot{v}_t(\boldsymbol{\theta})\dot{v}_t'(\boldsymbol{\theta})/v_t^2(\boldsymbol{\theta})\}$$

$$+ h\{X_t/v_t^{1/2}(\boldsymbol{\theta})\}\{v_t(\boldsymbol{\theta})\ddot{v}_t(\boldsymbol{\theta}) - \dot{v}_t(\boldsymbol{\theta})\dot{v}_t'(\boldsymbol{\theta})\}/v_t^2(\boldsymbol{\theta}),$$

and

$$\ddot{\hat{m}}_t(\boldsymbol{\theta}) = \left(\frac{1}{4}\right) G\{X_t/\hat{v}_t^{1/2}(\boldsymbol{\theta})\}\{\dot{\hat{v}}_t(\boldsymbol{\theta})\dot{\hat{v}}_t'(\boldsymbol{\theta})/\hat{v}_t^2(\boldsymbol{\theta})\}$$

$$+ h\{X_t/\hat{v}_t^{1/2}(\boldsymbol{\theta})\}\{\hat{v}_t(\boldsymbol{\theta})\ddot{\hat{v}}_t(\boldsymbol{\theta}) - \dot{\hat{v}}_t(\boldsymbol{\theta})\dot{\hat{v}}_t'(\boldsymbol{\theta})\}/\hat{v}_t^2(\boldsymbol{\theta}).$$

Hence (A.12) is satisfied. ■

Verification of (A.13). Here we consider the criterion function $\hat{M}_n(\boldsymbol{\theta})$ at $\boldsymbol{\alpha}_0 = \boldsymbol{\theta}_{0H}$ and show that $\dot{M}_n(\boldsymbol{\theta}_{0H})/n = o_p(1)$ and

$$\{\dot{\hat{M}}_n(\boldsymbol{\theta}) - \dot{M}_n(\boldsymbol{\theta})\}/n = o_p(1), \qquad \forall \boldsymbol{\theta} \in \Theta_0. \quad \textbf{(A.20)}$$

From (A.2)

$$X_t/\{v_t(\boldsymbol{\theta}_{0H})\}^{1/2} = X_t/\{c_H v_t(\boldsymbol{\theta}_0)\}^{1/2} = \epsilon_t/c_H^{1/2},$$

https://doi.org/10.1017/S0266466608080602 Published online by Cambridge University Press

and so we get from (A.19)

$$(1/n)\dot{\boldsymbol{M}}_n(\boldsymbol{\theta}_{0H}) = (1/n)\sum_{t=1}^{n}[h(\epsilon_t/c_H^{1/2})\{\dot{v}_t(\boldsymbol{\theta}_{0H})/v_t(\boldsymbol{\theta}_{0H})\}]. \tag{A.21}$$

Note that $\{h(\epsilon_t/c_H^{1/2})\{\dot{v}_t(\boldsymbol{\theta}_{0H})/v_t(\boldsymbol{\theta}_{0H})\}; t \geq 1\}$ is a stationary ergodic sequence of vectors. Moreover, the independence of $v_t(\boldsymbol{\theta}_{0H})$ and ϵ_t, $\forall t \in \mathcal{Z}$, (3.1), and (A.8) with $\nu = 1$ imply that the vectors have zero mean. Hence $\dot{M}_n(\boldsymbol{\theta}_{0H})/n \to \mathbf{0}$ a.s.

Because for vectors $\boldsymbol{v}, \hat{\boldsymbol{v}}$ and real numbers $h, \hat{h}$,

$$|h\boldsymbol{v} - \hat{h}\hat{\boldsymbol{v}}| \leq |h - \hat{h}|\,|\boldsymbol{v}| + |h|\,|\boldsymbol{v} - \hat{\boldsymbol{v}}| + |\boldsymbol{v} - \hat{\boldsymbol{v}}|\,|h - \hat{h}|,$$

we get from (A.19)

$$\begin{aligned}
|\dot{m}_t(\boldsymbol{\theta}) - \dot{\hat{m}}_t(\boldsymbol{\theta})| &\leq |h\{X_t/v_t^{1/2}(\boldsymbol{\theta})\} - h\{X_t/\hat{v}_t^{1/2}(\boldsymbol{\theta})\}|\,|\{\dot{v}_t(\boldsymbol{\theta})/v_t(\boldsymbol{\theta})\}| \\
&\quad + |h\{X_t/v_t^{1/2}(\boldsymbol{\theta})\}|\,|\{\dot{v}_t(\boldsymbol{\theta})/v_t(\boldsymbol{\theta})\} - \{\dot{\hat{v}}_t(\boldsymbol{\theta})/\hat{v}_t(\boldsymbol{\theta})\}| \\
&\quad + |\{\dot{v}_t(\boldsymbol{\theta})/v_t(\boldsymbol{\theta})\} - \{\dot{\hat{v}}_t(\boldsymbol{\theta})/\hat{v}_t(\boldsymbol{\theta})\}|\,|h\{X_t/v_t^{1/2}(\boldsymbol{\theta})\} - h\{X_t/\hat{v}_t^{1/2}(\boldsymbol{\theta})\}| \\
&= T_{1t}(\boldsymbol{\theta}) + T_{2t}(\boldsymbol{\theta}) + T_{3t}(\boldsymbol{\theta}), \quad \text{say.}
\end{aligned}$$

For T_{1t}, apply the mean value theorem on the function $g(v) = h\{X_t/v^{1/2}\}$ with derivative $\dot{g}(v) = (\frac{1}{2})(1/v)\{-X_t/v^{1/2}\}\dot{h}\{X_t/v^{1/2}\}$. Note that $-x\dot{h}(x) = G(x)/2$ and let $u_* \leq \hat{v}_t(\boldsymbol{\theta}) < \xi_t(\boldsymbol{\theta}) < v_t(\boldsymbol{\theta})$ denote the intermediate points of the mean value theorem. Then using (A.6) in the first inequality and (A.10) in the third inequality,

$$\begin{aligned}
T_{1t}(\boldsymbol{\theta}) &= \left|\left(\frac{1}{4}\right)\{v_t(\boldsymbol{\theta}) - \hat{v}_t(\boldsymbol{\theta})\}\{1/\xi_t(\boldsymbol{\theta})\}G\{X_t/\xi_t^{1/2}(\boldsymbol{\theta})\}\right|\,|\{\dot{v}_t(\boldsymbol{\theta})/v_t(\boldsymbol{\theta})\}| \\
&\leq C_1 Z_0 \rho^t\{|\dot{v}_t(\boldsymbol{\theta})/v_t(\boldsymbol{\theta})|\}|G\{X_t/\xi_t^{1/2}(\boldsymbol{\theta})\}| \\
&\leq C_1 Z_0 \rho^t|\{\dot{v}_t(\boldsymbol{\theta})/v_t(\boldsymbol{\theta})\}|\{|G\{X_t/\xi_t^{1/2}(\boldsymbol{\theta})\} - G(\epsilon_t/c_H^{1/2})| + |G(\epsilon_t/c_H^{1/2})|\} \\
&\leq CZ_0\rho^t\{|\dot{v}_t(\boldsymbol{\theta})/v_t(\boldsymbol{\theta})|\} \\
&\quad \times \big[|(\sigma_t c_H^{1/2})/\xi_t^{1/2}(\boldsymbol{\theta}) - 1|\big\{\{(\sigma_t c_H^{1/2})/\xi_t^{1/2}(\boldsymbol{\theta})\}|\epsilon_t/c_H^{1/2}|\Lambda(\epsilon_t/c_H^{1/2}) \\
&\qquad + |\epsilon_t/c_H^{1/2}|\,|\dot{H}(\epsilon_t/c_H^{1/2})|\big\} + |G(\epsilon_t/c_H^{1/2})|\big] \\
&\leq C_1\rho^t Z_0\{|\dot{v}_t(\boldsymbol{\theta})/v_t(\boldsymbol{\theta})|\}\big[(\sigma_t + C_2)\{\sigma_t|\epsilon_t/c_H^{1/2}|\Lambda(\epsilon_t/c_H^{1/2}) \\
&\qquad + 2|\epsilon_t/c_H^{1/2}|\,|\dot{H}(\epsilon_t/c_H^{1/2})|\}\big].
\end{aligned}$$

Using (A.8) and other moment conditions on Λ and $\dot{H}$, and so on, and finally (A.1), we get $\sum_{t=1}^{n} T_{1t}(\boldsymbol{\theta}) = O_p(1)$ and hence $n^{-1}\sum_{t=1}^{n} T_{1t}(\boldsymbol{\theta}) = o_p(1)$.

https://doi.org/10.1017/S0266466608080602 Published online by Cambridge University Press

For $T_{2t}(\boldsymbol{\theta})$ we first bound $|\{\dot{v}_t(\boldsymbol{\theta})/v_t(\boldsymbol{\theta})\} - \{\dot{\hat{v}}_t(\boldsymbol{\theta})/\hat{v}_t(\boldsymbol{\theta})\}|$. Because for vectors $\boldsymbol{w}, \hat{\boldsymbol{w}}$ and real numbers $v > 0, \hat{v} > 0$,

$$|(\boldsymbol{w}/v) - (\hat{\boldsymbol{w}}/\hat{v})| \le \{|\boldsymbol{w} - \hat{\boldsymbol{w}}|/\hat{v}\} + (1/\hat{v})|v - \hat{v}|\,|\boldsymbol{w}/v|,$$

we get using (A.6) and (A.7)

$$\begin{aligned}
&|\{\dot{v}_t(\boldsymbol{\theta})/v_t(\boldsymbol{\theta})\} - \{\dot{\hat{v}}_t(\boldsymbol{\theta})/\hat{v}_t(\boldsymbol{\theta})\}| \\
&\quad \le |\dot{v}_t(\boldsymbol{\theta}) - \dot{\hat{v}}_t(\boldsymbol{\theta})|/\hat{v}_t(\boldsymbol{\theta}) + \{1/\hat{v}_t(\boldsymbol{\theta})\}|v_t(\boldsymbol{\theta}) - \hat{v}_t(\boldsymbol{\theta})|\{|\dot{v}_t(\boldsymbol{\theta})|/v_t(\boldsymbol{\theta})\} \\
&\quad \le C\rho^t[Z_1 + Z_0\{|\dot{v}_t(\boldsymbol{\theta})|/v_t(\boldsymbol{\theta})\}]. \qquad \textbf{(A.22)}
\end{aligned}$$

Moreover, $|h\{X_t/v_t^{1/2}(\boldsymbol{\theta})\}|$ is bounded by

$$\begin{aligned}
&|h\{X_t/v_t^{1/2}(\boldsymbol{\theta})\} - h(\epsilon_t/c_H^{1/2})| + |h(\epsilon_t/c_H^{1/2})| \\
&\quad \le C[\{L(\epsilon_t/c_H^{1/2})(\sigma_t^2 + 1)\} + \{|H(\epsilon_t/c_H^{1/2})| + C_1\}].
\end{aligned}$$

Therefore

$$T_{2t}(\boldsymbol{\theta}) \le C\rho^t[Z_1 + Z_0\{|\dot{v}_t(\boldsymbol{\theta})|/v_t(\boldsymbol{\theta})\}]\{L(\epsilon_t/c_H^{1/2})(\sigma_t^2 + 1) + |H(\epsilon_t/c_H^{1/2}) + C_1|\},$$

and hence $\sum_{t=1}^n T_{2t}(\boldsymbol{\theta}) = O_p(1)$. Therefore $n^{-1}\sum_{t=1}^n T_{2t}(\boldsymbol{\theta}) = o_p(1)$. In a similar fashion, we can show that $n^{-1}\sum_{t=1}^n T_{3t}(\boldsymbol{\theta}) = o_p(1)$. Therefore (A.20) is satisfied. ■

Verification of (A.14). Note that

$$\begin{aligned}
\ddot{\boldsymbol{M}}_n(\boldsymbol{\theta}_{0H}) = \left(\frac{1}{4}\right)\sum_{t=1}^n &[\dot{H}(\epsilon_t/c_H^{1/2})(\epsilon_t/c_H^{1/2})\{\dot{v}_t(\boldsymbol{\theta}_{0H})\dot{v}_t'(\boldsymbol{\theta}_{0H})/v_t^2(\boldsymbol{\theta}_{0H})\}] \\
&+ \sum_{t=1}^n [h(\epsilon_t/c_H^{1/2})\{v_t(\boldsymbol{\theta}_{0H})\ddot{v}_t(\boldsymbol{\theta}_{0H}) - \dot{v}_t(\boldsymbol{\theta}_{0H})\dot{v}_t'(\boldsymbol{\theta}_{0H})\}/v_t^2(\boldsymbol{\theta}_{0H})].
\end{aligned}$$

Similar to (A.21), $\ddot{\boldsymbol{M}}_n(\boldsymbol{\theta}_{0H})/(2n) \to \boldsymbol{V}$, where

$$\boldsymbol{V} = \tilde{a}\boldsymbol{G} \quad \text{and} \quad \tilde{a} = \mathrm{E}\{(\epsilon/c_H^{1/2})\dot{H}(\epsilon/c_H^{1/2})\}/8 > 0, \qquad \textbf{(A.23)}$$

by (3.2). Moreover, similar to (A.20),

$$\{\ddot{\hat{\boldsymbol{M}}}_n(\boldsymbol{\theta}) - \ddot{\boldsymbol{M}}_n(\boldsymbol{\theta})\}/n = o_p(1), \quad \forall \boldsymbol{\theta} \in \Theta_0. \qquad \textbf{(A.24)}$$

Hence (A.14) is satisfied with $\boldsymbol{V} := \boldsymbol{V}_M = \tilde{a}\boldsymbol{G}$. ■

Verification of (A.15). Here, unlike using the triangulation by $M_n(\boldsymbol{\theta})$ as in (A.20) and (A.24), we will verify (A.15) directly for $\hat{M}_n(\boldsymbol{\theta})$. We will bound $(n\delta)^{-1}|\boldsymbol{T}_n(\boldsymbol{\theta})|$ by a sequence of random variables that are free from δ and of the form of the average of stationary random variables with finite expectation and hence are finite a.s.

Note that a δ-neighborhood of $\boldsymbol{\theta}_{0H}$ is $\mathcal{N}_\delta = \{\boldsymbol{\theta}_{0H} + \delta\boldsymbol{w}; \|\boldsymbol{w}\| \le 1\}$, $\delta > 0$. In what follows, the supremum is taken on $\boldsymbol{\theta} \in \mathcal{N}_\delta$.

https://doi.org/10.1017/S0266466608080602 Published online by Cambridge University Press

First, using the mean value theorem on the function $g(\boldsymbol{\theta}) = \hat{v}^{-1/2}(\boldsymbol{\theta})$ with intermediate points $\tilde{\boldsymbol{\theta}}_{0H}$,

$$\{1/\hat{v}_t(\boldsymbol{\theta})\}^{1/2} = \{1/\hat{v}_t(\boldsymbol{\theta}_{0H})\}^{1/2} - \left(\frac{1}{2}\right)(\boldsymbol{\theta} - \boldsymbol{\theta}_{0H})'\dot{\hat{v}}_t(\tilde{\boldsymbol{\theta}}_{0H})/\{\hat{v}_t(\tilde{\boldsymbol{\theta}}_{0H})\}^{3/2}.$$

Because the coefficients $\{c_j(\boldsymbol{\theta})\}$ are continuously differentiable functions,

$$\lim_{\boldsymbol{\theta}\to\boldsymbol{\theta}_{0H}} \hat{v}_t(\boldsymbol{\theta}) = \lim_{\boldsymbol{\theta}\to\boldsymbol{\theta}_{0H}} \left[c_0(\boldsymbol{\theta}) + \sum_{j=1}^{t-1} c_j(\boldsymbol{\theta})X_{t-j}^2\right] = \hat{v}_t(\boldsymbol{\theta}_{0H}).$$

Hence substituting $\boldsymbol{\theta} = \boldsymbol{\theta}_{0H} + \delta\boldsymbol{w}$, with $\|\boldsymbol{w}\| \le 1$,

$$\lim_{\delta\to 0+} \sup\{\delta^{-1}|\{1/\hat{v}_t(\boldsymbol{\theta})\}^{1/2} - \{1/\hat{v}_t(\boldsymbol{\theta}_{0H})\}^{1/2}|\} \le C|\dot{\hat{v}}_t(\boldsymbol{\theta}_{0H})|/\{\hat{v}_t(\boldsymbol{\theta}_{0H})\}^{3/2}. \tag{A.25}$$

Next

$$\begin{aligned}
&\ddot{\boldsymbol{M}}_n(\boldsymbol{\theta}) - \ddot{\boldsymbol{M}}_n(\boldsymbol{\theta}_{0H}) \\
&\quad= \left(\frac{1}{4}\right)\sum_{t=1}^{n}\left[\left\{G\{X_t/\hat{v}_t^{1/2}(\boldsymbol{\theta})\} - G\{X_t/\hat{v}_t^{1/2}(\boldsymbol{\theta}_{0H})\}\right\} \times \{\dot{\hat{v}}_t(\boldsymbol{\theta})\dot{\hat{v}}_t'(\boldsymbol{\theta})/\hat{v}_t^2(\boldsymbol{\theta})\}\right] \\
&\qquad+ \sum_{t=1}^{n}\left[G\{X_t/\hat{v}_t^{1/2}(\boldsymbol{\theta}_{0H})\}\left\{\{\dot{\hat{v}}_t(\boldsymbol{\theta})\dot{\hat{v}}_t'(\boldsymbol{\theta})/\hat{v}_t^2(\boldsymbol{\theta})\} - \{\dot{\hat{v}}_t(\boldsymbol{\theta}_{0H})\dot{\hat{v}}_t'(\boldsymbol{\theta}_{0H})/\hat{v}_t^2(\boldsymbol{\theta}_{0H})\}\right\}\right] \\
&\qquad+ \sum_{t=1}^{n}\left[\left\{h\{X_t/\hat{v}_t^{1/2}(\boldsymbol{\theta})\} - h\{X_t/\hat{v}_t^{1/2}(\boldsymbol{\theta}_{0H})\}\right\}\{\dot{\hat{v}}_t(\boldsymbol{\theta})\dot{\hat{v}}_t'(\boldsymbol{\theta}) - \hat{v}_t(\boldsymbol{\theta})\ddot{\hat{v}}_t(\boldsymbol{\theta})\}/\hat{v}_t^2(\boldsymbol{\theta})\right] \\
&\qquad+ \sum_{t=1}^{n}\Big[h\{X_t/\hat{v}_t^{1/2}(\boldsymbol{\theta}_{0H})\} \\
&\qquad\qquad\times\left\{\{\dot{\hat{v}}_t(\boldsymbol{\theta})\dot{\hat{v}}_t'(\boldsymbol{\theta}) - \hat{v}_t(\boldsymbol{\theta})\ddot{\hat{v}}_t(\boldsymbol{\theta})\}/\hat{v}_t^2(\boldsymbol{\theta})\right. \\
&\qquad\qquad\qquad\left.- \{\dot{\hat{v}}_t(\boldsymbol{\theta}_{0H})\dot{\hat{v}}_t'(\boldsymbol{\theta}_{0H}) - \hat{v}_t(\boldsymbol{\theta}_{0H})\ddot{\hat{v}}_t(\boldsymbol{\theta}_{0H})\}/\hat{v}_t^2(\boldsymbol{\theta}_{0H})\right\}\Big] \\
&\quad= \left(\frac{1}{4}\right)\boldsymbol{S}_1 + \boldsymbol{S}_2 + \boldsymbol{S}_3 + \boldsymbol{S}_4, \quad \text{say.}
\end{aligned}$$

To handle $\boldsymbol{S}_1 = \boldsymbol{S}_1(n,\delta)$, use (A.11) with $s_2 = c_H^{1/2}\sigma_t/\{\hat{v}_t(\boldsymbol{\theta}_{0H})\}^{1/2} = \{v_t(\boldsymbol{\theta}_{0H})/\hat{v}_t(\boldsymbol{\theta}_{0H})\}^{1/2}$ and $s_1 = c_H^{1/2}\sigma_t/\{\hat{v}_t(\boldsymbol{\theta})\}^{1/2}$ to get

$$\begin{aligned}
|\boldsymbol{S}_1| &\le c_H^{1/2}\sum_{t=1}^{n}\{|\{1/\hat{v}_t(\boldsymbol{\theta})\}^{1/2} - \{1/\hat{v}_t(\boldsymbol{\theta}_{0H})\}^{1/2}|\}\sigma_t \\
&\qquad\times\left[\left\{|c_H^{1/2}\sigma_t/\{\hat{v}_t(\boldsymbol{\theta})\}^{1/2} - 1|\,|\epsilon_t/c_H^{1/2}|\Lambda(\epsilon_t/c_H^{1/2})\right\} + \left\{|\epsilon_t/c_H^{1/2}|\,|\dot{H}(\epsilon_t/c_H^{1/2})|\right\}\right. \\
&\qquad\qquad\left.+ \left\{|\epsilon_t/c_H^{1/2}|\Lambda(\{\epsilon_t/c_H^{1/2}\}\{v_t(\boldsymbol{\theta}_{0H})/\hat{v}_t(\boldsymbol{\theta}_{0H})\}^{1/2})\right\}\right]\{|\dot{\hat{v}}_t(\boldsymbol{\theta})\dot{\hat{v}}_t'(\boldsymbol{\theta})/\hat{v}_t^2(\boldsymbol{\theta})|\}.
\end{aligned}$$

https://doi.org/10.1017/S0266466608080602 Published online by Cambridge University Press

Let $u_t = c_H^{1/2}\sigma_t/\{\hat{v}_t(\boldsymbol{\theta}_{0H})\}^{1/2} = \{v_t(\boldsymbol{\theta}_{0H})/\hat{v}_t(\boldsymbol{\theta}_{0H})\}^{1/2}$ and notice that

$$0 \le u_t^2 - 1 = \{v_t(\boldsymbol{\theta}_{0H}) - \hat{v}_t(\boldsymbol{\theta}_{0H})\}/\hat{v}_t(\boldsymbol{\theta}_{0H}) \le CZ_0\rho^t \tag{A.26}$$

and

$$0 \le u_t - 1 \le u_t^2 - 1 = \{v_t(\boldsymbol{\theta}_{0H}) - \hat{v}_t(\boldsymbol{\theta}_{0H})\}/\hat{v}_t(\boldsymbol{\theta}_{0H}) \le CZ_0\rho^t. \tag{A.27}$$

Now using (A.25),

$$\begin{aligned}
&n^{-1}\lim_{\delta\to 0+}\sup\{\delta^{-1}|\boldsymbol{S}_1|\}\\
&\quad\le Cn^{-1}\sum_{t=1}^{n}\Big[\big\{|\dot{\hat{v}}_t(\boldsymbol{\theta}_{0H})|/\{\hat{v}_t(\boldsymbol{\theta}_{0H})\}\big\}\big\{\sigma_t/\{\hat{v}_t(\boldsymbol{\theta}_{0H})\}^{1/2}\big\}\\
&\qquad\times\big\{\{|c_H^{1/2}\sigma_t/\{\hat{v}_t(\boldsymbol{\theta}_{0H})\}^{1/2} - 1|\,|\epsilon_t/c_H^{1/2}|\Lambda(\epsilon_t/c_H^{1/2})\}\\
&\qquad\quad+\{|\epsilon_t/c_H^{1/2}|\,|\dot{H}(\epsilon_t/c_H^{1/2})|\}\\
&\qquad\quad+\{|\epsilon_t/c_H^{1/2}|\Lambda(\{\epsilon_t/c_H^{1/2}\}\{v_t(\boldsymbol{\theta}_{0H})/\hat{v}_t(\boldsymbol{\theta}_{0H})\}^{1/2})\}\big\}\\
&\qquad\times\{|\dot{\hat{v}}_t(\boldsymbol{\theta}_{0H})\dot{\hat{v}}_t'(\boldsymbol{\theta}_{0H})/\hat{v}_t^2(\boldsymbol{\theta}_{0H})|\}\Big]\\
&\quad= Cn^{-1}\sum_{t=1}^{n}\Big[\big\{|\dot{\hat{v}}_t(\boldsymbol{\theta}_{0H})|/\{\hat{v}_t(\boldsymbol{\theta}_{0H})\}\big\}u_t\\
&\qquad\times\big\{\{|u_t - 1|\,|\epsilon_t/c_H^{1/2}|\Lambda(\epsilon_t/c_H^{1/2})\}\\
&\qquad\quad+\{|\epsilon_t/c_H^{1/2}|\,|\dot{H}(\epsilon_t/c_H^{1/2})|\}\big\}\times\{|\dot{\hat{v}}_t(\boldsymbol{\theta}_{0H})\dot{\hat{v}}_t'(\boldsymbol{\theta}_{0H})/\hat{v}_t^2(\boldsymbol{\theta}_{0H})|\}\Big]\\
&\quad\quad+Cn^{-1}\sum_{t=1}^{n}\Big[\{|\dot{\hat{v}}_t(\boldsymbol{\theta}_{0H})|/\hat{v}_t(\boldsymbol{\theta}_{0H})\}u_t\\
&\qquad\times\big\{|\epsilon_t/c_H^{1/2}|\Lambda(\{\epsilon_t/c_H^{1/2}\}u_t)\big\}\{|\dot{\hat{v}}_t(\boldsymbol{\theta}_{0H})\dot{\hat{v}}_t'(\boldsymbol{\theta}_{0H})/\hat{v}_t^2(\boldsymbol{\theta}_{0H})|\}\Big]\\
&\quad= S_{11} + S_{12}, \quad \text{say.}
\end{aligned}$$

Consider S_{11} first. As in (A.22),

$$|\dot{\hat{v}}_t(\boldsymbol{\theta}_{0H})|/\hat{v}_t(\boldsymbol{\theta}_{0H}) \le \{|\dot{v}_t(\boldsymbol{\theta}_{0H})|/v_t(\boldsymbol{\theta}_{0H})\} + C\rho^t[Z_1 + Z_0\{|\dot{v}_t(\boldsymbol{\theta}_{0H})|/v_t(\boldsymbol{\theta}_{0H})\}],$$

and in a similar fashion, one can get a bound on $|\dot{\hat{v}}_t(\boldsymbol{\theta}_{0H})\dot{\hat{v}}_t'(\boldsymbol{\theta}_{0H})/\hat{v}_t^2(\boldsymbol{\theta}_{0H})|$ in terms of $|\dot{v}_t(\boldsymbol{\theta}_{0H})\dot{v}_t'(\boldsymbol{\theta}_{0H})/v_t^2(\boldsymbol{\theta}_{0H})|$ plus additional terms involving ρ^t.

Next notice that

$$(C_1\rho^t + C_2)(C_3\rho^t + C_4) \le (C_5\rho^t + C_6)$$

for some large constants C_5 and C_6. Hence using (A.26) and (A.27), S_{11} can be bounded by averages of the variables of the form

https://doi.org/10.1017/S0266466608080602 Published online by Cambridge University Press

$$(Z_3\rho^t + Z_4)\{|\dot{v}_t(\boldsymbol{\theta}_{0H})|/v_t(\boldsymbol{\theta}_{0H})\}\{|\dot{v}_t(\boldsymbol{\theta}_{0H})\dot{v}_t'(\boldsymbol{\theta}_{0H})|/v_t^2(\boldsymbol{\theta}_{0H})\}$$
$$\times [\{|\epsilon_t/c_H^{1/2}|\Lambda(\epsilon_t/c_H^{1/2})\} + \{|\epsilon_t/c_H^{1/2}||\dot{H}(\epsilon_t/c_H^{1/2})|\}],$$

where Z_3 and Z_4 are finite a.s., $\{|\dot{v}_t(\boldsymbol{\theta}_{0H})|/v_t(\boldsymbol{\theta}_{0H})\}$ is independent of ϵ_t for each t, and the moments of $|\epsilon_t/c_H^{1/2}|\Lambda(\epsilon_t/c_H^{1/2})$, $\{|\epsilon_t/c_H^{1/2}||\dot{H}(\epsilon_t/c_H^{1/2})|\}$, and $\{|\dot{v}_t(\boldsymbol{\theta}_{0H})/v_t(\boldsymbol{\theta}_{0H})|^\nu\}$ exist. Taking random variables Z_3, Z_4 outside the summation and then taking expectation of the average of stationary random variables, we notice that the expectation is finite. Hence S_{11} is finite a.s.

Next consider $\boldsymbol{S}_{12}$. Using (3.8) in the second inequality and (A.27) in the third inequality,

$$|\boldsymbol{S}_{12}| \le (Z_5\rho^t + Z_6)\{|\dot{v}_t(\boldsymbol{\theta}_{0H})|/v_t(\boldsymbol{\theta}_{0H})\}\{|\dot{v}_t(\boldsymbol{\theta}_{0H})\dot{v}_t'(\boldsymbol{\theta}_{0H})|/v_t^2(\boldsymbol{\theta}_{0H})\}$$
$$\times \Big[\big\{|\epsilon_t/c_H^{1/2}|\{|\Lambda(\{\epsilon_t/c_H^{1/2}\}(1+(u_t-1))) - \Lambda(\epsilon_t/c_H^{1/2})|\}\big\}$$
$$+ |\epsilon_t/c_H^{1/2}|\Lambda(\epsilon_t/c_H^{1/2})\Big]$$
$$\le (Z_5\rho^t + Z_6)\{|\dot{v}_t(\boldsymbol{\theta}_{0H})|/v_t(\boldsymbol{\theta}_{0H})\}\{|\dot{v}_t(\boldsymbol{\theta}_{0H})\dot{v}_t'(\boldsymbol{\theta}_{0H})|/v_t^2(\boldsymbol{\theta}_{0H})\}$$
$$\times [|\epsilon_t/c_H^{1/2}|\Lambda^*(\epsilon_t/c_H^{1/2})[u_t - 1] + |\epsilon_t/c_H^{1/2}|\Lambda(\epsilon_t/c_H^{1/2})]$$
$$\le (Z_5\rho^t + Z_6)\{|\dot{v}_t(\boldsymbol{\theta}_{0H})|/v_t(\boldsymbol{\theta}_{0H})\}\{|\dot{v}_t(\boldsymbol{\theta}_{0H})\dot{v}_t'(\boldsymbol{\theta}_{0H})|/v_t^2(\boldsymbol{\theta}_{0H})\}$$
$$\times [|\epsilon_t/c_H^{1/2}|\Lambda^*(\epsilon_t/c_H^{1/2})\rho^t Z_0 + |\epsilon_t/c_H^{1/2}|\Lambda(\epsilon_t/c_H^{1/2})].$$

Hence S_{12} is finite a.s. A similar trick can be used to verify (A.15) for $\boldsymbol{S}_3$. To verify (A.15) for $\boldsymbol{S}_2$ we can get a δ-factor by applying the mean value theorem entrywise on the difference of matrices

$$\big\{\{\dot{\hat{v}}_t(\boldsymbol{\theta})\dot{\hat{v}}_t'(\boldsymbol{\theta})/\hat{v}_t^2(\boldsymbol{\theta})\} - \{\dot{\hat{v}}_t(\boldsymbol{\theta}_{0H})\dot{\hat{v}}_t'(\boldsymbol{\theta}_{0H})/\hat{v}_t^2(\boldsymbol{\theta}_{0H})\}\big\}$$

and then using triangulation, inequalities of the form (A.10), and expectation on the average of stationary terms. The same trick can be applied on $\boldsymbol{S}_4$.

Therefore, by Theorem KN(i), $\hat{\boldsymbol{\theta}}_n$ is consistent for $\boldsymbol{\theta}_{0H}$.

Now, using the independence of $v_t(\boldsymbol{\theta}_{0H})$ and ϵ_t and (3.1), it is easy to see that $\dot{\boldsymbol{M}}_n(\boldsymbol{\theta}_{0H})$ of (A.21) is a sum of the vectors of the martingale difference arrays. Hence using the martingale central limit theorem (see, e.g., Hall and Heyde, 1980), (A.17) is satisfied. To get $\boldsymbol{W}$ of (A.17), note that

$$(16n)^{-1}\sum_{t=1}^{n}\{\dot{v}_t(\boldsymbol{\theta}_{0H})\dot{v}_t'(\boldsymbol{\theta}_{0H})/v_t^2(\boldsymbol{\theta}_{0H})\}\{1 - H(\epsilon_t/c_H^{1/2})\}^2 \to \tilde{b}\boldsymbol{G}, \quad \text{a.s.,} \tag{A.28}$$

where $\tilde{b} = \mathrm{Var}\{H(\epsilon/c_H^{1/2})\}/16$. Hence $\boldsymbol{W} := \boldsymbol{W}_M = \tilde{b}\boldsymbol{G}$. Because $\tilde{b}/(\tilde{a})^2 = \sigma^2(H)$, we get from (A.23), (A.28), and (A.18),

$$n^{1/2}(\hat{\boldsymbol{\theta}}_n - \boldsymbol{\theta}_{0H}) \to N[\mathbf{0}, \sigma^2(H)\boldsymbol{G}^{-1}].$$

Hence (3.9) is proved. ■

https://doi.org/10.1017/S0266466608080602 Published online by Cambridge University Press

Proof of Proposition 3.1. The proof for the consistency of $\widehat{\sigma^2(H)}$ to $\sigma^2(H)$ follows by showing that the individual estimators converge. To illustrate, we only show that

$$(1/n)\sum_{t=1}^{n}\{H(X_t/\hat{v}_t^{1/2}(\hat{\boldsymbol{\theta}}_n))\}^2 = \mathrm{E}\{H(\epsilon/c_H^{1/2})\}^2 + o_p(1). \tag{A.29}$$

For details on the proof involving other terms, see Ndebu (2006). By the ergodic theorem, (A.29) follows by showing that

$$(1/n)\sum_{t=1}^{n}|\{H(X_t/\hat{v}_t^{1/2}(\hat{\boldsymbol{\theta}}_n))\}^2 - \{H(\epsilon_t/c_H^{1/2})\}^2| = o_p(1). \tag{A.30}$$

Using $|a^2 - b^2| \le |a-b|\{|a-b| + 2|b|\}$, the left-hand side of (A.30) is bounded by

$$(1/n)\sum_{t=1}^{n}|H\{(\epsilon_t/c_H^{1/2})s_t\} - H(\epsilon_t/c_H^{1/2})|$$
$$\times [|H\{(\epsilon_t/c_H^{1/2})s_t\} - H(\epsilon_t/c_H^{1/2})| + 2|H(\epsilon_t/c_H^{1/2})|],$$

where $X_t/\hat{v}_t^{1/2}(\hat{\boldsymbol{\theta}}_n) = (\epsilon_t/c_H^{1/2})s_t$ and $s_t = v_t^{1/2}(\boldsymbol{\theta}_{0H})/\hat{v}_t^{1/2}(\hat{\boldsymbol{\theta}}_n)$. Using the mean value theorem and (A.4) on the first term and (A.6) on the second term in the expression that follows,

$$|s_t^2 - 1| \le \{|v_t(\boldsymbol{\theta}_{0H}) - v_t(\hat{\boldsymbol{\theta}}_n)| + |v_t(\hat{\boldsymbol{\theta}}_n) - \hat{v}_t(\hat{\boldsymbol{\theta}}_n)|\}/\hat{v}_t(\hat{\boldsymbol{\theta}}_n)$$
$$\le C|\hat{\boldsymbol{\theta}}_n - \boldsymbol{\theta}_{0H}|\left\{C + \sum_{j=1}^{\infty}\rho^j X_{t-j}^2\right\} + \rho^t Z_0.$$

Hence from (3.3), the left-hand side of (A.30) is bounded by terms involving summations of the form $(1/n)\sum_{t=1}^{n}L(\epsilon_t/c_H^{1/2})\rho^t$, $|\hat{\boldsymbol{\theta}}_n - \boldsymbol{\theta}_{0H}|(1/n)\sum_{t=1}^{n}L(\epsilon_t/c_H^{1/2})\sum_{j=1}^{\infty}\rho^t X_{t-j}^2$, and so on. Terms involving $(1/n)\sum_{t=1}^{n}L(\epsilon_t/c_H^{1/2})\rho^t = o_p(1)$ using (A.1) as in the proof of $n^{-1}\sum_{t=1}^{n}T_{2t}(\boldsymbol{\theta}) = o_p(1)$. Terms involving $(1/n)\sum_{t=1}^{n}L(\epsilon_t/c_H^{1/2})\sum_{j=1}^{\infty}\rho^t X_{t-j}^2$ have bounded expectation using (3.10) and (3.11) and the independence between ϵ_t and X_{t-j}^2; hence (A.30) follows.

To prove the consistency of $\hat{\boldsymbol{G}}$ to $\boldsymbol{G}$, write $\hat{\boldsymbol{G}} - \boldsymbol{G}$ as

$$n^{-1}\sum_{t=1}^{n}[\{\dot{\hat{v}}_t(\hat{\boldsymbol{\theta}}_n)\dot{\hat{v}}_t'(\hat{\boldsymbol{\theta}}_n)/\hat{v}_t^2(\hat{\boldsymbol{\theta}}_n)\} - \{\dot{\hat{v}}_t(\boldsymbol{\theta}_{0H})\dot{\hat{v}}_t'(\boldsymbol{\theta}_{0H})/\hat{v}_t^2(\boldsymbol{\theta}_{0H})\}]$$
$$+ n^{-1}\sum_{t=1}^{n}[\{\dot{\hat{v}}_t(\boldsymbol{\theta}_{0H})\dot{\hat{v}}_t'(\boldsymbol{\theta}_{0H})/\hat{v}_t^2(\boldsymbol{\theta}_{0H})\} - \{\dot{v}_t(\boldsymbol{\theta}_{0H})\dot{v}_t'(\boldsymbol{\theta}_{0H})/v_t^2(\boldsymbol{\theta}_{0H})\}] + o_p(1).$$

The first term is $o_p(1)$ as in the verification of (A.15), and the second term is $o_p(1)$ as in the verification of (A.13). For details, see Ndebu (2006). Hence the proof of the proposition is complete. ■

https://doi.org/10.1017/S0266466608080602 Published online by Cambridge University Press