Econometric Theory, **20**, 2004, 1203–1226. Printed in the United States of America.
DOI: 10.1017/S0266466604206065

ASYMPTOTIC INFERENCE FOR NONSTATIONARY GARCH

Søren Tolver Jensen and Anders Rahbek
University of Copenhagen

Consistency and asymptotic normality are established for the highly applied quasi-maximum likelihood estimator in the GARCH(1,1) model. Contrary to existing literature we allow the parameters to be in the region where no stationary version of the process exists. This has the important implication that the likelihood-based estimator for the GARCH parameters is consistent and asymptotically normal in the entire parameter region including both stationary and explosive behavior. In particular, there is no "knife edge result like the unit root case" as hypothesized in Lumsdaine (1996, *Econometrica* 64, 575–596).

1. INTRODUCTION

This paper considers the asymptotic behavior of the likelihood-based estimators in the generalized autoregressive conditional heteroskedastic (GARCH) model or better the "workhorse of the industry" (Lee and Hansen, 1994). The GARCH(1,1) or simply the GARCH model is given by

$$y_t = \sqrt{h_t(\theta)}z_t, \tag{1}$$

$$h_t(\theta) = \omega + \alpha y_{t-1}^2 + \beta h_{t-1}(\theta), \tag{2}$$

with $t = 1,\dots,T$ and z_t an independent and identically distributed (i.i.d.) (0,1) sequence. As to initial values the analysis is conditional on the observed value y_0, whereas the unobserved variance, $h_0(\theta)$, is parametrized by γ, $h_0(\theta) = \gamma$. The parameter θ of the GARCH model is therefore

$$\theta = (\alpha,\beta,\omega,\gamma) \tag{3}$$

with $\alpha,\beta,\omega,$ and γ all positive. Denote henceforth the positive true parameter values by $\theta_0 = (\alpha_0,\beta_0,\omega_0,\gamma_0)$.

The GARCH model was introduced by Bollerslev (1986), extending the autoregressive conditional heteroskedastic (ARCH) model of Engle (1982). Asymp-

Anders Rahbek is grateful for support from the Danish Social Sciences Research Council, the Centre for Analytical Finance (CAF), and the EU network DYNSTOCH. Both authors thank the two anonymous referees and the editor for highly valuable and detailed comments that have, we believe, led to a much improved version of the paper, both in terms of the econometric theory and of the presentation. Address correspondence to Anders Rahbek, Department of Applied Mathematics and Statistics, University of Copenhagen, Universitetsparken 5, DK-2100 Copenhagen Ø, Denmark; e-mail: rahbek@math.ku.dk.

 0266-4666/04 $12.00

totic inference for the ARCH and GARCH type models has been studied in, e.g., Kristensen and Rahbek (2002), Lee and Hansen (1994), Lumsdaine (1996), and Weiss (1986). Common to these is the assumption that the (G)ARCH process y_t is suitably ergodic or stationary such that appropriate laws of large numbers apply. Moreover the generic assumption for asymptotic normality is that the squared error process, z_t^2, has a finite (conditional) variance, $\kappa = V(z_t^2) = E(z_t^2 - 1) < \infty$. In the case of i.i.d. innovations z_t the results in Lee and Hansen (1994) establish asymptotic normality essentially under the assumption that

$$E \log(\alpha_0 z_t^2 + \beta_0) < 0. \tag{4}$$

This condition is necessary and sufficient for stationarity of the GARCH process as argued in Nelson (1990) and Bougerol and Picard (1992). Recall that the assumption in (4) is implied by the well-known sufficient condition $\alpha_0 + \beta_0 \leq 1$, which includes the much studied case of integrated GARCH where $\alpha_0 + \beta_0 = 1$.

Our contribution is to relax this and work under the following assumption, which permits explosive and nonstationary behavior of the GARCH process.

Assumption 1. Assume that with z_t i.i.d.(0,1), the true parameters satisfy

$$E \log(\alpha_0 z_t^2 + \beta_0) \geq 0. \tag{5}$$

Clearly this extends the parameter region for which asymptotic normality holds. Our results show that whether or not the parameters are such that the process is ergodic, integrated, or even explosive, asymptotic normality of the likelihood-based estimators applies. Thus there is no “knife edge result like the unit root case” when entering the parameter region in Assumption 1 as hypothesized in Lumsdaine (1996, p. 580). Indeed our results imply, in particular, that requirements for existence of moments and stationarity for the GARCH process can be ignored when reporting, e.g., standard deviations and test statistics involving the likelihood-based estimators here, which until now has caused concern in the literature on GARCH inference. To this end unreported simulations indicate that in fact the convergence of the estimators to the Gaussian distribution is faster in the explosive case than in the stationary. Note that Jensen and Rahbek (2004) relax the condition about stability of the y_t process in the ARCH(1) model, where $\beta = 0$, and allow the ARCH process to be nonstationary and to have no moments. The added complexity here due to the parameter β and hence lagged variance, $h_{t-1}(\theta)$, in (2) implies that results regarding inference require different types of arguments when compared to the ARCH model. This is also noted by Lee and Hansen (1994, p. 35) for the stationary case, where it is emphasized that inference with respect to β is the most difficult.

The paper is structured as follows. Section 2 presents the two main theorems of the paper. Theorem 1 establishes asymptotic normality when the parameter that parametrizes the initial unobserved variance $h_0(\theta) = \gamma$ is set equal to the true value, $\gamma = \gamma_0$, and, furthermore, the scale parameter equals its true value, $\omega = \omega_0$. Theorem 2 shows that the asymptotics hold independently of this choice, that is, independently of the initial values. Sections 3 and 4 establish the proofs of Theorems 1 and 2, respectively.

2. MAIN THEOREMS

As in Lee and Hansen (1994) and most of the literature, we consider the likelihood estimators based on minimization of

$$\ell_T(\theta) = \frac{1}{T}\sum_{t=1}^{T}\left[\log h_t(\theta) + \frac{y_t^2}{h_t(\theta)}\right] \tag{6}$$

with $h_t(\theta)$ defined in (2). Throughout this is referred to as the (quasi-)likelihood function, and likewise the first and second derivatives are referred to as the score and observed information, respectively. Note that it is the true log-likelihood function (multiplied by minus two) if z_t is indeed Gaussian. Our first main result is the following.

THEOREM 1. *With (ω,γ) fixed at their true values, (ω_0,γ_0), consider the model given by the (quasi-)likelihood function $\ell_T(\alpha,\beta) := \ell_T(\alpha,\beta,\omega_0,\gamma_0)$ as given by (6). Assume that at the true parameter $\theta_0 = (\alpha_0,\beta_0,\omega_0,\gamma_0)$, y_t given by (1) satisfies Assumption 1 such that no stationary version exists. Assume further that for the i.i.d.(0,1) process z_t, $V(z_t^2) = \kappa < \infty$.*

Under these assumptions there exists a fixed open neighborhood $U = U(\alpha_0,\beta_0)$ of (α_0,β_0) such that with probability tending to one as $T \to \infty$, $\ell_T(\alpha,\beta)$ has a unique minimum point $(\hat\alpha_T,\hat\beta_T)$ in U. Furthermore, $(\hat\alpha_T,\hat\beta_T)$ is consistent and asymptotically Gaussian,

$$\sqrt{T}[(\hat\alpha_T,\hat\beta_T) - (\alpha_0,\beta_0)]' \xrightarrow{D} N(0,\Omega).$$

Here $\Omega > 0$ and is given by $\Omega = \kappa\Sigma^{-1}$, with $\mu_i = E(\beta_0/(\alpha_0 z_t^2 + \beta_0))^i$, $i = 1,2$, and

$$\Sigma = \begin{pmatrix} \dfrac{1}{\alpha_0^2} & \dfrac{\mu_1}{\alpha_0\beta_0(1-\mu_1)} \\ \dfrac{\mu_1}{\alpha_0\beta_0(1-\mu_1)} & \dfrac{(1+\mu_1)\mu_2}{\beta_0^2(1-\mu_1)(1-\mu_2)} \end{pmatrix}. \tag{7}$$

Remark 1. The covariance matrix Ω in Theorem 1 provides a lower bound for the implicitly given variance in the stationary and ergodic case analyzed in Lee and Hansen (1994, Theorem 3). This follows by the fact that the information matrix provides an upper bound as seen by the proof of Lemma 6 in Section 3.3, which applies the inequalities in Lemma 4. These inequalities hold independently of whether the process y_t is stationary or not. Similarly for the two-dimensional information discussed in Section 3.5.

Remark 2. Note that if z_t is Gaussian explicit expressions for μ_1 and μ_2 can be computed for the covariance matrix in Theorem 1.

Next, we state the result that the values of $h_0(\theta) = \gamma$ and ω (and also the initial value y_0) are asymptotically negligible. In particular, $(\hat{\alpha}_T, \hat{\beta}_T)$ is a minimum point in U of $\ell_T(\theta) = \ell_T(\alpha, \beta, \omega, \gamma)$ for any arbitrary values of ω and γ.

THEOREM 2. *Assume that* $E \log(\alpha_0 z_t^2 + \beta_0) > 0$. *Then the results in Theorem 1 hold for any arbitrary values of* $\gamma > 0$ *and* $\omega > 0$.

Theorem 2 states that (α, β) can be estimated consistently by taking any arbitrary values of ω and γ and that the asymptotic distribution of the estimator does not depend on the arbitrary values of ω and γ. After the parameter (α, β) has been (consistently) estimated, one may estimate (ω, γ), but the estimator would not be consistent, and hence there is no need to reestimate the parameter (α, β).

Remark 3. Note that the results in Theorem 2 exclude the boundary case in Assumption 1 of $E \log(\alpha_0 z_t^2 + \beta_0) = 0$. We do not know whether there exists a consistent estimator for ω or if the asymptotic distribution of $\hat{\alpha}$ and $\hat{\beta}$ does not depend on ω in this case.

3. PROOF OF THEOREM 1

The proof of Theorem 1 is based on applying Lemma 1, which follows. Note that conditions (A.1)–(A.3) are similar to conditions stated in the literature on asymptotic likelihood-based inference (see, e.g., Lehmann, 1999; Basawa, Feigin, and Heyde, 1976). The difference is that (A.1)–(A.3) explicitly exploit the fact that (the negative) log-likelihood function is three times continuously differentiable in the parameter. Furthermore, Lemma 1 establishes uniqueness (convexity) in addition to existence of the consistent and asymptotically Gaussian estimator.

LEMMA 1. *Consider* $\ell_T(\varphi)$, *which is a function of the observations* $X_1, \ldots, X_T$ *and the parameter* $\varphi \in \Phi \subseteq \mathbb{R}^k$. *Introduce furthermore* φ_0, *which is an interior point of* Φ. *Assume that* $\ell_T(\cdot) : \mathbb{R}^k \to \mathbb{R}$ *is three times continuously differentiable in* φ *and that*

(A.1) *As* $T \to \infty$, $\sqrt{T}\,\partial \ell_T(\varphi_0)/\partial \varphi \xrightarrow{D} N(0, \Omega_S)$, $\Omega_S > 0$.
(A.2) *As* $T \to \infty$, $\partial^2 \ell_T(\varphi_0)/\partial \varphi \partial \varphi' \xrightarrow{P} \Omega_I > 0$.
(A.3) $\max_{h,i,j=1,\ldots,k} \sup_{\varphi \in N(\varphi_0)} |\partial^3 \ell_T(\varphi)/\partial \varphi_h \partial \varphi_i \partial \varphi_j| \leq c_T$,

where $N(\varphi_0)$ is a neighborhood of φ_0 and $0 \leq c_T \xrightarrow{P} c$, $0 < c < \infty$. Then there exists a fixed open neighborhood $U(\varphi_0) \subseteq N(\varphi_0)$ of φ_0 such that

(B.1) With probability tending to one as $T \to \infty$, there exists a minimum point $\hat{\varphi}_T$ of $\ell_T(\varphi)$ in $U(\varphi_0)$ and $\ell_T(\varphi)$ is convex in $U(\varphi_0)$. In particular, $\hat{\varphi}_T$ is unique and solves $\partial \ell_T(\hat{\varphi}_T)/\partial\varphi = 0$.
(B.2) As $T \to \infty$, $\hat{\varphi}_T \xrightarrow{P} \varphi_0$.
(B.3) As $T \to \infty$, $\sqrt{T}(\hat{\varphi}_T - \varphi_0) \xrightarrow{D} N(0, \Omega_I^{-1}\Omega_S\Omega_I^{-1})$.

The proof of Lemma 1 is given in the Appendix.

Next, with $\varphi := (\alpha, \beta)'$ and $\ell_T(\varphi) = \ell_T(\alpha, \beta)$ defined in Theorem 1, the results in Theorem 1 follow by establishing conditions (A.1)–(A.3) in Lemma 1. For exposition only we initially focus on the GARCH parameter β in Sections 3.1–3.4. The first-, second-, and third-order derivatives of the likelihood function with respect to β are given in Section 3.1. Upon some initial results in Section 3.1, the behavior of the score and observed information evaluated at the true value, $\theta = \theta_0$, are studied in Section 3.3. In Section 3.4 it is shown that the third derivative is uniformly bounded by a suitably integrable majorant. The derivations concerning the ARCH parameter α are simple when compared with the ones with respect to β and are outlined in Section 3.5. It is also there that the asymptotic results for the joint parameter (α, β) are given. Note finally that (A.1)–(A.3) hold by Lemmas 5, 6, and 10 in Sections 3.3 and 3.4, together with the comments in Section 3.5. It follows that $\Omega_S = \kappa\Sigma$ in (A.1) (see Lemma 5 and (38)), whereas in (A.2), $\Omega_I = \Sigma$ (see Lemma 6 and (39)). Finally, Σ is given by (40).

3.1. Variation with Respect to β

In this section we derive the first-, second-, and third-order derivatives of the likelihood function with respect to β.

The likelihood function is given by (6) in terms of θ, and it follows that

$$\frac{\partial \ell_T}{\partial \beta}(\theta) = \frac{1}{T}\sum_{t=1}^{T}\left[1 - \frac{y_t^2}{h_t(\theta)}\right]h_{1t}(\theta), \tag{8}$$

$$\frac{\partial^2 \ell_T}{\partial \beta^2}(\theta) = \frac{1}{T}\sum_{t=1}^{T}\left[1 - \frac{y_t^2}{h_t(\theta)}\right]h_{2t}(\theta) + \frac{1}{T}\sum_{t=1}^{T}\left[2\frac{y_t^2}{h_t(\theta)} - 1\right]h_{1t}^2(\theta), \tag{9}$$

$$\begin{aligned}\frac{\partial^3 \ell_T}{\partial \beta^3}(\theta) = {} & \frac{1}{T}\sum_{t=1}^{T}\left[1 - \frac{y_t^2}{h_t(\theta)}\right]h_{3t}(\theta) + \frac{1}{T}\sum_{t=1}^{T}3\left[2\frac{y_t^2}{h_t(\theta)} - 1\right]h_{1t}(\theta)h_{2t}(\theta) \\ & + \frac{1}{T}\sum_{t=1}^{T}2\left[1 - 3\frac{y_t^2}{h_t(\theta)}\right]h_{1t}^3(\theta).\end{aligned} \tag{10}$$

Here, applying simple recursions,

$$h_{1t}(\theta) := \frac{\partial h_t(\theta)/\partial\beta}{h_t(\theta)} = \sum_{j=1}^{t} \beta^{j-1} \frac{h_{t-j}(\theta)}{h_t(\theta)}, \tag{11}$$

$$h_{2t}(\theta) := \frac{\partial^2 h_t(\theta)/\partial\beta^2}{h_t(\theta)} = 2\sum_{j=1}^{t} (j-1)\beta^{j-2} \frac{h_{t-j}(\theta)}{h_t(\theta)}, \quad \text{and} \tag{12}$$

$$h_{3t}(\theta) := \frac{\partial^3 h_t(\theta)/\partial\beta^3}{h_t(\theta)} = 3\sum_{j=1}^{t} \beta^{j-3}(j-1)(j-2) \frac{h_{t-j}(\theta)}{h_t(\theta)}. \tag{13}$$

3.2. Some Initial Results

For the asymptotic likelihood analysis in the following discussion the first and the second derivatives of the likelihood function are evaluated at the true value θ_0, and we introduce therefore the notation

$$h_t := h_t(\theta_0), \qquad h_{it} := h_{it}(\theta_0) \quad \text{and} \quad \frac{\partial^i \ell_T}{\partial \beta^i} := \frac{\partial^i \ell_T}{\partial \beta^i}(\theta)\bigg|_{\theta=\theta_0} \quad \text{for } i = 1,2,3. \tag{14}$$

Underlying parts of the asymptotics is first of all the following observation from Nelson (1990, Theorem 2).

LEMMA 2. *Under Assumption 1, as $t \to \infty$,*

$$h_t \xrightarrow{\text{a.s.}} \infty.$$

Next, to study the asymptotics of the central quantities h_{1t}, h_{2t}, and h_{3t} in (11)–(13) it is useful to introduce the stationary processes $u_{it}(\cdot)$ for $i = 1,\dots,4$ defined in terms of the i.i.d. innovations z_t. Note that the processes and their properties are well defined for the entire parameter region. In particular, Assumption 1 is not required in the lemma.

LEMMA 3. *Define the processes*

$$u_{mt}(a,b) = m\sum_{j=1}^{\infty} a^{j-m} \prod_{n=1}^{m-1} (j-n) \prod_{k=1}^{j} \frac{1}{\alpha_0 z_{t-k}^2 + b} \tag{15}$$

for $m = 1,\dots,4$ and with the notational convention that $\prod_{n=1}^{0} = 1$. For all $p \geq 1$ and $m = 1,\dots,4$, $u_{mt} := u_{mt}(\beta_0,\beta_0)$ is ergodic and

$$Eu_{mt}^{p} < \infty. \tag{16}$$

Furthermore, for each $p \geq 1$ there exist β_L and β_U with $\beta_L < \beta_0 < \beta_U$ such that $u_{mt}(\beta_0, \beta_L)$ and $u_{mt}(\beta_U, \beta_0)$ are ergodic and

$$E[u_{mt}(\beta_0,\beta_L)]^p < \infty \quad \text{and} \quad E[u_{mt}(\beta_U,\beta_0)]^p < \infty. \tag{17}$$

Proof of Lemma 3. Without loss of generality consider the case of $m = 2$. Define

$$q_p := E\left(\frac{\beta_0}{\alpha_0 z_1^2 + \beta_0}\right)^p < 1 \tag{18}$$

as z_t is nondegenerate. Using Minkowski's inequality, (16) follows by

$$\begin{aligned}[E(u_{2t}^p)]^{1/p} &\leq 2\sum_{j=2}^{\infty}\frac{1}{\beta_0^2}(j-1)\left[E\left(\prod_{k=1}^{j}\frac{\beta_0}{\alpha_0 z_{t-k}^2 + \beta_0}\right)^p\right]^{1/p}\\ &= \frac{2}{\beta_0^2}\sum_{j=2}^{\infty}(j-1)q_p^{j/p} = \frac{2q_p^{2/p}}{\beta_0^2(1-q_p^{1/p})^2} < \infty.\end{aligned}$$

Next, consider, say, $u_{2t}(\beta_0,\beta_L)$:

$$u_{2t}(\beta_0,\beta_L) = 2\sum_{j=1}^{\infty}\frac{1}{\beta_0^2}(j-1)\prod_{k=1}^{j}\frac{\beta_0}{\alpha_0 z_{t-k}^2 + \beta_L}.$$

Then as before, $E[u_{2t}(\beta_0,\beta_L)]^p < \infty$, provided

$$E\left(\frac{\beta_0}{\alpha_0 z_1^2 + \beta_L}\right)^p < 1,$$

which is the case for some $\beta_L < \beta_0$ (as the innovations z_t are nondegenerate). Likewise for $u_{2t}(\beta_U,\beta_0)$, which ends the proof. ■

Next, we show how the h_{it} and u_{it} are related.

LEMMA 4. *Consider h_{1t} and h_{2t} defined by (11) and (12), respectively, with the notational convention in (14). Then for $i = 1,2$,*

$$0 \leq h_{it} \leq u_{it}, \tag{19}$$

where the u_{it} are defined in Lemma 3. Furthermore, under Assumption 1, for $i = 1,2$, then as $t \to \infty$,

$$h_{it} - u_{it} \xrightarrow{L^p} 0, \tag{20}$$

$$\frac{1}{T}\sum_{t=1}^{T}(h_{1t}^2 - u_{1t}^2) \xrightarrow{L^p} 0 \quad \text{and} \quad \frac{1}{T}\sum_{t=1}^{T}(h_{2t} - u_{2t}) \xrightarrow{L^p} 0 \tag{21}$$

for all $p \geq 1$.

It is important that the inequality in (19) holds independently of Assumption 1.

Proof of Lemma 4. From the recursions in (11) note that

$$h_{1t} = \sum_{j=1}^{t} \beta^{j-1} \prod_{k=1}^{j} \frac{h_{t-k}}{h_{t-k+1}}. \tag{22}$$

Next, use that

$$\frac{\beta_0 h_{t-k}}{\omega_0 + (\alpha_0 z_{t-k}^2 + \beta_0) h_{t-k}} \le \frac{\beta_0}{\alpha_0 z_{t-k}^2 + \beta_0} \le 1 \tag{23}$$

to establish the desired inequality,

$$h_{1t} = \sum_{j=1}^{t} \frac{1}{\beta_0} \prod_{k=1}^{j} \frac{\beta_0 h_{t-k}}{\omega_0 + (\alpha_0 z_{t-k}^2 + \beta_0) h_{t-k}} \le \sum_{j=1}^{t} \frac{1}{\beta_0} \prod_{k=1}^{j} \frac{\beta_0}{\alpha_0 z_{t-k}^2 + \beta_0} \le u_{1t}.$$

Likewise,

$$h_{2t} = 2 \sum_{j=2}^{t} (j-1) \beta^{j-2} \prod_{k=1}^{j} \frac{h_{t-k}}{h_{t-k+1}}, \tag{24}$$

which shows (19).

Turn to h_{1t} again. By (23) and Lemma 2, then as $t \to \infty$,

$$\frac{\beta_0}{(\alpha_0 z_{t-k}^2 + \beta_0)} - \frac{\beta_0 h_{t-k}}{\omega_0 + (\alpha_0 z_{t-k}^2 + \beta_0) h_{t-k}} \xrightarrow{a.s.} 0,$$

and therefore

$$1 \ge \prod_{k=1}^{j} \frac{\beta_0}{(\alpha_0 z_{t-k}^2 + \beta_0)} - \frac{\beta_0^j h_{t-j}}{h_t} \xrightarrow{a.s.} 0. \tag{25}$$

By dominated convergence also L^1 convergence holds in (25). Now let $t_0 < t$ be arbitrary and consider

$$\limsup_{t\to\infty} E(u_{1t} - h_{1t}) \le \limsup_{t\to\infty} \sum_{j=1}^{t_0} \frac{1}{\beta_0} E\left[\prod_{k=1}^{j} \left(\frac{\beta_0}{\alpha_0 z_{t-j}^2 + \beta_0} \right) - \frac{\beta_0^j h_{t-j}}{h_t} \right] + \limsup_{t\to\infty} \sum_{j=t_0+1}^{\infty} \frac{1}{\beta_0} \prod_{k=1}^{j} E\left(\frac{\beta_0}{\alpha_0 z_{t-j}^2 + \beta_0} \right).$$

With t_0 fixed, then as $t \to \infty$ the first term equals zero by the just established L^1 convergence. The second term equals $(1/\beta_0)(q_1^{t_0+1}/(1 - q_1))$ with q_1 defined in (18). As t_0 was arbitrary, the second term then tends to zero as $t_0 \to \infty$ because $q_1 < 1$. Hence $E(u_{1t} - h_{1t})$ tends to zero as $t \to \infty$. Next, note that

$$u_{1t}^p \ge (u_{1t} - h_{1t})^p$$

implies that the latter is uniformly integrable because the distribution of u_{1t} does not depend on t. The just established L^1 convergence implies convergence in probability of $(u_{1t} - h_{1t})^p$ and hence by the uniform integrability,

$$E(u_{1t} - h_{1t})^p \to 0.$$

Likewise for u_{2t} and h_{2t}, which establishes (20). Finally, the claimed L^p convergence in (21) follows by

$$\left(E\left|\frac{1}{T}\sum_{t=1}^{T}(u_{1t}^2 - h_{1t}^2)\right|^p\right)^{1/p} = \left(E\left|\frac{1}{T}\sum_{t=1}^{T}(u_{1t} + h_{1t})(u_{1t} - h_{1t})\right|^p\right)^{1/p}$$

$$\leq \frac{2}{T}\sum_{t=1}^{T}(E(u_{1t}(u_{1t} - h_{1t}))^p)^{1/p} \leq \frac{2}{T}(Eu_{1t}^{2p})^{1/2p}\sum_{t=1}^{T}(E(u_{1t} - h_{1t})^{2p})^{1/2p},$$

which tends to zero as $E(u_{1t} - h_{1t})^{2p}$ in particular tends to zero. This ends the proof of Lemma 4. ■

3.3. Asymptotics of the Score and Observed Information

This section establishes asymptotic normality of the score and convergence of the observed information in probability under the true value, θ_0. As noted, the idea is, asymptotically, to replace the h_{it} terms with the corresponding u_{it} terms in the expressions (8) and (9), respectively; see also Lemma 4.

Consider first the score.

LEMMA 5. *Under Assumption 1 the score given by (8) evaluated at $\theta = \theta_0$ is asymptotically Gaussian,*

$$\sqrt{T}\,\frac{\partial \ell_T}{\partial \beta} = \frac{1}{\sqrt{T}}\sum_{t=1}^{T}\left[1 - \frac{y_t^2}{h_t}\right]h_{1t} \xrightarrow{D} N(0, \kappa\varpi),$$

$$\varpi = Eu_{1t}^2 = \frac{(1+\mu_1)\mu_2}{\beta_0^2(1-\mu_1)(1-\mu_2)} \quad \text{and} \quad \kappa = V(z_t^2),$$

where u_{1t} is given by (15) and $\mu_i = E(\beta_0/(\alpha_0 z_t^2 + \beta_0))^i$, $i = 1,2$.

Proof of Lemma 5. Evaluated at $\theta = \theta_0$ the score is given by

$$\frac{\partial \ell_T}{\partial \beta} = \frac{1}{T}\sum_{t=1}^{T}[1 - z_t^2]h_{1t} = \frac{1}{T}\sum_{t=1}^{T} v_t$$

such that $E(v_t|\mathcal{F}_{t-1}) = 0$, where $\mathcal{F}_t = \sigma(z_t, z_{t-1}, \ldots)$. Applying the central limit theorem for martingale differences in Brown (1971), consider first

$$\frac{1}{T}\sum_{t=1}^{T} E(v_t^2|\mathcal{F}_{t-1}) = \kappa\frac{1}{T}\sum_{t=1}^{T} h_{1t}^2 = \kappa\frac{1}{T}\sum_{t=1}^{T}(h_{1t}^2 - u_{1t}^2) + \kappa\frac{1}{T}\sum_{t=1}^{T} u_{1t}^2$$
$$\to \kappa E(u_{1t}^2) > 0$$

in probability (and L^1) as $T \to \infty$, using Lemma 4 and the ergodic theorem. Turning to the Lindeberg condition, as $h_{1t} \leq u_{1t}$,

$$\frac{1}{T}\sum_{t=1}^{T} E(v_t^2 1\{|v_t^2| > \delta\sqrt{T}\}) \leq \frac{1}{T}\sum_{t=1}^{T} E((1-z_t^2)^2 u_{1t}^2 1\{|(1-z_t^2)^2 u_{1t}^2| > \delta\sqrt{T}\})$$
$$= E((1-z_t^2)^2 u_{1t}^2 1\{|(1-z_t^2)^2 u_{1t}^2| > \delta\sqrt{T}\}) \to 0$$

for all δ as $T \to \infty$ because u_{1t} (and also z_t^2) is stationary and has finite second-order moments. ∎

Next we establish the asymptotic limit of the observed information.

LEMMA 6. *Under Assumption 1 the observed information given by (9) evaluated at $\theta = \theta_0$ converges in probability,*

$$\frac{\partial^2 \ell_T}{\partial\beta^2} \xrightarrow{P} \varpi = Eu_{1t}^2 > 0,$$

where ϖ is given in Lemma 5.

Proof of Lemma 6. For $\theta = \theta_0$ the observed information is given by

$$\frac{\partial^2 \ell_T}{\partial\beta^2} = \frac{1}{T}\sum_{t=1}^{T}[1-z_t^2]h_{2t} + \frac{1}{T}\sum_{t=1}^{T}[2z_t^2 - 1]h_{1t}^2$$
$$= \frac{1}{T}\sum_{t=1}^{T}[1-z_t^2](h_{2t} - u_{2t}) + \frac{1}{T}\sum_{t=1}^{T}[2z_t^2 - 1](h_{1t}^2 - u_{1t}^2)$$
$$+ \frac{1}{T}\sum_{t=1}^{T}[1-z_t^2]u_{2t} + \frac{1}{T}\sum_{t=1}^{T}[2z_t^2 - 1]u_{1t}^2.$$

The last two terms on the right-hand side converge by the ergodic theorem to $E[2z_t^2 - 1]Eu_{1t}^2 = Eu_{1t}^2$ using the independence of u_{it} and z_t. As $E|1 - z_t^2|$ and $E|2z_t^2 - 1|$ are finite constants, Lemma 4 implies that the first two terms on the right-hand side converge in (L^1 and hence) probability to zero. ∎

Remark 4. Note that the arguments for the score and information in Lemmas 5 and 6 carry over to the stationary case by using the ergodic theorem for the observed information as in Lee and Hansen (1994) and Lumsdaine (1996).

3.4. Third Derivative of the Likelihood Function

In this section the third derivative of the likelihood function is shown to be uniformly bounded in a neighborhood around the true value θ_0.

Introduce lower and upper values for each parameter in θ, $\alpha_L < \alpha_0 < \alpha_U$, $\beta_L < \beta_0 < \beta_U$, $\omega_L < \omega_0 < \omega_U$, and $\gamma_L < \gamma_0 < \gamma_U$, and, in terms of these, the neighborhood $N(\theta_0)$ around the true value θ_0 defined as

$$N(\theta_0) = \{\theta | \alpha_L \leq \alpha \leq \alpha_U, \beta_L \leq \beta \leq \beta_U, \omega_L \leq \omega \leq \omega_U, \text{ and } \gamma_L \leq \gamma \leq \gamma_U\}. \tag{26}$$

The next lemma establishes that the individual terms of the third derivative $(\partial^3 \ell_T / \partial \beta^3)(\theta)$ in (10) are uniformly bounded in the neighborhood $N(\theta_0)$ by the corresponding terms as a function of β alone. With

$$\theta(\beta) = (\alpha_0, \beta, \omega_0, \gamma_0) \tag{27}$$

introduce the notation $h_t(\beta)$ and $h_{it}(\beta)$ for $h_t(\theta(\beta))$ and $h_{it}(\theta(\beta))$, respectively, with $i = 1,2,3$. Then the following lemma holds.

LEMMA 7. *With $N(\theta_0)$ defined in (26), then for any t, s*

$$\sup_{\theta \in N(\theta_0)} \frac{h_t(\theta)}{h_s(\theta)} \leq \kappa_1 \sup_{\beta_L \leq \beta \leq \beta_U} \frac{h_t(\beta)}{h_s(\beta)}, \tag{28}$$

$$\sup_{\theta \in N(\theta_0)} \frac{1}{h_t(\theta)} \leq \kappa_2 \sup_{\beta_L \leq \beta \leq \beta_U} \frac{1}{h_t(\beta)}, \tag{29}$$

and, furthermore,

$$\sup_{\theta \in N(\theta_0)} h_{it}(\theta) \leq \kappa_1 \sup_{\beta_L \leq \beta \leq \beta_U} h_{it}(\beta) \quad \textit{for } i = 1,2,3, \tag{30}$$

where the constants κ_i are given by,

$$\kappa_1 = \frac{\max\left(\frac{\alpha_U}{\alpha_0}, \frac{\gamma_U}{\gamma_0}, \frac{\omega_U}{\omega_0}\right)}{\min\left(\frac{\alpha_L}{\alpha_0}, \frac{\gamma_L}{\gamma_0}, \frac{\omega_L}{\omega_0}\right)} \quad \text{and} \quad \kappa_2 = \min\left(\frac{\alpha_L}{\alpha_0}, \frac{\gamma_L}{\gamma_0}, \frac{\omega_L}{\omega_0}\right).$$

Proof of Lemma 7. Note that with $h_0 = \gamma$ then by simple recursion,

$$h_t(\theta) = \omega \sum_{j=1}^{t} \beta^{j-1} + \alpha \sum_{j=1}^{t} \beta^{j-1} y_{t-j}^2 + \gamma \beta^t.$$

Hence,

$$\min\left(\frac{\alpha_L}{\alpha_0}, \frac{\gamma_L}{\gamma_0}, \frac{\omega_L}{\omega_0}\right) h_t(\beta) \leq h_t(\theta) \leq \max\left(\frac{\alpha_U}{\alpha_0}, \frac{\gamma_U}{\gamma_0}, \frac{\omega_U}{\omega_0}\right) h_t(\beta), \tag{31}$$

which implies (28) and (29). Next, (30) follows by applying (28) to the definition of $h_{it}(\theta)$ in (11)–(13). ∎

To establish bounds for $h_t(\beta)$ and $h_{it}(\beta)$ in Lemma 9 we start with two fundamental identities concerning h_t and $h_t(\beta)$.

LEMMA 8.

$$h_t(\beta) = h_t + (\beta - \beta_0) \sum_{s=1}^{t} \beta^{s-1} h_{t-s}, \tag{32}$$

$$h_t = h_t(\beta) + (\beta_0 - \beta) \sum_{s=1}^{t} \beta_0^{s-1} h_{t-s}(\beta). \tag{33}$$

Proof of Lemma 8. Rewriting the equations for $h_t(\beta)$ and h_t gives

$$h_t(\beta) - \beta h_{t-1}(\beta) = h_t - \beta_0 h_{t-1},$$

and the results follow immediately by noting that $h_0 = h_0(\beta)(= \gamma_0)$. ∎

Next turn to Lemma 9, which holds independently of Assumption 1.

LEMMA 9. *With* $\beta_L < \beta_0 < \beta_U$,

(i) $$\frac{h_t}{h_t(\beta)} \leq \begin{cases} 1 & \text{for } \beta_0 \leq \beta \leq \beta_U \\ 1 + (\beta_0 - \beta_L) u_{1t}(\beta_0, \beta_L) & \text{for } \beta_L \leq \beta \leq \beta_0 \end{cases}$$

(ii) $$h_{1t}(\beta) \leq \begin{cases} u_{1t}(\beta_U, \beta_0) + \frac{1}{2}(\beta_U - \beta_0) u_{2t}(\beta_U, \beta_0) & \text{for } \beta_0 \leq \beta \leq \beta_U \\ u_{1t}(\beta_0, \beta_L) & \text{for } \beta_L \leq \beta \leq \beta_0 \end{cases}$$

(iii) $$h_{2t}(\beta) \leq \begin{cases} u_{2t}(\beta_U, \beta_0) + \frac{1}{3}(\beta_U - \beta_0) u_{3t}(\beta_U, \beta_0) & \text{for } \beta_0 \leq \beta \leq \beta_U \\ u_{2t}(\beta_0, \beta_L) & \text{for } \beta_L \leq \beta \leq \beta_0 \end{cases}$$

(iv) $$h_{3t}(\beta) \leq \begin{cases} u_{3t}(\beta_U, \beta_0) + \frac{1}{4}(\beta_U - \beta_0) u_{4t}(\beta_U, \beta_0) & \text{for } \beta_0 \leq \beta \leq \beta_U \\ u_{3t}(\beta_0, \beta_L) & \text{for } \beta_L \leq \beta \leq \beta_0 \end{cases}$$

where the $u_{it}(\cdot)$ *are defined in (15).*

Proof of Lemma 9. Consider first $h_t/h_t(\theta)$. For $\beta_0 \leq \beta < \beta_U$, using (32)

$$\frac{h_t}{h_t(\beta)} = \frac{h_t}{h_t + (\beta - \beta_0)\sum_{s=1}^{t}\beta^{s-1}h_{t-s}} \leq 1.$$

For the case of $\beta_L < \beta \leq \beta_0$, use (33) to see that

$$\frac{h_t}{h_t(\beta)} = \frac{h_t(\beta) + (\beta_0 - \beta)\sum_{s=1}^{t}\beta_0^{s-1}h_{t-s}(\beta)}{h_t(\beta)} = 1 + (\beta_0 - \beta)\sum_{s=1}^{t}\beta_0^{s-1}\frac{h_{t-s}(\beta)}{h_t(\beta)}.$$

Then, similar to the proof of Lemma 4,

$$\begin{aligned}\frac{h_{t-s}(\beta)}{h_t(\beta)} &= \prod_{k=1}^{s}\frac{h_{t-k}(\beta)}{h_{t-k+1}(\beta)} = \prod_{k=1}^{s}\frac{h_{t-k}(\beta)}{\omega_0 + \alpha_0 y_{t-k}^2 + \beta h_{t-k}(\beta)}\\ &\leq \prod_{k=1}^{s}\frac{1}{\alpha_0 z_{t-k}^2\dfrac{h_{t-k}}{h_{t-k}(\beta)} + \beta} \leq \prod_{k=1}^{s}\frac{1}{\alpha_0 z_{t-k}^2 + \beta_L}\end{aligned} \quad \textbf{(34)}$$

because $h_{t-k}/h_{t-k}(\beta) \geq 1$ for $\beta \leq \beta_0$. Inserting, it follows that

$$\frac{h_t}{h_t(\beta)} \leq 1 + (\beta_0 - \beta_L)\sum_{s=1}^{\infty}\frac{1}{\beta_0}\prod_{k=1}^{s}\frac{\beta_0}{\alpha_0 z_{t-k}^2 + \beta_L} = 1 + (\beta_0 - \beta_L)u_{1t}(\beta_0, \beta_L),$$

where the right-hand side is independent of β. This establishes the inequality in (i). The inequalities (ii)–(iv) hold by identical arguments, and we give the proof only for (iv), which is the most complicated of these. By definition,

$$h_{3t}(\beta) = 3\sum_{j=3}^{t}\beta^{j-3}(j-1)(j-2)\frac{h_{t-j}(\beta)}{h_t(\beta)},$$

and the inequality for $\beta_L < \beta \leq \beta_0$ holds by (34). Next, for $\beta_0 \leq \beta < \beta_U$, $h_t = h_t(\theta_0) \leq h_t(\beta)$, and using this in addition to (32) gives

$$
\begin{aligned}
h_{3t}(\beta) &= 3\sum_{j=3}^{t}(j-1)(j-2)\beta^{j-3}\,\frac{h_{t-j}(\beta)}{h_t}\\
&= 3\sum_{j=3}^{t}(j-1)(j-2)\beta^{j-3}\,\frac{h_{t-j}}{h_t}\\
&\quad + 3(\beta-\beta_0)\sum_{j=3}^{t}(j-1)(j-2)\beta^{j-3}\sum_{s=1}^{t-j}\beta^{s-1}\,\frac{h_{t-j-s}}{h_t}\\
&= 3\sum_{j=3}^{t}(j-1)(j-2)\beta^{j-3}\,\frac{h_{t-j}}{h_t}\\
&\quad + (\beta-\beta_0)\sum_{j=4}^{t}(j-1)(j-2)(j-3)\beta^{j-4}\,\frac{h_{t-j}}{h_t}\\
&\leq 3\sum_{j=3}^{t}(j-1)(j-2)\beta_U^{j-3}\prod_{k=1}^{j}\frac{1}{\alpha_0 z_{t-k}^2+\beta_0}\\
&\quad + (\beta_U-\beta_0)\sum_{j=4}^{t}(j-1)(j-2)(j-3)\beta_U^{j-4}\prod_{k=1}^{j}\frac{1}{\alpha_0 z_{t-k}^2+\beta_0},
\end{aligned}
$$

where the last inequality follows as in (34). This shows (iv) and completes the proof of Lemma 9. ∎

We are now in a position to address the third derivative of the likelihood function, $(\partial^3\ell_T/\partial\beta^3)(\theta)$, as given by (10). We show that, independently of Assumption 1, it is uniformly bounded in a region around the true value β_0.

LEMMA 10. *There exists a neighborhood $N(\theta_0)$ given by (26) for which*

$$
\sup_{\theta\in N(\theta_0)}\left|\frac{\partial^3\ell_T}{\partial\beta^3}(\theta)\right| \leq \frac{1}{T}\sum_{t=1}^{T} w_t, \tag{35}
$$

where w_t is stationary and has finite moment, $Ew_t = M < \infty$. Furthermore,

$$
\frac{1}{T}\sum_{t=1}^{T} w_t \xrightarrow{a.s.} M. \tag{36}
$$

Proof of Lemma 10. Noting that by definition $y_t^2/h_t(\theta) = z_t^2(h_t/h_t(\theta))$, the expression for $(\partial^3\ell_T/\partial\beta^3)(\theta)$ in (10) implies that

$$
\left|\frac{\partial^3\ell_T}{\partial\beta^3}(\theta)\right| \leq \frac{1}{T}\sum_{t=1}^{T} w_t(\theta), \quad \text{where}
$$

$$
\begin{aligned}
w_t(\theta) = \left(1+z_t^2\frac{h_t}{h_t(\theta)}\right)h_{3t}(\theta) + 3\left(2z_t^2\frac{h_t}{h_t(\theta)}+1\right)h_{1t}(\theta)h_{2t}(\theta)\\
+ 2\left(1+3z_t^2\frac{h_t}{h_t(\theta)}\right)h_{1t}^3(\theta).
\end{aligned}
$$

Lemma 7 implies that

$$\sup_{\theta \in N(\theta_0)} w_t(\theta) \leq c \sup_{\beta_L \leq \beta \leq \beta_U} w_t(\beta)$$

with c a constant. By Lemma 9, the quantities $h_{it}(\beta)$, $i = 1,\ldots,3$, and $h_t/h_t(\beta)$ are bounded by functions that by Lemma 3 have any desired moments. Hence, $\sup_{\beta_L \leq \beta \leq \beta_U} w_t(\beta) \leq w_t$ as desired. The convergence in (36) follows by the ergodic theorem, which ends the proof of Lemma 10. ■

3.5. Introducing α

The arguments with respect to α, and hence the joint variation in terms of both α and β, are completely analogous to the ones in Sections 3.1–3.4, and we emphasize only the important steps. Simple computations lead to

$$\frac{\partial \ell_T(\theta)}{\partial \alpha} = \frac{1}{T} \sum_{t=1}^{T} \left(1 - \frac{y_t^2}{h_t(\theta)}\right) h_{1t}^*(\theta), \quad \text{where}$$

$$h_{1t}^*(\theta) = \frac{\partial h_t(\theta)/\partial \alpha}{h_t(\theta)} = \frac{\sum_{j=1}^{t} \beta^{j-1} y_{t-j}^2}{h_t(\theta)}.$$

Hence

$$h_{1t}^* = h_{1t}^*(\theta_0) = \sum_{j=1}^{t} z_{t-j}^2 \beta_0^{j-1} \frac{h_{t-j}}{h_t},$$

which as in Lemma 3 leads to the definition of the ergodic process,

$$u_{1t}^* := \sum_{j=1}^{\infty} z_{t-j}^2 \frac{1}{\beta_0} \prod_{k=1}^{j} \frac{\beta_0}{\alpha_0 z_{t-k}^2 + \beta_0}. \tag{37}$$

As in the proofs of Lemmas 4–6 it follows that

$$\sqrt{T} \left. \frac{\partial \ell_T(\theta)}{\partial (\alpha, \beta)'} \right|_{\theta = \theta_0} \xrightarrow{D} N_2(0, \kappa \Sigma) \tag{38}$$

and

$$\frac{1}{T} \left. \frac{\partial^2 \ell_T(\theta)}{\partial (\alpha, \beta) \partial (\alpha, \beta)'} \right|_{\theta = \theta_0} \xrightarrow{P} \Sigma > 0, \tag{39}$$

where $\kappa = V(z_t^2)$ and

$$\Sigma = \begin{pmatrix} E(u_{1t}^*)^2 & E(u_{1t}^* u_{1t}) \\ E(u_{1t}^* u_{1t}) & E(u_{1t})^2 \end{pmatrix}.$$

We note the surprisingly simple relationship,

$$
\begin{aligned}
u_{1t}^{*} &= \sum_{j=1}^{\infty} \frac{1}{\alpha_0} \frac{\alpha_0 z_{t-j}^2}{\alpha_0 z_{t-j}^2 + \beta_0} \prod_{k=1}^{j-1} \frac{\beta_0}{\alpha_0 z_{t-k}^2 + \beta_0} \\
&= \sum_{j=1}^{\infty} \frac{1}{\alpha_0} \left(1 - \frac{\beta_0}{\alpha_0 z_{t-j}^2 + \beta_0}\right) \prod_{k=1}^{j} \frac{\beta_0}{\alpha_0 z_{t-k}^2 + \beta_0} \\
&= \sum_{j=1}^{\infty} \frac{1}{\alpha_0} \left(\prod_{k=1}^{j-1} \frac{\beta_0}{\alpha_0 z_{t-k}^2 + \beta_0} - \prod_{k=1}^{j} \frac{\beta_0}{\alpha_0 z_{t-k}^2 + \beta_0}\right) \\
&= \lim_{T\to\infty} \sum_{j=1}^{T} \frac{1}{\alpha_0} \left(\prod_{k=1}^{j-1} \frac{\beta_0}{\alpha_0 z_{t-k}^2 + \beta_0} - \prod_{k=1}^{j} \frac{\beta_0}{\alpha_0 z_{t-k}^2 + \beta_0}\right) \\
&= \lim_{T\to\infty} \frac{1}{\alpha_0} \left(1 - \prod_{k=1}^{T} \frac{\beta_0}{\alpha_0 z_{t-k}^2 + \beta_0}\right) = \frac{1}{\alpha_0},
\end{aligned}
$$

which implies

$$
\Sigma = \begin{pmatrix} \dfrac{1}{\alpha_0^2} & \dfrac{\mu_1}{\alpha_0 \beta_0 (1-\mu_1)} \\ \dfrac{\mu_1}{\alpha_0 \beta_0 (1-\mu_1)} & \dfrac{(1+\mu_1)\mu_2}{\beta_0^2 (1-\mu_1)(1-\mu_2)} \end{pmatrix}, \tag{40}
$$

where $\mu_i = E(\beta_0/(\alpha_0 z_t^2 + \beta_0))^i$, $i = 1,2$.

Finally, straightforward differentiation shows that inequalities completely analogous to (35) in Lemma 10 hold for the third derivatives $\partial^3 \ell_T(\theta)/\partial\alpha^3$, $\partial^3 \ell_T(\theta)/\partial\alpha^2\partial\beta$, and $\partial^3 \ell_T(\theta)/\partial\alpha\partial\beta^2$.

Remark 5. Note that the nonstationary condition in Assumption 1 is not needed to establish these bounds (see also Lemma 10), and hence the uniform bounds can be applied in the stationary case also. Furthermore, the bounds of the third derivatives establish inequalities of the form, $E\sup_{\theta\in N(\theta_0)} |\partial^3 \ell_T(\theta)/\partial\alpha^i \partial\beta^j| < \infty$ for $i + j = 3$, $i, j = 0,1,2,3$, where $N(\theta_0)$ denotes a neighborhood of the true parameter values, θ_0 (see also the condition (A.3) in Lemma 1). Observe that the proofs in Lee and Hansen (1994, p. 51, l.13–14 from below, regarding stochastic equicontinuity) and Lumsdaine (1996, p. 594) apply, we believe, different insufficient inequalities of the form, $\sup_{\theta\in\theta\in N(\theta_0)} E|\partial^3 \ell_T(\theta)/\partial\alpha^i \partial\beta^j| < \infty$.

4. PROOF OF THEOREM 2

We address here the asymptotic independence of the initial values of (ω, γ).

Recall that in the proof of Theorem 1, Lemma 1 was applied with $\varphi := (\alpha, \beta)'$ and $\ell_T(\varphi) = \ell_T(\alpha, \beta)$ defined in Theorem 1. If there is an extra parameter ω, say, in the likelihood function as in Theorem 2, it follows that the results are unchanged in Theorem 1 provided the conditions in Lemma 13 in Section 4.2 are fulfilled. To see this note first that in the proof of Lemma 1 in the Appendix, $D\ell_T(\varphi_0)$ can be replaced by $D\ell_T(\varphi_0, \omega)$ on the right-hand side of equation (A.2) using (44) in Lemma 13. Likewise in equation (A.2), $D^2\ell_T(\varphi^*)$ can be replaced by $D^2\ell_T(\varphi^*, \omega)$ using (45) in Lemma 13. In equation (A.3), on the left-hand side, $\sqrt{T}D\ell_T(\varphi_0)$ can be replaced by $\sqrt{T}D\ell_T(\varphi_0, \omega)$ using (44) in Lemma 13, and on the right-hand side, $v_1' A_T(\hat{\varphi}_T) v_2 = v_1'(D^2\ell_T(\varphi_T^*) - \Omega_I) v_2$ can be replaced by $v_1'(D^2\ell_T(\varphi_T^*, \omega) - \Omega_I) v_2$ using (45) in Lemma 13.

The derivations that follow are given in detail for ω, whereas for the case of γ we provide one lemma and note that the proof follows as in the ω case but is simpler. For presentational purposes we focus on the variation with respect to β in addition to ω and γ, respectively, which as in the proof of Theorem 1 in Section 3 is without loss of generality. To emphasize which parameters vary the following notation is adopted:

$$\theta(\beta, \omega) = (\alpha_0, \beta, \omega, \gamma_0).$$

Similarly $\theta(\beta, \gamma) = (\alpha_0, \beta, \omega_0, \gamma)$; see also (27). We further emphasize the dependence on, e.g., ω by adopting the notation $\ell_T(\beta, \omega) := \ell_T(\theta(\beta, \omega))$ for the likelihood function and also for other functions previously defined.

Section 4.1 states an initial general lemma that is useful in the derivations. Section 4.2 addresses ω and Section 4.3 γ.

4.1. An Initial Lemma

The following lemma addresses the average of products of stochastic processes.

LEMMA 11. *Let* $(X_t)_{t=1,2,\ldots}$ *and* $(Y_t)_{t=1,2,\ldots}$ *be stochastic processes on* $\mathbb{R}$ *for which*

$$E|X_t| < c_x, \quad E|Y_t|^a < c_y t^d \rho^t,$$

where c_x, c_y, a, *and* d *are positive constants and* $|\rho| < 1$. *Then with* $\delta > 0$ *and as* $T \to \infty$,

$$\frac{1}{T^\delta} \sum_{t=1}^{T} X_t Y_t \xrightarrow{P} 0.$$

Proof of Lemma 11. We establish L^s convergence for $s = a/(1+a) < 1$. First, as $s < 1$,

$$E\left|\frac{1}{T^\delta}\sum_{t=1}^{T} X_t Y_t\right|^s \leq \frac{1}{T^{s\delta}}\sum_{t=1}^{T} E|X_t Y_t|^s.$$

Next, with $p = 1 + a > 1$ and $q = p/(p-1) > 1$ apply Hölder's inequality to get that this is bounded by

$$\begin{aligned}
&\frac{1}{T^{s\delta}}\sum_{t=1}^{T}(E|X_t|^{sq})^{1/q}(E|Y_t|^{sp})^{1/p}\\
&\quad = \frac{1}{T^{s\delta}}\sum_{t=1}^{T}(E|X_t|)^{1/q}(E|Y_t|^{a})^{1/p}\\
&\quad \leq \frac{1}{T^{s\delta}}\sum_{t=1}^{T} c_x^{1/q} c_y^{1/p} t^{d/p}\rho^{t/p} \leq \frac{1}{T^{s\delta}} c_x^{1/q} c_y^{1/p} c,
\end{aligned}$$

which tends to zero as $T \to \infty$ and where $c = \sum_{t=1}^{\infty} t^{d/p}\rho^{t/p} < \infty$. ∎

4.2. The Role of the Scale Parameter ω

Initially we provide upper bounds for the terms appearing in the score and observed information in (8) and (9). To this end we need the following proposition.

PROPOSITION 1. *Assume that Assumption 1 holds with strict inequality and with $\beta_0 < 1$. Then, for some $p > 0$,*

$$E(\alpha_0 z_t^2 + \beta_0)^{-p} < 1.$$

Proof of Proposition 1. Set $v_t = \alpha_0 z_t^2 + \beta_0$ and note that $v_t \geq \beta_0$. Define the function $f_p(v) = (v^{-p} - 1)/p = (\exp(-p\log v) - 1)/p \to -\log v$ as $p \to 0$. Note that on $A_1 = [\beta_0, 1](= \emptyset \text{ if } \beta_0 > 1)$, $0 \leq f_p(v) \leq (1/\beta_0 - 1)$ for $0 \leq p \leq 1$, whereas on $A_2 = (1,\infty)$, $-f_p(v) \geq 0$ and increasing in p as $p \to 0$. Finally, $Ef_p(v_t) = E[f_p(v_t)1_{A_1}(v_t)] - E[-f_p(v_t)1_{A_2}(v_t)]$, which by dominated and monotone convergence respectively, converge to $E[-\log v_t 1_{A_1}(v_t)] - E[\log v_t 1_{A_2}(v_t)] = -E\log v_t$, which is negative by Assumption 1. Hence for p small enough the result holds. ∎

Next consider individual terms appearing in the likelihood function in (6) and also terms of the score and observed information in (8) and (9) and their variation with respect to ω.

LEMMA 12. *Assume that Assumption 1 holds with strict inequality. Then for any $\omega > 0$, there exist $\beta_L < \beta_0 < \beta_U$ such that*

$$\sup_{\beta_L \le \beta \le \beta_U} \frac{\partial h_t(\beta,\omega)/\partial\omega}{h_t} = \sup_{\beta_L \le \beta \le \beta_U} \frac{\sum_{i=0}^{t-1} \beta^i}{h_t} \le r_{1\omega t},$$

$$\sup_{\beta_L \le \beta \le \beta_U} \frac{\partial^2 h_t(\beta,\omega)/\partial\beta\partial\omega}{h_t} = \sup_{\beta_L \le \beta \le \beta_U} \frac{\sum_{i=0}^{t-2} (i+1)\beta^i}{h_t} \le r_{2\omega t} t,$$

$$\sup_{\beta_L \le \beta \le \beta_U} \frac{\partial^3 h_t(\beta,\omega)/\partial\beta^2\partial\omega}{h_t} = \sup_{\beta_L \le \beta \le \beta_U} \frac{\sum_{i=0}^{t-3} (i+1)(i+2)\beta^i}{h_t} \le r_{3\omega t} t^2.$$

Here, with i = 1, 2, and 3,

$$r_{i\omega t} = \frac{1}{\gamma_0}\begin{cases} \dfrac{1}{(\beta_U - 1)^i} \displaystyle\prod_{k=1}^{t} \frac{\beta_U}{\alpha_0 z_{t-k}^2 + \beta_0} & \text{for } \beta_0 \ge 1 \\ \dfrac{1}{(1-\beta_U)^i} \displaystyle\prod_{k=1}^{t} \frac{1}{\alpha_0 z_{t-k}^2 + \beta_0} & \text{for } \beta_0 < 1, \end{cases} \tag{41}$$

$[Er_{i\omega t}^p]^{1/p} = \rho_i^t,$

where $\rho_i < 1$ and $p > 1$ for $\beta_0 \ge 1$, whereas $p > 0$ for $\beta_0 < 1$.

Proof of Lemma 12. We give only the proof for $i = 1$ as the other cases follow analogously. Note that $h_t \ge \prod_{j=0}^{t}(\alpha_0 z_{t-1}^2 + \beta_0)\gamma_0$ and hence

$$\frac{\sum_{i=0}^{t-1} \beta^i}{h_t} \le \frac{\sum_{i=0}^{t-1} \beta_U^i}{\gamma_0} \prod_{j=0}^{t} \frac{1}{\alpha_0 z_{t-j}^2 + \beta_0}. \tag{42}$$

Consider first the case of $\beta_0 \ge 1$, which implies $\beta_U > 1$ and in particular,

$$\frac{\sum_{i=0}^{t-1} \beta^i}{h_t} \le \frac{\beta_U^t}{\gamma_0(\beta_U - 1)} \prod_{j=0}^{t} \frac{1}{\alpha_0 z_{t-j}^2 + \beta_0} = \frac{1}{\gamma_0(\beta_U - 1)} \prod_{j=0}^{t} \frac{\beta_U}{\alpha_0 z_{t-j}^2 + \beta_0}.$$

This function has exponentially decreasing absolute pth-order moment, $p \ge 1$, provided

$$E\left(\frac{\beta_U}{\alpha_0 z_t^2 + \beta_0}\right)^p < 1, \tag{43}$$

which is the case for some $\beta_U > \beta_0$; see Lemma 3.

Next turn to the case of $\beta_0 < 1$. In this case, without loss of generality, it can be assumed that $\beta_U < 1$, and we find

$$\frac{\sum_{i=0}^{t-1}\beta^i}{h_t} \leq \frac{\sum_{i=0}^{t-1}\beta_U^i}{\gamma_0}\prod_{j=0}^{t}\frac{1}{\alpha_0 z_{t-1}^2+\beta_0} \leq \frac{1}{\gamma_0(1-\beta_U)}\prod_{j=0}^{t}\frac{1}{\alpha_0 z_{t-1}^2+\beta_0}.$$

Applying Proposition 1 finishes the proof of Lemma 12. ■

Next, turn to the main lemma.

LEMMA 13. *Assume that Assumption 1 holds with strict inequality. Then for any arbitrary* $\omega > 0$, *there exist* β_L *and* β_U, $\beta_L < \beta_0 < \beta_U$, *such that*

$$\sqrt{T}\left(\frac{\partial \ell_T(\beta_0,\omega)}{\partial\beta} - \frac{\partial \ell_T}{\partial\beta}\right) \xrightarrow{P} 0, \tag{44}$$

$$\sup_{\beta_L\leq\beta\leq\beta_U}\left(\frac{\partial^2 \ell_T(\beta,\omega)}{\partial\beta^2} - \frac{\partial^2 \ell_T(\beta,\omega_0)}{\partial\beta^2}\right) \xrightarrow{P} 0. \tag{45}$$

Proof of Lemma 13. Given the arbitrary value $\omega_{\text{arb}}(=\omega) > 0$, define $\omega_L = \min(\omega_0,\omega_{\text{arb}})$ and $\omega_U = \max(\omega_0,\omega_{\text{arb}})$. By Taylor expansions, (44) and (45) follow by showing that

$$\sup_{\omega_L\leq\omega\leq\omega_U}\sqrt{T}\,\frac{\partial^2 \ell_T(\beta_0,\omega)}{\partial\beta\partial\omega} \xrightarrow{P} 0, \tag{46}$$

$$\sup_{\beta_L\leq\beta\leq\beta_U}\ \sup_{\omega_L\leq\omega\leq\omega_U}\left|\frac{\partial^3 \ell_T(\beta,\omega)}{\partial\beta^2\partial\omega}\right| \xrightarrow{P} 0. \tag{47}$$

Simple computations give

$$\begin{aligned}
\sqrt{T}\,\frac{\partial^2 \ell_T(\beta,\omega)}{\partial\beta\partial\omega}\bigg|_{\beta=\beta_0} &= \frac{\partial}{\partial\omega}\left(\frac{1}{\sqrt{T}}\sum_{t=1}^{T}\left[1-\frac{y_t^2}{h_t(\beta_0,\omega)}\right]h_{1t}(\beta_0,\omega)\right)\\
&= \frac{1}{\sqrt{T}}\sum_{t=1}^{T}\left[2\,\frac{y_t^2}{h_t(\beta_0,\omega)}-1\right]\left(\frac{\partial h_t(\beta_0,\omega)/\partial\omega}{h_t(\beta_0,\omega)}\right)h_{1t}(\beta_0,\omega)\\
&\quad+\frac{1}{\sqrt{T}}\sum_{t=1}^{T}\left[1-\frac{y_t^2}{h_t(\beta_0,\omega)}\right]\frac{\partial^2 h_t(\beta_0,\omega)/\partial\beta\partial\omega}{h_t(\beta_0,\omega)}.
\end{aligned} \tag{48}$$

Likewise,

$$
\begin{aligned}
\frac{\partial^3 \ell_T(\beta,\omega)}{\partial\beta^2\partial\omega} &= \frac{1}{T}\sum_{t=1}^{T}\left[1-\frac{y_t^2}{h_t(\beta,\omega)}\right]\frac{\partial^3 h_t(\beta,\omega)/\partial\beta^3\partial\omega}{h_t(\beta,\omega)} \\
&\quad + \frac{1}{T}\sum_{t=1}^{T}\left[2\,\frac{y_t^2}{h_t(\beta,\omega)}-1\right]h_{2t}(\beta,\omega)\left(\frac{\partial h_t(\beta,\omega)/\partial\omega}{h_t(\beta,\omega)}\right) \\
&\quad + \frac{1}{T}\sum_{t=1}^{T}\left[4\,\frac{y_t^2}{h_t(\beta,\omega)}-2\right]h_{1t}(\beta,\omega)\left(\frac{\partial^2 h_t(\beta,\omega)/\partial\beta\partial\omega}{h_t(\beta,\omega)}\right) \\
&\quad + \frac{1}{T}\sum_{t=1}^{T}\left[2-6\,\frac{y_t^2}{h_t(\beta,\omega)}\right](h_{1t}(\beta,\omega))^2\left(\frac{\partial h_t(\beta,\omega)/\partial\omega}{h_t(\beta,\omega)}\right). \qquad \textbf{(49)}
\end{aligned}
$$

All the preceding terms are bounded as they are all of a form that can be expressed in the form described in Lemma 11: a typical term in (48) and (49) is given by

$$
\begin{aligned}
&\left|\frac{1}{T}\sum_{t=1}^{T}\left[2\,\frac{y_t^2}{h_t(\beta,\omega)}-1\right]h_{2t}(\beta,\omega)\left(\frac{\partial h_t(\beta,\omega)/\partial\omega}{h_t(\beta,\omega)}\right)\right| \\
&\qquad \le \frac{1}{T}\sum_{t=1}^{T}\left|2\,\frac{y_t^2}{h_t}\left(\frac{h_t}{h_t(\beta,\omega)}\right)+1\right|\left|\frac{h_t}{h_t(\beta,\omega)}\right||h_{2t}(\beta,\omega)|\left|\frac{\partial h_t(\beta,\omega)/\partial\omega}{h_t}\right|. \qquad \textbf{(50)}
\end{aligned}
$$

By Lemma 12, $\sup_{\beta_L\le\beta\le\beta_U}\sup_{\omega_L\le\omega\le\omega_U}|(\partial h_t(\beta,\omega)/\partial\omega)/h_t|$ has exponentially decreasing moments, and this factor plays the role of Y_t in Lemma 11. Using Lemmas 7 and 9, the remaining three factors are bounded by variables, which by Lemma 3 have finite moments of any desired order. Hence the product of these variables plays the role of X_t in Lemma 11. This ends the proof of Theorem 2 regarding the role of ω. ■

4.3. The Role of the Initial Value $h_0(\theta) = \gamma$

As mentioned, the proof, although simpler, follows exactly the outline of the proof of the independence on the scale parameter ω given in Section 4.2. Recall that to emphasize the dependence on γ, we adopt the notation $\ell_T(\beta,\gamma) = \ell_T(\theta(\beta,\gamma))$ for the likelihood function and other functions. To establish the results in Lemma 13 with ω replaced by γ we need only the following lemma, which corresponds to Lemma 12.

LEMMA 14. *Under Assumption 1, for any* $\gamma > 0$ *there exist* β_L *and* β_U, $\beta_L < \beta_0 < \beta_U$ *such that*

$$\sup_{\beta_L \le \beta \le \beta_U} \frac{\partial h_t(\beta,\gamma)/\partial\gamma}{h_t} = \sup_{\beta_L \le \beta \le \beta_U} \frac{\beta^t}{h_t} \le \frac{1}{\gamma_0} r_{h\gamma t},$$

$$\sup_{\beta_L \le \beta \le \beta_U} \frac{\partial^2 h_t(\beta,\gamma)/\partial\beta\partial\gamma}{h_t} = \sup_{\beta_L \le \beta \le \beta_U} \frac{t\beta^{t-1}}{h_t} \le \frac{t}{\gamma_0 \beta_0} r_{h\gamma t},$$

$$\sup_{\beta_L \le \beta \le \beta_U} \frac{\partial^3 h_t(\beta,\gamma)/\partial\beta^2\partial\gamma}{h_t} = \sup_{\beta_L \le \beta \le \beta_U} \frac{t(t-1)\beta^{t-2}}{h_t} \le \frac{t(t-1)}{\gamma_0 \beta_0} r_{h\gamma t}$$

with

$$r_{h\gamma t} = \prod_{k=1}^{t} \frac{\beta_U}{\alpha_0 z_{t-k}^2 + \beta_0}, \tag{51}$$

$$[E r_{h\gamma t}^2]^{1/2} = \rho^t \quad \text{where } \rho < 1. \tag{52}$$

Proof of Lemma 14. The results follow as in the proof of Lemma 12 for the case of $\beta_0 > 1$. ■

REFERENCES

Basawa, I.V., P.D. Feigin, & C.C. Heyde (1976) Asymptotic properties of maximum likelihood estimators for stochastic processes. *Sankya Series A* 38, 259–270.

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.

Brown, B.M. (1971) Martingale central limit theorems. *Annals of Mathematical Statistics* 42, 59–66.

Engle, R.F. (1982) Autoregressive conditional heteroscedasticity with estimates of United Kingdom inflation. *Econometrica* 50, 987–1007.

Jensen, S.T. & A. Rahbek (2004) Non-stationary and no-moments asymptotics for the ARCH(1) model. *Econometrica* 72, 641–646.

Kristensen, D. & A. Rahbek (2002) Asymptotics of the QMLE for a Class of ARCH(q) models. Preprint 10, Department of Applied Mathematics and Statistics, University of Copenhagen.

Lee, S.-W. & B. Hansen (1994) Asymptotic theory for the GARCH(1,1) quasi-maximum likelihood estimator. *Econometric Theory* 10, 29–53.

Lehmann, E.L. (1999) *Elements of Large-Sample Theory*. Springer-Verlag.

Lumsdaine, R. (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.

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

Weiss, A. (1986) Asymptotic theory for ARCH models. *Econometric Theory* 2, 107–131.

APPENDIX: PROOF OF LEMMA 1

Note first that by condition (A.3) in Lemma 1, it follows that for any vectors $v_1, v_2 \in \mathbb{R}^k$, and any $\varphi \in N(\varphi_0)$,

$$|v_1'(D^2\ell_T(\varphi) - D^2\ell_T(\varphi_0))v_2| \leq \|v_1\|\|v_2\|\|\varphi - \varphi_0\|\tilde{c}_T, \tag{A.1}$$

where $D\ell_T(\varphi) = \partial\ell_T(\varphi)/\partial\varphi$, $D^2\ell_T(\varphi) = \partial^2\ell_T(\varphi)/\partial\varphi\partial\varphi'$, and $\tilde{c}_T = k^{3/2}c_T$. To see this, note that the left-hand side of expression (A.1) is $|f(1) - f(0)| = |\partial f(\lambda^*)/\partial\lambda|$ for some $0 \leq \lambda^* \leq 1$, where $f(\lambda) = v_1'[D^2\ell_T(\varphi_0 - \lambda(\varphi - \varphi_0))]v_2$, $0 \leq \lambda \leq 1$. By Taylor's formula and condition (A.3) in Lemma 1,

$$\begin{aligned}|\partial f(\lambda^*)/\partial\lambda| &= \left|\sum_{i,j,l=1}^{k} v_{1,i}v_{2,j}(\varphi_l - \varphi_{0,l})\partial^3\ell_T(\varphi_0 - \lambda^*(\varphi - \varphi_0))/\partial\varphi_i\partial\varphi_j\partial\varphi_l\right| \\ &\leq c_T\sum_{i=1}^{k}|v_{1,i}|\sum_{j=1}^{k}|v_{2,j}|\sum_{l=1}^{k}|\varphi_l - \varphi_{0,l}| \leq \tilde{c}_T\|v_1\|\|v_2\|\|\varphi - \varphi_0\|.\end{aligned}$$

Next, by definition the continuous function $\ell_T(\varphi)$ attains its minimum in any compact neighborhood $K(\varphi_0, r) = \{\varphi|\ \|\varphi - \varphi_0\| \leq r\} \subseteq N(\varphi_0)$ of φ_0. With $v_\varphi = (\varphi - \varphi_0)$, and φ^* on the line from φ to φ_0, Taylor's formula gives

$$\begin{aligned}\ell_T(\varphi) - \ell_T(\varphi_0) &= D\ell_T(\varphi_0)v_\varphi + \frac{1}{2}v_\varphi'D^2\ell_T(\varphi^*)v_\varphi \\ &= D\ell_T(\varphi_0)v_\varphi + \frac{1}{2}v_\varphi'[\Omega_I + (D^2\ell_T(\varphi_0) - \Omega_I) \\ &\qquad + (D^2\ell_T(\varphi^*) - D^2\ell_T(\varphi_0))]v_\varphi.\end{aligned} \tag{A.2}$$

Denote by ρ_T and ρ, $\rho > 0$, the smallest eigenvalues of $[D^2\ell_T(\varphi_0) - \Omega_I]$ and Ω_I, respectively. Note that $\rho_T \xrightarrow{P} 0$ by condition (A.2) in Lemma 1 and the fact that the smallest eigenvalue of a $k \times k$ symmetric matrix M, $\inf_{\{v\in\mathbb{R}^k|\|v\|=1\}} v'Mv$, is continuous in M. Then conditions (A.1) and (A.3) in Lemma 1, with $\tilde{c} = k^{3/2}c$, and equation (A.2) imply that $\inf_{\varphi:v_\varphi=r}[\ell_T(\varphi) - \ell_T(\varphi_0)]$ is greater than or equal to

$$-\|D\ell_T(\varphi_0)\|r + \frac{1}{2}[\rho + \rho_T - \tilde{c}_T r]r^2 \xrightarrow{P} \frac{1}{2}[\rho - \tilde{c}r]r^2.$$

Therefore, if $r < \rho/\tilde{c}$, the probability that $\ell_T(\varphi)$ attains its minimum on the boundary of $K(\varphi_0, r)$ tends to zero. Next, for $\varphi \in K(\varphi_0, r)$ and $v \in \mathbb{R}^k$, rewriting $v'D^2\ell_T(\varphi)v$ as in equation (A.2), $v'D^2\ell_T(\varphi)v \geq \|v\|^2(\rho + \rho_T - r\tilde{c}_T)$, which tends in probability to $\|v\|^2(\rho - r\tilde{c})$. Hence, if $r < \rho/\tilde{c}$ the probability that $\ell_T(\varphi)$ is strongly convex in the interior of $K(\varphi_0, r)$ tends to one, and therefore it has at most one stationary point. This establishes condition (B.1) in Lemma 1: if $r < \rho/\tilde{c}$ and $K(\varphi_0, r) \subseteq N(\varphi_0)$, there is with a probability tending to one exactly one solution $\hat{\varphi}_T$ to the likelihood equation in the interior $U(\varphi_0) = \operatorname{int} K(\varphi_0, r)$. It is the unique minimum point of $\ell_T(\varphi)$ in $U(\varphi_0)$ and, as it is a stationary point, it solves $D\ell_T(\varphi) = 0$.

By the same argument, for any δ, $0 < \delta < r$ there is with a probability tending to one a solution to the likelihood equation in $K(\varphi_0, \delta)$. As $\hat{\varphi}_T$ is the unique solution to the likelihood equation in $K(\varphi_0, r)$, it must therefore be in $K(\varphi_0, \delta)$ with a probability tending to one. Hence we have proved that $\hat{\varphi}_T$ is consistent. That is, for any $0 < \delta < r$, the probability that $\hat{\varphi}_T$ is a unique solution to $D\ell_T(\varphi) = 0$ in $K(\varphi_0, r)$ and $\|\hat{\varphi}_T - \varphi_0\| \leq \delta$ tends to one, which establishes (B.2).

That $\hat{\varphi}_T$ is asymptotically Gaussian follows from condition (A.1) in Lemma 1 and by Taylor's formula for the functions $\partial \ell_T(\varphi)/\partial \varphi_j, j = 1, \ldots, k$:

$$\sqrt{T} D\ell_T(\varphi_0) = (\Omega_I + A_T(\hat{\varphi}_T))\sqrt{T}(\hat{\varphi}_T - \varphi_0). \tag{A.3}$$

Here the elements in the matrix $A_T(\hat{\varphi}_T)$ are of the form $v_1'(D^2\ell_T(\varphi_T^*) - \Omega_I)v_2$ with v_1, v_2 unit vectors in $\mathbb{R}^k$ and φ_T^* a point on the line from φ_0 to $\hat{\varphi}_T$. Note that φ_T^* depends on the first vector v_1. Next, by expression (A.1),

$$|v_1'(D^2\ell_T(\varphi_T^*) - \Omega_I)v_2| \leq |v_1'(D^2\ell_T(\varphi_0) - \Omega_I)v_2| + \|v_1\|\|v_2\|\|\varphi_T^* - \varphi_0\|\tilde{c}_T.$$

Because $\varphi_T^* \xrightarrow{P} \varphi_0$ and $\tilde{c}_T \xrightarrow{P} \tilde{c} < \infty$ it follows from condition (A.2) in Lemma 1 that the right-hand side tends in probability to zero. Hence $A_T(\hat{\varphi}_T) \xrightarrow{P} 0$, and condition (B.3) follows by expression (A.3) using condition (A.1) in Lemma 1. ∎