Econometric Theory, **10**, 1994, 29-52. Printed in the United States of America.

ASYMPTOTIC THEORY FOR THE GARCH(1,1) QUASI-MAXIMUM LIKELIHOOD ESTIMATOR

SANG-WON LEE
Indiana University

BRUCE E. HANSEN
University of Rochester

This paper investigates the sampling behavior of the quasi-maximum likelihood estimator of the Gaussian GARCH(1,1) model. The rescaled variable (the ratio of the disturbance to the conditional standard deviation) is not required to be Gaussian nor independent over time, in contrast to the current literature. The GARCH process may be integrated ($\alpha + \beta = 1$), or even mildly explosive ($\alpha + \beta > 1$). A bounded conditional fourth moment of the rescaled variable is sufficient for the results. Consistent estimation and asymptotic normality are demonstrated, as well as consistent estimation of the asymptotic covariance matrix.

1. INTRODUCTION

Explicit models of heteroskedasticity have a long history in statistics and econometrics. Engle [9] proposed a popular time-series model of heteroskedasticity. His concept of autoregressive conditional heteroskedasticity (ARCH) literally revolutionized empirical work in financial economics, primarily in the modeling of stock returns, interest rates, and foreign exchange rates. ARCH specifies the conditional variance as a linear function of past squared disturbances, and suggests estimation by maximum likelihood. Recent contributions have extended the ARCH model to a wider class of specifications, the most important of which is the generalized ARCH (GARCH) model of Bollerslev [5]. The Gaussian GARCH(1,1) model has become the workhorse of the industry, with the largest number of applications. For a recent survey of the enormous number of empirical applications of the ARCH methodology, see Bollerslev, Chou, and Kroner [7].

Despite the large empirical literature, there is a rather sparse literature investigating the sampling properties of the estimation techniques. Yet it is by no means obvious that the maximum likelihood estimator (MLE) will be

Our thanks go to Don Andrews, Adrian Pagan, and two referees for helpful comments. Bruce Hansen thanks the National Science Foundation for research support. This paper was written while Sang-Won Lee was a Sloan Foundation Doctoral Dissertation Fellow.

 0266-4666/94 $5.00 + .00

consistent or asymptotically normal in relevant modeling situations. As documented by Bollerslev and Engle [6], estimated GARCH models typically suggest that the underlying data are "integrated." Nelson [13] studied the behavior of the "IGARCH" process. One well-known finding is that the unconditional mean of the IGARCH conditional variance is not finite. Thus, the model implies that the data may have an infinite variance, and it is known that regression with infinite variance processes typically leads to non-normal limiting distributions of the estimator. The existing literature for infinite variance processes, however, concerns linear analysis, which is not applicable to the context of GARCH models which involve nonlinear functions of the entire past history of the data.

Weiss [16] provided the first study of the asymptotic properties of the ARCH MLE. He showed that the MLE is consistent and asymptotically normal, requiring that the unnormalized data have finite fourth moments. This effectively rules out most interesting GARCH processes (and of course all IGARCH processes).

Bollerslev and Wooldridge [8] derived the large sample distribution of the quasi-maximum likelihood estimator under high-level assumptions (asymptotic normality of the score vector and uniform weak convergence of the likelihood and its second derivative). They did not verify these conditions or show how they might be verified for actual GARCH models.

Lumsdaine [12] was the first to study the asymptotic theory for GARCH models allowing for possibly integrated processes. She showed that there exists a consistent root of the likelihood equation, and that this root is asymptotically normally distributed. Her innovation over Weiss was to impose assumptions upon the *rescaled variable*—the ratio of the disturbance to the conditional standard deviation—rather than upon the observed data. As auxiliary assumptions, Lumsdaine assumed that the rescaled variable is independent and identically distributed (i.i.d.) and drawn from a symmetric unimodal density with the 32nd moment finite.

In this paper, we extend this literature to encompass a much broader class of GARCH processes. We are motivated by the concept of quasi-maximum likelihood estimation. In our framework, the researcher has specified the conditional mean and variance equations correctly, and then uses the Gaussian likelihood as a vehicle to estimate these parameters. In this framework, there is no reason to assume that all of the conditional dependence is contained in the conditional mean and variance, so the rescaled variable need not be (in fact, is unlikely to be) independent over time. Therefore, it is necessary to allow this variable to be time dependent. We do so by specifying that it is strictly stationary and ergodic. Note in contrast that the previous literature has assumed that the rescaled variable is i.i.d.

In addition, we do not restrict our analysis to GARCH and IGARCH processes. Instead, we only require that the series satisfies the necessary and sufficient condition for stationarity as given by Nelson [13], which allows for mildly explosive GARCH processes. No conditions on the shape of the distribution are required other than conditional moment bounds, and the innovations are not required to possess a density.

As in Lumsdaine [12], we are only able to prove the existence of a consistent root of the likelihood if we allow for IGARCH. For this result, we require that the conditional $2 + \delta$ moment of the rescaled variable is uniformly bounded. Restricting attention to nonintegrated GARCH models, we provide the first consistency proof for the quasi-maximum likelihood estimator. (Recall, Weiss [16] studied ARCH estimation.) Asymptotic normality is proved (again including the IGARCH case) by adding the assumption that the conditional fourth moment of the rescaled variable is uniformly bounded.

The order of this paper is as follows. Section 2 presents the model and the likelihood function. Section 3 provides two consistency proofs: one that allows for IGARCH processes and one that does not. Section 4 derives the large sample distribution theory for the quasi-MLE and demonstrates that a robust covariance matrix estimator is consistent for the asymptotic variance of the quasi-MLE. Section 5 concludes. The mathematical proofs are presented in the Appendix.

We use the following notation throughout the paper. $|A| = (\mathrm{tr}(A'A))^{1/2}$ denotes the Euclidean norm of a matrix or vector, and $\|A\|_r = (E|A|^r)^{1/r}$ denotes the L^r-norm of a random matrix or vector. All limits are taken as the sample size n diverges to positive infinity.

2. QUASI-LIKELIHOOD

Suppose that we observe some sequence $\{y_t\}$ with

$$y_t = \gamma_0 + \epsilon_t, \qquad t = 1, \ldots, n,$$

where $E(\epsilon_t | \mathcal{F}_{t-1}) = 0$ a.s. and $\mathcal{F}_t = \sigma(\epsilon_t, \epsilon_{t-1}, \ldots)$. Define the conditional variance $h_{0t} \equiv E(\epsilon_t^2 | \mathcal{F}_{t-1})$. We assume that h_{0t} follows a GARCH process:

$$h_{0t} = \omega_0(1 - \beta_0) + \alpha_0 \epsilon_{t-1}^2 + \beta_0 h_{0t-1} \text{ a.s.}$$

An equivalent expression for the conditional variance can be derived as

$$h_{0t} = \omega_0 + \alpha_0 \sum_{k=0}^{\infty} \beta_0^k \epsilon_{t-1-k}^2 \text{ a.s.}$$

This process is described by the parameter vector $\theta_0 \equiv [\gamma_0 \quad \omega_0 \quad \alpha_0 \quad \beta_0]'$.

The model for the unknown parameters $\theta \equiv [\gamma \quad \omega \quad \alpha \quad \beta]'$ is

$$y_t = \gamma + e_t, \qquad t = 1, \ldots, n,$$

and

$$h_t^*(\theta) = \omega(1 - \beta) + \alpha e_{t-1}^2 + \beta h_{t-1}^*(\theta), \qquad t = 2, \ldots, n$$

with the start-up condition

$$h_1^*(\theta) = \omega.$$

This gives the convenient expression for the variance process

$$h_t^*(\theta) = \omega + \alpha \sum_{k=0}^{t-2} \beta^k e_{t-1-k}^2.$$

Define the compact parameter space

$$\Theta \equiv \{\theta : \gamma_l \le \gamma \le \gamma_u,\ 0 < \omega_l \le \omega \le \omega_u,\ 0 < \alpha_l \le \alpha \le \alpha_u,$$
$$0 < \beta_l \le \beta \le \beta_u < 1\}.$$

We assume that $\theta_0 \in \Theta$. This implies that $\alpha_0 > 0$ and $\beta_0 > 0$, which means that ϵ_t is strictly a GARCH process, ruling out the possibility of ϵ_t being either a pure ARCH process or even an i.i.d. process.

We can define the rescaled variable $z_t = \epsilon_t / h_{0t}^{1/2}$. By construction, $E(z_t \mid \mathfrak{F}_{t-1}) = 0$ a.s. and $E(z_t^2 \mid \mathfrak{F}_{t-1}) = 1$ a.s. Estimation of GARCH models is frequently done under the assumption that $z_t \sim$i.i.d. $N(0,1)$ so that the likelihood is easily specified. We follow this practice by assuming that the Gaussian likelihood is used to form the estimator. Thus, the log likelihood takes the form (ignoring constants)

$$L_n^*(\theta) = \frac{1}{2n} \sum_{t=1}^{n} l_t^*(\theta) \qquad \text{where} \quad l_t^*(\theta) \equiv -\left(\ln h_t^*(\theta) + \frac{e_t^2}{h_t^*(\theta)}\right).$$

Since the likelihood need not be the correct density, it is typically referred to as a quasi-likelihood, although we will sometimes refer to it as a likelihood for brevity.

3. CONSISTENCY OF THE QUASI-MLE

It will be convenient at times to work with the unobserved variance processes

$$h_t(\theta) = \omega + \alpha \sum_{k=0}^{\infty} \beta^k e_{t-k-1}^2$$

$$h_t^\epsilon(\theta) = \omega + \alpha \sum_{k=0}^{\infty} \beta^k \epsilon_{t-k-1}^2$$

and the unobserved likelihood

$$L_n(\theta) = \frac{1}{2n} \sum_{t=1}^{n} l_t(\theta), \qquad \text{where} \quad l_t(\theta) \equiv -\left(\ln h_t(\theta) + \frac{e_t^2}{h_t(\theta)}\right).$$

The process $h_t(\theta)$ is the model of the conditional variance when the infinite past history of the data is observed. $h_t^\epsilon(\theta)$ is the same, except that it is a function of the true innovations ϵ_t rather than the residuals e_t. It turns out that these two variance processes are close up to scale:

LEMMA 1. *$B^{-1}h_t^{\epsilon}(\theta) \le h_t(\theta) \le Bh_t^{\epsilon}(\theta)$ a.s., where*

$$B = 1 + 2(1-\beta_u)^{-1/2}(\gamma_u - \gamma_l)\cdot\max\left(\frac{\alpha_u}{\omega_l}, 1\right) + \frac{\alpha_u}{\omega_l(1-\beta_u)}(\gamma_u - \gamma_l)^2. \quad \blacksquare$$

Assumption A.1

(i) z_t is strictly stationary and ergodic.
(ii) z_t^2 is nondegenerate.
(iii) For some $\delta > 0$, there exists an $S_\delta < \infty$ such that

$$E(z_t^{2+\delta} \mid \mathfrak{F}_{t-1}) \le S_\delta < \infty \text{ a.s.}$$

(iv) $\sup_t E(\ln(\beta_0 + \alpha_0 z_t^2) \mid \mathfrak{F}_{t-1}) < 0$ a.s. $\blacksquare$

The conditions imposed in (A.1) are fairly weak, although they are probably not the weakest possible. It would be desirable to relax the strict stationarity of (i) to some form of weak dependence, such as mixing or near-epoch dependence, but this is beyond our technical capabilities. Strict stationarity does, however, greatly relax the universally made assumption of the preexisting literature that z_t is independent and identically distributed. Condition (iii) might also be stronger than necessary. It trivially holds for i.i.d. data when $E(z_t^{2+\delta}) < \infty$ for some $\delta > 0$. The strengthening to uniformity over the conditional expectation controls conditional heterogeneity.

Nelson [13] showed that an analog of condition (iv), $E\ln(\beta_0 + \alpha_0 z_t^2) < 0$ is necessary and sufficient for the strict stationarity of z_t. Note by Jensen's inequality that (iv) holds if $\beta_0 + \alpha_0 \le 1$ and $\sup_t P(|z_t^2 - 1| \le \epsilon \mid \mathfrak{F}_{t-1}) < 1$ for some $\epsilon > 0$, since $E(z_t^2 \mid \mathfrak{F}_{t-1}) = 1$ a.s. But the condition does not *require* that $\beta_0 + \alpha_0 \le 1$, as pointed out by Nelson [13]. Thus, we are allowing for the possibility of mildly explosive GARCH, in addition to integrated GARCH.

The method of proof for consistency and asymptotic normality is standard in the sense that the results of Amemiya ([1], Theorem 4.1.1, Theorem 4.1.2, and Theorem 4.1.3) can be applied. Further, we can use general forms of the ULLN (specifically, those in Andrews [2]) to verify the required conditions. Therefore, the main task involved in the proof is to show the finiteness of various moments, such as those of the log likelihood, the score, and higher order derivatives. These proofs are quite demanding because these functions are nonlinear functions of the underlying innovations, where the latter need not have finite second moments. Most expressions can be reduced, however, to ratios of geometric averages in the squared innovations, and these can be bounded in expectation.

One potential difficulty is that the observed likelihood process $l_t^*(\theta)$ is not a stationary process due to the presence of the startup condition. The unobserved process $l_t(\theta)$ is, however, and this observation will be instrumental in our proof methods. This fact is stated in the following result, which is a generalization of the results of Nelson [13] in the context of i.i.d. z_t.

LEMMA 2. *Under* (A.1):

(1) h_{0t} *is strictly stationary and ergodic for all* t.
(2) $l_t(\theta)$ *and its first and second derivatives are strictly stationary and ergodic for all* $\theta \in \Theta$.
(3) *For some* $1 > p > 0$, *and all* $\theta \in \Theta$, $E|h_t(\theta)|^p \le H_p < \infty$. ■

The analysis of the unobserved process $l_t(\theta)$ can only be justified if the difference between it and the observed likelihood process is asymptotically negligible. This will hold only if the initial conditions become negligible (in some well-defined sense). Asymptotic negligibility of initial conditions in an integrated GARCH model is not immediately obvious since the conditional variance is a martingale plus drift. We find, however, that this is unimportant for the quasi-likelihood function.

LEMMA 3. *Under* (A.1), $\sup_{\theta\in\Theta}|L_n(\theta) - L_n^*(\theta)| \to_p 0$. ■

While Lemma 3 allows us to neglect the initial conditions in the asymptotic analysis, it still may be the case that proper handling of the initial conditions may have important effects in finite samples.

The following lemma provides some basic results which will be used repeatedly in the subsequent proofs.

LEMMA 4. *Under* (A.1):

(1) $P\{z_t^2 \le \frac{1}{2} | \mathcal{F}_{t-1}\} \le \mathcal{P}$ *a.s. where* $\mathcal{P} \equiv 1 - [1/(2^{(2+\delta)/\delta} S_\delta^{2/\delta})] \in (0,1)$.
(2) *For all* $\psi > 0$ *and all* $r \ge 1$,

$$\beta^r E\left(\left(\frac{1}{\beta + \psi z_t^2}\right)^r \middle| \mathcal{F}_{t-1}\right) \le E\left(\frac{1}{1 + \psi z_t^2} \middle| \mathcal{F}_{t-1}\right) \le \mathcal{R}(\psi) \text{ a.s.}$$

where $\mathcal{R}(\psi) \equiv (2 + \psi\mathcal{P})/(2 + \psi) < 1$.

(3) *For all finite* r,

$$E\left(\left(\frac{h_{0t-k}}{h_{0t}}\right)^r \middle| \mathcal{F}_{t-k-1}\right) \le \left(\frac{\mathcal{R}_0}{\beta_0^r}\right)^k \text{ a.s.}$$

where $\mathcal{R}_0 = \mathcal{R}(\alpha_0) < 1$.

(4) *If* $\beta \le \beta_0$, $\dfrac{h_t^\epsilon(\theta)}{h_{0t}} \le K_l \equiv \dfrac{\omega_u}{\omega_0} + \dfrac{\alpha_u}{\alpha_0} < \infty$ a.s.

(5) *If* $\beta \ge \beta_0$, $\dfrac{h_{0t}}{h_t^\epsilon(\theta)} \le H_u \equiv \dfrac{\omega_0}{\omega_l} + \dfrac{\alpha_0}{\alpha_l} < \infty$ a.s. ■

Part (2) of this lemma is of major importance for the proofs of this paper. One may question why the need for the supremum over t in part (2), since for all t,

$$E\left(\frac{1}{1+\psi z_t^2}\,\middle|\,\mathcal{F}_{t-1}\right) < 1.$$

With non-i.i.d. z_t, however, the conditional expectation is a random variable which (in principle) could take on values arbitrarily close to unity, which we need to exclude.

The inequality in part (1) is sufficient for part (2), and this is the only place in the paper where we use part (1). Hence, if the bound in part (2) could be established using other means, then the inelegant part (1) could be avoided. We feel that the large role the bound $\mathcal{P}$ plays in Lemma 4 and hence the resulting asymptotic theory is more an artifact of the proof of parts (1) and (2) of Lemma 4 than an important condition for distributional theory. It should be noted that the approach taken here differs fundamentally from the approach taken by the earlier literature. Lumsdaine [12], for example, uses the theory of expectations of ratios of quadratic forms of independent random variables, which does not have an immediate generalization to the dependent case.

We now establish some further bounds on the ratio of the unobserved variance process h_{0t} to its "estimate" $h_t(\theta)$. Part of the difficulty in dealing with potentially IGARCH processes is that when $\beta_0 + \alpha_0 = 1$, $Eh_t(\theta)$ is infinite. Thus, it is necessary to always work with appropriately selected ratios of random variables. It turns out to be particularly important to be able to bound the ratio $h_{0t}/h_t(\theta)$ and its inverse uniformly in $\theta \in \Theta$. The most difficult element of the parameter vector θ to handle is β. Note that $h_t(\theta)$ is a weighted average of squares of the data, with β controlling this weighting. Our method of establishing the needed bounds is to split the parameter space.

Set

$$\mathcal{R}_l = \mathcal{R}(K_l^{-1}\alpha_l) < 1$$

and pick positive constants η_l and η_u, which satisfy

$$\eta_l < \beta_0(1 - \mathcal{R}_l^{1/6})$$

and

$$\eta_u < \beta_0(1 - \mathcal{R}_0^{1/6}).$$

($\mathcal{R}_0$ is defined in Lemma 4(3).) Define for $1 \le r \le 6$ the constants

$$\beta_{rl} = \beta_0 \mathcal{R}_l^{1/r} + \eta_l < \beta_0,$$

$$\beta_{ru} = \frac{\beta_0 - \eta_u}{\mathcal{R}_0^{1/r}} > \beta_0,$$

and the subspaces

$$\Theta_l^r = \{\theta \in \Theta : \beta_{rl} \leq \beta \leq \beta_0\}$$

$$\Theta_u^r = \{\theta \in \Theta : \beta_0 \leq \beta \leq \beta_{ru}\}$$

$$\Theta_r = \Theta_l^r \cup \Theta_u^r.$$

In the theory that follows, it is clear that small values of η_l and η_u detract from accuracy of the asymptotic approximations. The ability to allow η_l and η_u to be large depends on the constants $\mathcal{R}_l$ and $\mathcal{R}_0$ which are function of the parameter space Θ and the constant $\mathcal{P}$.

LEMMA 5. *Under* (A.1), *for* $1 \leq r \leq 6$:

(1) $\beta^k \left\| \dfrac{h_{t-k}^\epsilon(\theta)}{h_t^\epsilon(\theta)} \right\|_r \leq (\mathcal{R}_l^{1/r})^k$ *uniformly in* $\theta \in \Theta_l^r$.

(2) $\left\| \dfrac{h_{0t}}{h_t^\epsilon(\theta)} \right\|_r \leq H_c \equiv \dfrac{\omega_0}{\omega_l} + \dfrac{\alpha_0}{\alpha_l \eta_l} < \infty$ *uniformly in* $\theta \in \Theta_r$. ■

LEMMA 6. *Under* (A.1), *for* $1 \leq r \leq 6$:

(1) $\left\| \dfrac{h_t^\epsilon(\theta)}{h_{0t}} \right\|_r \leq K_u \equiv \dfrac{\omega_u}{\omega_0} + \dfrac{\alpha_u}{\eta_u \alpha_0} < \infty$ *uniformly in* $\theta \in \Theta_u^r$.

(2) $\left\| \dfrac{h_{t-k}^\epsilon(\theta)}{h_t^\epsilon(\theta)} \right\|_r \leq K_u H_u \left(\dfrac{\mathcal{R}_0^{1/r}}{\beta_0} \right)^k$ *uniformly in* $\theta \in \Theta_u^r$.

(3) $\left\| \dfrac{h_{t-k}^\epsilon(\theta)}{h_t^\epsilon(\theta)} \right\|_r \leq K_u H_u \left(\dfrac{\mathcal{R}_0^{1/r}}{\beta_0} \right)^k$ *uniformly in* $\theta \in \Theta_u^r$. ■

We are now in a position to establish the pointwise convergence of the unobserved likelihood process.

LEMMA 7. *Under* (A.1), *for all* $\theta \in \Theta_1$:

(1) $E \dfrac{e_t^2}{h_t(\theta)} \leq H_1 \equiv \dfrac{(\gamma_u - \gamma_l)^2}{\omega_l} + BH_c$.

(2) $L_n(\theta) \to_p L(\theta) = E \dfrac{l_t(\theta)}{2}$. ■

The pointwise convergence of Lemma 7(2) is not sufficient for consistency of the quasi-MLE. It is necessary as well to establish uniform convergence and identification. A sufficient condition for uniform convergence is uniform boundedness of the expected value of the score (the vector of derivatives of the likelihood with respect to θ) which we denote by $\nabla l_t(\theta)$. We now explicitly derive such bounds. First, we need to bound the derivatives of the variance process. Lumsdaine [12, Lemma 1] showed that the quantities

$$\frac{\partial h_t(\theta)}{\partial \gamma}\frac{1}{h_t(\theta)}, \qquad \frac{\partial h_t(\theta)}{\partial \omega}\frac{1}{h_t(\theta)}, \qquad \frac{\partial h_t(\theta)}{\partial \alpha}\frac{1}{h_t(\theta)}$$

are all naturally bounded by functions of the parameter space Θ. This is not true for the derivative with respect to β, however. To simplify the notation, set

$$h_{\beta t}(\theta) = \frac{\partial h_t(\theta)}{\partial \beta}\frac{1}{h_t(\theta)}.$$

LEMMA 8. *Under* (A.1):

(1) *For* $1 \le r \le 6$, $\sup_{\theta\in\Theta_r}\|h_{\beta t}(\theta)\|_r \le B^2 \max\left(\frac{1}{(1-\Re_l)}, \frac{K_u H_u}{\eta_u}\right) \equiv H_\beta < \infty$.

(2) $\sup_{\theta\in\Theta_2} E|\nabla l_t(\theta)| < \infty$. ■

Now we can prove the local consistency of the quasi-MLE. Define the estimator

$$\hat{\theta}_n = \arg\max_{\theta\in\Theta_2} L_n^*(\theta).$$

$\hat{\theta}_n$ is the parameter value which maximizes the likelihood in the restricted region $\Theta_2 \subset \Theta$. It may or may not equal the global MLE

$$\tilde{\theta}_n = \arg\max_{\theta\in\Theta} L_n^*(\theta).$$

We can, however, establish the consistency of $\hat{\theta}_n$.

THEOREM 1. *Under* (A.1), $\hat{\theta}_n \to_p \theta_0$. ■

Theorem 1 shows that there exists a consistent root of the likelihood equation. Unfortunately, this is not sufficient for consistency of the global quasi-MLE, $\tilde{\theta}_n$. We conjecture that $\tilde{\theta}_n$ is indeed consistent, but a formal proof appears to be quite challenging.

If we restrict attention to nonintegrated GARCH processes, we can establish the consistency of the quasi-MLE under the same conditions. Since this has not been established before, we state this result formally.

THEOREM 2. *Under* (A.1) *and* $\alpha_0 + \beta_0 < 1$, $\tilde{\theta}_n \to_p \theta_0$. ■

4. ASYMPTOTIC NORMALITY OF THE QUASI-MLE

We now turn to the derivation of the large-sample distribution of the quasi-maximum likelihood estimator. For this development, we need a stronger condition.

Assumption A.2

(i) $E(z_t^4 \mid \mathcal{F}_{t-1}) \le \mathcal{K} < \infty$ a.s.
(ii) θ_0 is in the interior of Θ. ■

Assumption A.2(i) strengthens condition (iii) of Assumption A.1 to a uniformly bounded conditional fourth moment for the rescaled innovations z_t. Note that it is trivially satisfied by any i.i.d. process with finite fourth moment.

We first establish a functional central limit theorem for the score function evaluated at θ_0, which implies the standard CLT required for the proof of asymptotic normality. This extra generality will not be used in this paper, but comes at no cost and could be useful in other applications. Let $\Rightarrow$ denote weak convergence of probability measures with respect to the uniform metric, let $W(r)$ denote a Brownian motion with covariance matrix I_4, and let $[\cdot]$ denote integer part.

LEMMA 9. *Under* (A.1) *and* (A.2):

(1) *For all* $\theta \in \Theta_4$, $E|\nabla l_t(\theta)\nabla l_t(\theta)'| < \infty$.

(2) $\frac{1}{\sqrt{n}} A_0^{-1/2} \sum_{t=1}^{[nr]} \nabla l_t^*(\theta_0) \Rightarrow W(r)$ *where* $A_0 = E(\nabla l_t(\theta_0)\nabla l_t(\theta_0)')$. ■

To complete the derivation of the large-sample distribution of $\hat{\theta}_n$, we need to examine the 4×4 matrix of second derivatives of the likelihood which we denote by $\nabla^2 l_t(\theta)$. We can simplify the notation somewhat by defining

$$h_{\beta\beta t}(\theta) = \frac{\partial^2 h_t(\theta)}{\partial \beta^2}\,\frac{1}{h_t(\theta)}, \qquad h_{\beta\beta\beta t}(\theta) = \frac{\partial^3 h_t(\theta)}{\partial \beta^3}\,\frac{1}{h_t(\theta)}.$$

LEMMA 10. *Under* (A.1), *for* $1 \le r \le 6$:

(1) $\sup_{\theta \in \Theta_{2r}} \| h_{\beta\beta t}(\theta) \|_r \le 2H_\beta^2 < \infty$.
(2) $\sup_{\theta \in \Theta_{3r}} \| h_{\beta\beta\beta t}(\theta) \|_r \le 6H_\beta^3 < \infty$. ■

Set

$$\hat{B}_n(\theta) = -\frac{1}{n} \sum_{t=1}^{n} \nabla^2 l_t^*(\theta).$$

and

$$B(\theta) = -E\nabla^2 l_t(\theta).$$

LEMMA 11. *Under* (A.1) *and* (A.2):

(1) *For all* $\theta \in \Theta_4$ $E|\nabla^2 l_t(\theta)| < \infty$.

(2) *For all* $\theta \in \Theta_4$ *and* $i = 1,2,3,4$, $E\left|\frac{\partial}{\partial \theta_i} \nabla^2 l_t(\theta)\right| < \infty$ *where* θ_i *is the ith element of* θ.

(3) $\sup_{\theta \in \Theta_4} |\hat{B}_n(\theta) - B(\theta)| \to_p 0$ *and* $B(\theta)$ *is continuous in* Θ_4. ■

We are now in a position to state our main distributional result.

THEOREM 3. *Under* (A.1) *and* (A.2):

$$\sqrt{n}(\hat{\theta}_n - \theta_0) \rightarrow_D N(0, V_0)$$

where $V_0 = B_0^{-1} A_0 B_0^{-1}$ *and* $B_0 = B(\theta_0) = -E\nabla^2 l_t(\theta_0)$.

Set

$$\hat{B}_n = \hat{B}_n(\hat{\theta}_n). \qquad \blacksquare$$

The matrix $\hat{B}_n^{-1}$ would be the standard estimate of the covariance matrix for $\hat{\theta}_n$ in the context of maximum likelihood. In the more general context of quasi-likelihood estimation, however, the asymptotic covariance matrix is $B_0^{-1}A_0B_0^{-1}$ not B_0^{-1}, so $\hat{B}_n^{-1}$ will not be a consistent estimate for this quantity. To construct a consistent estimate, define

$$\hat{A}_n(\theta) = \frac{1}{n}\sum_{t=1}^{n} \nabla l_t^*(\theta)\, \nabla l_t^*(\theta)',$$

$$\hat{A}_n = \hat{A}_n(\hat{\theta}_n),$$

and

$$A(\theta) = E\nabla l_t(\theta)\, \nabla l_t(\theta)'.$$

LEMMA 12. *Under* (A.1) *and* (A.2):

(1) $\sup_{\theta \in \Theta_6} |\hat{A}_n(\theta) - A(\theta)| \rightarrow_p 0$, *and* $A(\theta)$ *is continuous in* Θ_6.

(2) $\hat{V}_n = \hat{B}_n^{-1}\hat{A}_n\hat{B}_n^{-1} \rightarrow_p V_0 = B_0^{-1}A_0B_0^{-1}$. $\blacksquare$

Lemma 12 completes our characterization of the classical properties of the quasi-MLE for the GARCH(1,1) model. It shows that the conventional "robust" covariance matrix estimator is consistent for the asymptotic variance of the parameter estimator.

5. CONCLUSION

Models of conditional heteroskedasticity are routinely used in applied econometrics. This paper has explored the distributional theory for one simple example—the Gaussian GARCH(1,1) model with an intercept. The potential generalizations of this simple case are numerous. To state a few: inclusion of additional regressors, GARCH(p,q) models, non-Gaussian likelihoods, and various nonlinear GARCH models. Unfortunately, the methods used in this paper to analyze the Gaussian GARCH(1,1) model are quite cumbersome and may not easily generalize to more complicated models. We believe that different approaches may be required in such cases.

This paper has shown that the Gaussian likelihood will consistently estimate the parameters of the GARCH(1,1) model even if the rescaled variable is neither Gaussian nor independent. This is certainly an advance over the existing literature, which has universally assumed that the rescaled variable is independent. We are maintaining another important and unsatisfactory assumption, however. Throughout the analysis, we assumed that the true conditional variance is described by the GARCH equation. This is less than satisfactory, because most applied work implicitly views the GARCH equation as an *approximation* to the true conditional variance. It would be interesting to know the properties of the quasi-MLE when the true conditional variance does not satisfy the GARCH equation. This would be a challenging, yet rewarding, task for future research.

REFERENCES

1. Amemiya, T. *Advanced Econometrics.* Cambridge: Harvard University Press, 1985.
2. Andrews, D.W.K. Generic uniform convergence. *Econometric Theory* 8 (1992): 241–257.
3. Billingsley, P. *Convergence of Probability Measures.* New York: Wiley, 1968.
4. Billingsley, P. *Probability and Measure.* New York: Wiley, 1979.
5. Bollerslev, T. Generalized autoregressive conditional heteroskedasticity. *Journal of Econometrics* 31 (1986): 307–327.
6. Bollerslev, T. & R.F. Engle. Common persistence in conditional variances. *Econometrica* 61 (1993): 167–186.
7. Bollerslev, T., R.Y. Chou & K.F. Kroner. ARCH modelling in finance: A review of the theory and empirical evidence. *Journal of Econometrics* 52 (1992): 5–59.
8. Bollerslev, T. & J.M. Wooldridge. Quasi-maximum likelihood estimation and inference in dynamic models with time-varying covariances. *Econometric Reviews* 11 (1992): 143–172.
9. Engle, R.F. Autoregressive conditional heteroskedasticity with estimates of the variance of United Kingdom inflation. *Econometrica* 50 (1982): 987–1007.
10. Engle, R.F. & T. Bollerslev. Modelling the persistence of conditional variances. *Econometric Reviews* 5 (1986): 1–50.
11. Hardy, G.H., J.E. Littlewood & G. Polya. *Inequalities.* Cambridge: Cambridge University Press, 1952.
12. Lumsdaine, R.L. Asymptotic Properties of the Quasi-Maximum Likelihood Estimator in GARCH(1,1) and IGARCH(1,1) Models. Mimeo, Princeton University, 1991.
13. Nelson, D.B. Stationarity and persistence in the GARCH(1,1) model. *Econometric Theory* 6 (1990): 318–334.
14. Pantula, S.G. Autoregressive Conditionally Heteroscedastic Models. Mimeo, North Carolina State University, 1984.
15. Stout, W.F. *Almost Sure Convergence.* New York: Academic Press, 1974.
16. Weiss, A.A. Asymptotic theory for ARCH models: Estimation and testing. *Econometric Theory* 2 (1986): 107–131.

APPENDIX

In this Appendix, all of the equalities or inequalities hold *almost surely* if applicable.

Proof of Lemma 1. Define $g \equiv \gamma_0 - \gamma$, so $\epsilon_t = e_t + \gamma - \gamma_0 = e_t - g$. Then

$$
\begin{aligned}
\frac{h_t^\epsilon(\theta)}{h_t(\theta)} &= \frac{\omega + \alpha \sum_{k=0}^{\infty} \beta^k (e_{t-1-k} - g)^2}{h_t(\theta)} \\
&\le \frac{\omega + \alpha \sum_{k=0}^{\infty} \beta^k e_{t-1-k}^2 + 2\alpha |g| \sum_{k=0}^{\infty} \beta^k |e_{t-1-k}| + \alpha g^2/(1-\beta)}{h_t(\theta)} \\
&\le 1 + \frac{2\alpha \sum_{k=0}^{\infty} \beta^k |e_{t-1-k}|}{h_t(\theta)} |g| + \frac{\alpha}{\omega(1-\beta)} g^2,
\end{aligned}
$$

where the first inequality uses the triangle inequality and the second inequality uses the fact that $h_t(\theta) \ge \omega$. Using Schwarz's inequality,

$$
\begin{aligned}
\frac{\alpha \sum_{k=0}^{\infty} \beta^k |\epsilon_{t-1-k}|}{h_t(\theta)} &= \frac{\alpha \sum_{k=0}^{\infty} \beta^{k/2} (\beta^{k/2} |e_{t-1-k}|)}{h_t(\theta)} \\
&\le \frac{\alpha \left(\sum_{k=0}^{\infty} \beta^k \right)^{1/2} \left(\sum_{k=0}^{\infty} \beta^k e_{t-1-k}^2 \right)^{1/2}}{h_t(\theta)} \\
&= (1-\beta)^{-1/2} \frac{\alpha \left(\sum_{k=0}^{\infty} \beta^k e_{t-1-k}^2 \right)^{1/2}}{h_t(\theta)}.
\end{aligned}
$$

(i) If $0 \le \sum_{k=0}^{\infty} \beta^k e_{t-1-k}^2 \le 1$,

$$
\frac{\alpha \left(\sum_{k=0}^{\infty} \beta^k e_{t-1-k}^2 \right)^{1/2}}{h_t(\theta)} \le \frac{\alpha}{\omega}.
$$

(ii) If $\sum_{k=0}^{\infty} \beta^k e_{t-1-k}^2 > 1$,

$$
\frac{\alpha \left(\sum_{k=0}^{\infty} \beta^k e_{t-1-k}^2 \right)^{1/2}}{h_t(\theta)} \le \frac{\alpha \sum_{k=0}^{\infty} \beta^k e_{t-1-k}^2}{\omega + \alpha \sum_{k=0}^{\infty} \beta^k e_{t-1-k}^2} \le 1.
$$

Therefore,

$$
\frac{\alpha \sum_{k=0}^{\infty} \beta^k |e_{t-1-k}|}{h_t(\theta)} \le (1-\beta)^{-1/2} \max\left(\frac{\alpha}{\omega}, 1 \right).
$$

We have shown that

$$
\frac{h_t^\epsilon(\theta)}{h_t(\theta)} \le 1 + 2(1-\beta_u)^{-1/2} \max\left(\frac{\alpha_u}{\omega_l}, 1 \right) (\gamma_u - \gamma_l) + \frac{\alpha_u}{\omega_l(1-\beta_u)} (\gamma_u - \gamma_l)^2 \equiv \boldsymbol{B}.
$$

Similarly, noting that

$$h_t(\theta) = \omega + \alpha \sum_{k=0}^{\infty} \beta^k (\epsilon_{t-1-k} + g)^2,$$

analogous calculations reveal that $h_t(\theta)/h_t^\epsilon(\theta) \leq B$ as stated. ■

Proof of Lemma 2. (1) This is proved in Nelson ([13], Theorem 2) under the assumption of i.i.d. z_t. But the SLLN for stationary and ergodic processes (Stout [15], Theorem 3.5.7) can be used to cover the present case.

(2) From Billingsley ([4], Theorem 13.3), $h_t(\theta)$, e_t^2, $l_t(\theta)$ and its derivatives are measurable functions of ϵ_t, and thus are strictly stationary and ergodic as shown by Stout ([15], Theorem 3.5.8).

(3) Theorem 4 in Nelson [13] shows that under condition (iv) of Assumption (A.1), there exists some $0 < p < 1$ such that $Eh_{0t}^p < \infty$. Hence, using Lemma 1 and the fact that $\epsilon_{t-k-1}^2 \leq \alpha_0^{-1} h_{0t}$,

$$E(h_t(\theta))^p \leq B^p E(h_t^\epsilon(\theta))^p \leq B^p \left(\omega^p + \alpha^p \sum_{k=0}^{\infty} \beta^{kp} E\epsilon_{t-k-1}^{2p} \right)$$

$$\leq B^p \left(\omega_u^p + \frac{\alpha_u^p}{\alpha_0^p} \frac{1}{1-\beta_u^p} Eh_{0t}^p \right) < \infty. \quad ■$$

Proof of Lemma 3. First, note that

$$h_t(\theta) = h_t^*(\theta) + \beta^{t-1}\alpha \sum_{k=0}^{\infty} \beta^k e_{-k}^2 = h_t^*(\theta) + \beta^{t-1}(h_1(\theta) - \omega) \leq h_t^*(\theta) + \beta^{t-1} h_1(\theta).$$

Second,

$$L_n^*(\theta) - L_n(\theta) = \frac{1}{2n} \sum_{t=1}^{n} \ln\left(\frac{h_t(\theta)}{h_t^*(\theta)} \right) - \frac{1}{2n} \sum_{t=1}^{n} \left(\frac{e_t^2}{h_t^*(\theta)} - \frac{e_t^2}{h_t(\theta)} \right).$$

Third,

$$0 \leq \sum_{t=1}^{n} \ln\left(\frac{h_t(\theta)}{h_t^*(\theta)} \right) \leq \sum_{t=1}^{n} \ln\left(1 + \frac{\beta^{t-1} h_1(\theta)}{h_t^*(\theta)} \right) \leq \sum_{t=1}^{n} \frac{\beta^{t-1} h_1(\theta)}{\omega} \leq \frac{h_1(\theta)}{\omega_l (1-\beta_u)}.$$

Since $Eh_1(\theta)^p < \infty$ by Lemma 2(3), it follows by Markov's inequality that

$$0 \leq \frac{1}{n} \sum_{t=1}^{n} \ln\left(\frac{h_t(\theta)}{h_t^*(\theta)} \right) \leq \frac{1}{n} \frac{h_1(\theta)}{\omega_l(1-\beta_u)} \to_p 0.$$

Fourth,

$$0 \leq \sum_{t=1}^{n} \left(\frac{e_t^2}{h_t^*(\theta)} - \frac{e_t^2}{h_t(\theta)} \right) = \sum_{t=1}^{n} \frac{h_t - h_t^*}{h_t^*} \frac{e_t^2}{h_t} \leq \frac{h_1(\theta) - \omega}{\alpha \omega_l^2} \sum_{t=1}^{\infty} \beta^{t-1} h_{t+1} = X_2,$$

say. But

$$\| X_2 \|_{p/2} \leq \left(\frac{\| h_1(\theta) \|_p + \omega}{\alpha \omega_l^2} \right) \sum_{t=1}^{\infty} \beta^{t-1} \| h_t(\theta) \|_p \leq \left(\frac{H_p^{1/p} + \omega_u}{\alpha_l \omega_l^2 (1-\beta_l)} \right) H_p^{1/p} < \infty$$

by Lemma 2(3), and thus by Markov's inequality

$$0 \le \frac{1}{n}\sum_{t=1}^{n}\left(\frac{e_t^2}{h_t^*(\theta)} - \frac{e_t^2}{h_t(\theta)}\right) \le \frac{1}{n}X_2 \to_p 0$$

completing the proof. ■

Proof of Lemma 4. (1) Set $P_t = P\{z_t^2 \le \frac{1}{2} \mid \mathcal{F}_{t-1}\}$. Note that $P_t < 1$ for all t since $E(z_t^2 \mid \mathcal{F}_{t-1}) = 1$. Denote the distribution function of z_t conditional on $\mathcal{F}_{t-1}$ by F_t. Now we can define a new conditional distribution function $F_t^* = F_t/(1 - P_t)$ with positive support only on $\{z_t^2 > \frac{1}{2}\}$. Then using Assumption A.1(iii) and Lyapounov's inequality we find that with probability 1,

$$\begin{aligned}
S_\delta \ge E(z_t^{2+\delta} \mid \mathcal{F}_{t-1}) &\ge \int_{\{z_t^2 > 1/2\}} z_t^{2+\delta}\, dF_t \\
&= (1 - P_t)\int_{\{z_t^2 > 1/2\}} z_t^{2+\delta}\, dF_t^* \\
&\ge (1 - P_t)\left(\int_{\{z_t^2 > 1/2\}} z_t^2\, dF_t^*\right)^{(2+\delta)/2} \\
&= (1 - P_t)^{-\delta/2}\left(\int z_t^2\, dF_t - \int_{\{z_t^2 \le 1/2\}} z_t^2\, dF_t\right)^{(2+\delta)/2} \\
&\ge (1 - P_t)^{-\delta/2}(1 - \tfrac{1}{2}P_t)^{(2+\delta)/2} \\
&\ge (1 - P_t)^{-\delta/2}2^{-(2+\delta)/2}.
\end{aligned}$$

The stated result obtains by rearranging the terms. Note that $\mathcal{P} = 1 - 1/[2^{(2+\delta)/\delta}S_\delta^{2/\delta}] < 1$ since $S_\delta < \infty$, and $\mathcal{P} > 0$ since $S_\delta \ge E(z_t^{2+\delta} \mid \mathcal{F}_{t-1}) \ge (E(z_t^2 \mid \mathcal{F}_{t-1}))^{(2+\delta)/2} = 1$.

(2) First, since $x^r \le x$ when $x \le 1$ and $r \ge 1$,

$$\begin{aligned}
\beta^r E\left(\left(\frac{1}{\beta + \psi z_t^2}\right)^r \middle| \mathcal{F}_{t-1}\right) = E\left(\left(\frac{\beta}{\beta + \psi z_t^2}\right)^r \middle| \mathcal{F}_{t-1}\right) &\le E\left(\frac{\beta}{\beta + \psi z_t^2} \middle| \mathcal{F}_{t-1}\right) \\
&\le E\left(\frac{1}{1 + \psi z_t^2} \middle| \mathcal{F}_{t-1}\right)
\end{aligned}$$

a.s., where the final inequality follows since $\beta < 1$. Second, by part (1) of this lemma, with probability 1,

$$\begin{aligned}
E\left(\frac{1}{1 + \psi z_t^2} \middle| \mathcal{F}_{t-1}\right) &= \int_{\{z_t^2 \le 1/2\}} \frac{1}{1 + \psi z_t^2}\, dF_t + \int_{\{z_t^2 > 1/2\}} \frac{1}{1 + \psi z_t^2}\, dF_t \\
&\le \int_{\{z_t^2 \le 1/2\}} dF_t + \int_{\{z_t^2 > 1/2\}} \frac{1}{1 + \psi \frac{1}{2}}\, dF_t \\
&= P\left(z_t^2 \le \frac{1}{2} \middle| \mathcal{F}_{t-1}\right) + \frac{2}{2 + \psi} P\left(\left\{z_t^2 > \frac{1}{2}\right\} \middle| \mathcal{F}_{t-1}\right) \\
&= \frac{2 + \psi P_t}{2 + \psi} \le \frac{2 + \psi \mathcal{P}}{2 + \psi}.
\end{aligned}$$

(3) Since $\epsilon_{t-1}^2 = h_{0t-1} z_{t-1}^2$,

$$\frac{h_{0t-1}}{h_{0t}} = \frac{h_{0t-1}}{\omega_0(1-\beta_0) + \beta_0 h_{0t-1} + \alpha_0 h_{0t-1} z_{t-1}^2} \le \frac{1}{\beta_0 + \alpha_0 z_{t-1}^2},$$

and

$$\frac{h_{0t-k}}{h_{0t}} = \frac{h_{0t-1}}{h_{0t}} \frac{h_{0t-2}}{h_{0t-1}} \cdots \frac{h_{0t-k}}{h_{0t-k+1}} \le \prod_{i=1}^{k} \frac{1}{\beta_0 + \alpha_0 z_{t-i}^2}.$$

Applying part (2) of this lemma,

$$E\left(\left(\frac{h_{0t-k}}{h_{0t}}\right)^r \middle| \mathcal{F}_{t-k-1}\right) \le \prod_{i=1}^{k} E\left(\left(\frac{1}{\beta_0 + \alpha_0 z_{t-i}^2}\right)^r \middle| \mathcal{F}_{t-k-1}\right) \le \prod_{i=1}^{k} \frac{\mathcal{R}(\alpha_0)}{\beta_0^r} = \left(\frac{\mathcal{R}_0}{\beta_0^r}\right)^k.$$

(4) Using the fact that $\beta \le \beta_0$,

$$\frac{h_t^\epsilon(\theta)}{h_{0t}} = \frac{\omega + \alpha \sum_{k=0}^{\infty} \beta^k \epsilon_{t-1-k}^2}{\omega_0 + \alpha_0 \sum_{k=0}^{\infty} \beta_0^k \epsilon_{t-1-k}^2} \le \frac{\omega_u}{\omega_0} + \frac{\alpha_u}{\alpha_0} \equiv K_l.$$

(5) Since $\beta \ge \beta_0$,

$$\frac{h_{0t}}{h_t^\epsilon(\theta)} = \frac{\omega_0 + \alpha_0 \sum_{k=0}^{\infty} \beta_0^k \epsilon_{t-1-k}^2}{\omega + \alpha \sum_{k=0}^{\infty} \beta^k \epsilon_{t-1-k}^2} \le \frac{\omega_0}{\omega_l} + \frac{\alpha_0}{\alpha_l} \equiv H_u. \qquad \blacksquare$$

Proof of Lemma 5. (1) By Lemma 4(4),

$$\frac{h_t^\epsilon}{h_{t+1}^\epsilon} = \frac{h_t^\epsilon}{\omega(1-\beta) + \alpha\epsilon_t^2 + \beta h_t^\epsilon} \le \frac{1}{\alpha \dfrac{h_{0t}}{h_t^\epsilon} z_t^2 + \beta} \le \frac{1}{\beta + \alpha_l K_l^{-1} z_t^2}$$

and thus by Lemma 4(2)

$$E\left(\left(\frac{h_t^\epsilon}{h_{t+1}^\epsilon}\right)^r \middle| \mathcal{F}_{t-1}\right) \le E\left(\left(\frac{1}{\beta + \alpha_l K_l^{-1} z_t^2}\right)^r \middle| \mathcal{F}_{t-1}\right) \le \frac{\mathcal{R}(\alpha_l K_l^{-1})}{\beta^r} = \frac{\mathcal{R}_l}{\beta^r}.$$

Hence,

$$E\left(\left(\frac{h_{t-k}^\epsilon}{h_t^\epsilon}\right)^r \middle| \mathcal{F}_{t-k-1}\right) = E\left(\prod_{i=1}^{k} \left(\frac{h_{t-i}^\epsilon}{h_{t-i+1}^\epsilon}\right)^r \middle| \mathcal{F}_{t-k-1}\right)$$
$$= \prod_{i=1}^{k} E\left(\left(\frac{h_{t-i}^\epsilon}{h_{t-i+1}^\epsilon}\right)^r \middle| \mathcal{F}_{t-i-1}\right) \le \left(\frac{\mathcal{R}_l}{\beta^r}\right)^k.$$

Taking expectations of both sides yields the desired result.

(2) First consider $\theta \in \Theta_l^r$. Using Minkowski's inequality, the fact that $\epsilon_{t-1}^2 \le \alpha^{-1} h_t^\epsilon$ and part (1) of this lemma,

$$\left\| \frac{h_{0t}}{h_t^\epsilon(\theta)} \right\|_r \le \left\| \frac{\omega_0}{\omega} + \frac{\alpha_0 \sum_{k=0}^{\infty} \beta_0^k \epsilon_{t-k-1}^2}{h_t^\epsilon(\theta)} \right\|_r$$

$$\le \frac{\omega_0}{\omega} + \frac{\alpha_0}{\alpha} \sum_{k=0}^{\infty} \beta_0^k \left\| \frac{h_{t-k}^\epsilon}{h_t^\epsilon} \right\|_r$$

$$\le \frac{\omega_0}{\omega} + \frac{\alpha_0}{\alpha} \sum_{k=0}^{\infty} \left(\frac{\beta_0 \mathfrak{R}_l^{1/r}}{\beta_{rl}} \right)^k$$

$$= \frac{\omega_0}{\omega} + \frac{\alpha_0}{\alpha} \frac{\beta_{rl}}{\beta_{rl} - \beta_0 \mathfrak{R}_l^{1/r}}$$

$$= \frac{\omega_0}{\omega} + \frac{\alpha_0 \beta_{rl}}{\alpha \eta_l} \le \frac{\omega_0}{\omega_l} + \frac{\alpha_0}{\alpha_l \eta_l} \equiv H_c.$$

Lemma 4(5) established that

$$\frac{h_{0t}}{h_t^\epsilon(\theta)} \le \frac{\omega_0}{\omega_l} + \frac{\alpha_0}{\alpha_l} \equiv H_u \le H_c \qquad \text{for} \quad \beta \ge \beta_0,$$

so the bound holds trivially in Θ_u^r as well. ∎

Proof of Lemma 6. (1) Since $\epsilon_{t-1}^2 \le \alpha_0^{-1} h_{0t}$,

$$\frac{h_t^\epsilon}{h_{0t}} = \frac{\omega + \alpha \sum_{k=0}^{\infty} \beta^k \epsilon_{t-1-k}^2}{h_{0t}} \le \frac{\omega}{\omega_0} + \frac{\alpha}{\alpha_0} \sum_{k=0}^{\infty} \beta^k \frac{h_{0t-k}}{h_{0t}}.$$

By Minkowski's inequality and Lemma 4(3),

$$\left\| \frac{h_t^\epsilon}{h_{0t}} \right\|_r \le \frac{\omega_u}{\omega_0} + \frac{\alpha_u}{\alpha_0} \sum_{k=0}^{\infty} \beta_{ru}^k \left\| \frac{h_{0t-k}}{h_{0t}} \right\|_r \le \frac{\omega_u}{\omega_0} + \frac{\alpha_u}{\alpha_0} \sum_{k=0}^{\infty} \left(\frac{\beta_{ru} \mathfrak{R}_0^{1/r}}{\beta_0} \right)^k$$

$$= \frac{\omega_u}{\omega_0} + \frac{\alpha_u}{\alpha_0} \frac{\beta_0}{\beta_0 - \beta_{ru} \mathfrak{R}_0^{1/r}} \le \frac{\omega_u}{\omega_0} + \frac{\alpha_u}{\eta_u \alpha_0} \equiv K_u.$$

(2) Since by Lemma 4(5)

$$\frac{h_{t-k}^\epsilon}{h_t^\epsilon} = \frac{h_{0t}}{h_t^\epsilon} \frac{h_{t-k}^\epsilon}{h_{0t-k}} \frac{h_{0t-k}}{h_{0t}} \le H_u \frac{h_{t-k}^\epsilon}{h_{0t-k}} \frac{h_{0t-k}}{h_{0t}},$$

and $(h_{t-k}^\epsilon / h_{0t-k})^r$ is $\mathfrak{F}_{t-k-1}$ measurable, we have by Lemma 4(3) and part (1) of this lemma,

$$\left\| \frac{h_{t-k}^\epsilon}{h_t^\epsilon} \right\|_r \le H_u \left(E\left[\left(\frac{h_{t-k}^\epsilon}{h_{0t-k}} \right)^r \left(\frac{h_{0t-k}}{h_{0t}} \right)^r \right] \right)^{1/r}$$

$$\le H_u \left(E\left[\left(\frac{h_{t-k}^\epsilon}{h_{0t-k}} \right)^r E\left(\left(\frac{h_{0t-k+1}}{h_{0t}} \right)^r \middle| \mathfrak{F}_{t-k} \right) \right] \right)^{1/r}$$

$$\le H_u \left(E\left[\left(\frac{h_{t-k}^\epsilon}{h_{0t-k}} \right)^r \left(\frac{\mathfrak{R}_0}{\beta_0^r} \right)^k \right] \right)^{1/r} \le H_u \left(\frac{\mathfrak{R}_0^{1/r}}{\beta_0} \right)^k K_u.$$ ∎

Proof of Lemma 7. (1) Using the facts that $E(\epsilon_t|\mathcal{F}_{t-1}) = 0$, $h_t(\theta) \geq \omega_l$, Lemma 1, and Lemma 5(2),

$$E\left(\frac{e_t^2}{h_t(\theta)}\right) \leq BE\left(\frac{\epsilon_t^2}{h_t^\epsilon(\theta)}\right) + \frac{g^2}{\omega_l} = BE\left(\frac{h_{0t}}{h_t^\epsilon(\theta)}\right) + \frac{(\gamma_u - \gamma_l)^2}{\omega_l}$$
$$\leq BH_c + \frac{(\gamma_u - \gamma_l)^2}{\omega_l}.$$

(2) Note that by Jensen's inequality, part (1) of this lemma, and Lemma 2(3)

$$E|l_t(\theta)| \leq E|\ln h_t(\theta)| + E\left(\frac{e_t^2}{h_t(\theta)}\right) \leq \frac{1}{p}\ln E|h_t^p(\theta)| + H_1 \leq \frac{1}{p}\ln H_p + H_1 < \infty.$$

This allows the application of the strong law of large numbers for stationary and ergodic sequences, for example, Stout ([15], p. 181), which yields the desired result. ■

Proof of Lemma 8. (1) Differentiating both sides of $h_t = \omega + \beta h_{t-1} + \alpha e_{t-1}^2$ with respect to β we find

$$\frac{\partial h_t}{\partial \beta} = \beta \frac{\partial h_{t-1}}{\partial \beta} + h_{t-1}$$

or

$$h_{\beta t}(\theta) = \sum_{k=0}^{\infty} \beta^k \frac{h_{t-k-1}(\theta)}{h_t(\theta)}.$$

Using Lemma 1, and Minkowski's inequality,

$$\| h_{\beta t}(\theta) \|_r \leq B^2 \sum_{k=0}^{\infty} \beta^k \left\| \frac{h_{t-k-1}^\epsilon(\theta)}{h_t^\epsilon(\theta)} \right\|_r.$$

When $\theta \in \Theta_l^r$, this is bounded by $B^2 \sum_{k=0}^{\infty} (\mathcal{R}_l^{1/r})^k = B^2/(1 - \mathcal{R}_l^{1/r})$ using Lemma 5(1). When $\theta \in \Theta_u^r$, we have instead from Lemma 6(2) the bound $B^2 \sum_{k=0}^{\infty} \beta^k H_u K_u (\mathcal{R}_0^{1/r}/\beta_0) \leq B^2 K_u H_u/\eta_u$.

(2) First, note that

$$\frac{\partial l_t(\theta)}{\partial \omega} = \left(\frac{e_t^2}{h_t(\theta)} - 1\right)\frac{\partial h_t(\theta)}{\partial \omega}\frac{1}{h_t(\theta)}$$

and

$$\frac{\partial l_t(\theta)}{\partial \alpha} = \left(\frac{e_t^2}{h_t(\theta)} - 1\right)\frac{\partial h_t(\theta)}{\partial \alpha}\frac{1}{h_t(\theta)}.$$

These two gradients are therefore bounded in expectation since $E|(e_t^2/h_t(\theta)) - 1| < 1 + H_1$ by Lemma 7(1) and the quantities $(\partial h_t(\theta)/\partial\omega)/h_t(\theta)$ and $(\partial h_t(\theta)/\partial\alpha)/h_t(\theta)$ are naturally bounded as shown by Lumsdaine [12, Lemma 1].

Second, note that

$$\frac{\partial l_t(\theta)}{\partial \gamma} = \left(\frac{e_t^2}{h_t(\theta)} - 1\right)\frac{\partial h_t(\theta)}{\partial \gamma}\frac{1}{h_t(\theta)} - 2\frac{e_t}{h_t(\theta)}$$

and again $(\partial h_t(\theta)/\partial\gamma)/h_t(\theta)$ is bounded by Lumsdaine [12, Lemma 1], so it remains to show that $E|e_t/h_t(\theta)| < \infty$. Indeed,

$$E\left|\frac{e_t}{h_t(\theta)}\right| \le \left(E\left|\frac{e_t}{h_t(\theta)}\right|^2\right)^{1/2} \le \left(\frac{1}{\omega}E\frac{e_t^2}{h_t(\theta)}\right)^{1/2} \le \left(\frac{1}{\omega_l}H_1\right)^{1/2} < \infty.$$

Finally, note that

$$\frac{\partial l_t(\theta)}{\partial\beta} = \left(\frac{e_t^2}{h_t(\theta)} - 1\right)\frac{\partial h_t(\theta)}{\partial\beta}\,\frac{1}{h_t(\theta)}.$$

By the triangle inequality, Hölder's inequality, Lemma 5(2), and part (1) of this lemma,

$$\begin{aligned}
E\left|\frac{\partial l_t(\theta)}{\partial\beta}\right| &\le E\left|\frac{(\epsilon_t+g)^2}{h_t(\theta)}h_{\beta t}(\theta)\right| + E|h_{\beta t}(\theta)| \\
&\le E\left|\frac{h_{0t}}{h_t(\theta)}h_{\beta t}(\theta)\right| + E\left|\frac{\epsilon_t}{h_t(\theta)}h_{\beta t}(\theta)\right| + H_\beta\left(1+\frac{g^2}{\omega_l}\right) \\
&\le \left\|\frac{h_{0t}}{h_t(\theta)}\right\|_2 \|h_{\beta t}(\theta)\|_2 + E^{1/2}\left(\frac{\epsilon_t^2}{h_t(\theta)^2}\right)\|h_{\beta t}(\theta)\|_2 + H_\beta\left(1+\frac{g^2}{\omega_l}\right) \\
&\le H_c H_\beta + \left(\frac{H_c}{\omega_l}\right)^{1/2} H_\beta + H_\beta\left(1+\frac{\gamma_u^2}{\omega_l}\right) < \infty.
\end{aligned}$$

We have shown that each derivative is bounded in expectation which completes the proof. ■

Proof of Theorem 1. First, note that Θ_2 is compact. Second, in Lemma 7(2) we showed that $L_n(\theta) \to_p L(\theta)$ pointwise. Third, Lemma 8(2) implies that $L_n(\theta)$ satisfies the weak Lipschitz condition of Andrews [2]. Hence by Theorem 3 of that paper, $L_n(\theta) \to_p L(\theta)$ uniformly in Θ_2 and $L(\theta)$ is continuous in Θ_2. Combined with Lemma 3, we find

$$\sup_{\theta\in\Theta_2}|L_n^*(\theta) - L(\theta)| \le \sup_{\theta\in\Theta_2}|L_n^*(\theta) - L_n(\theta)| + \sup_{\theta\in\Theta_2}|L_n(\theta) - L(\theta)| \to_p 0.$$

Fourth, Lumsdaine [12, Lemma 5 and Theorem 1], generalizing the proof of Weiss [16], showed that the limiting likelihood $L(\theta)$ is uniquely maximized at θ_0, and the proof carries over to the present case. We have established the standard conditions for consistency in nonlinear estimation, for example, Amemiya [1, Theorem 4.1.1]. We conclude that $\hat{\theta}_n \to_p \theta_0$. ■

Proof of Theorem 2. The entire difficulty of the proof of Theorem 1 is the fact that when $\alpha_0 + \beta_0 \ge 1$, $Eh_{0t} = \infty$. In the present case, however, we can establish that

$$\begin{aligned}
Eh_t(\theta) &= \omega + \alpha\sum_{k=0}^{\infty}\beta^k Ee_{t-k-1}^2 \\
&= \omega + \alpha\frac{Eh_{0t}+g^2}{1-\beta} \le \omega_u + \alpha_u\frac{\dfrac{\omega_0}{1-\alpha_0-\beta_0} + (\gamma_u-\gamma_l)^2}{1-\beta_u} = \bar{h} < \infty
\end{aligned}$$

uniformly in $\theta \in \Theta$.

There are few (but key) places in the proof of Theorem 1 where the restricted parameter space Θ_2 is used. The first appears in the proof of Lemma 7(1) where we need a bound on $E(\epsilon_t^2/h_t(\theta))$. In the present case it is simple to see that

$$E\frac{\epsilon_t^2}{h_t(\theta)} \le \omega_l^{-1}E\epsilon_t^2 = \omega_l^{-1}Eh_{0t} \le \omega_l^{-1}\bar{h} < \infty.$$

Thus, $E|l_t(\theta)| < \infty$ and the pointwise convergence of Lemma 7(2) holds. To establish uniform convergence, we need the derivative bounds in the proof of Lemma 8(2), which hold for $E|\partial l_t(\theta)/\partial\gamma|$, $E|\partial l_t(\theta)/\partial\omega|$, and $E|\partial l_t(\theta)/\partial\alpha|$ trivially. It only remains to establish $E|\partial l_t(\theta)/\partial\beta| < \infty$. Indeed,

$$E\left|\frac{\partial l_t(\theta)}{\partial\beta}\right| \le E\left|\frac{h_{0t}}{h_t(\theta)}h_{\beta t}(\theta)\right| + E|h_{\beta t}(\theta)|$$

and

$$E|h_{\beta t}(\theta)| \le \frac{B}{\omega}\sum_{k=0}^{\infty}\beta^k Eh_{t-k-1}^{\epsilon}(\theta) \le \frac{B}{\omega}\sum_{k=0}^{\infty}\beta^k Eh_{t-k-1}^{\epsilon}(\theta) \le \frac{B\bar{h}}{\omega_l(1-\beta_u)} < \infty.$$

For $\beta \ge \beta_0$, from Lemma 1 and Lemma 4(5),

$$E\left|\frac{h_{0t}}{h_t(\theta)}h_{\beta t}(\theta)\right| \le BH_uE|h_{\beta t}(\theta)| < \infty.$$

For $\beta \le \beta_0$, we use the decomposition

$$\begin{aligned}\frac{h_{0t}}{h_t^{\epsilon}(\theta)} &= \frac{\omega_0(1-\beta_0^k)}{h_t^{\epsilon}(\theta)} + \alpha_0\sum_{i=0}^{k-1}\beta_0^i\frac{\epsilon_{t-1-i}^2}{h_t^{\epsilon}(\theta)} + \beta_0^k\frac{h_{0t-1-k}}{h_t^{\epsilon}(\theta)}\\ &\le \frac{\omega_0}{\omega} + \frac{\alpha_0}{\alpha}k\left(\frac{\beta_0}{\beta}\right)^k + \beta_0^k\frac{h_{0t-1-k}}{\omega},\end{aligned}$$

since $\alpha\beta^i\epsilon_{t-1-i}^2 \le h_t^{\epsilon}(\theta)$ and $\sum_{i=0}^{k-1}(\beta_0/\beta)^i \le k(\beta_0/\beta)^k$ as $(\beta_0/\beta) \ge 1$. Thus,

$$\begin{aligned}E\frac{h_{0t}}{h_t(\theta)}h_{\beta t}(\theta) &\le BE\frac{h_{0t}}{h_t^{\epsilon}(\theta)}h_{\beta t}(\theta)\\ &\le B\sum_{k=0}^{\infty}\beta^kE\left(\frac{h_{t-k-1}(\theta)}{h_t(\theta)}\left(\frac{\omega_0}{\omega} + \frac{\alpha_0}{\alpha}k\left(\frac{\beta_0}{\beta}\right)^k + \beta_0^k\frac{h_{0t-1-k}}{\omega}\right)\right)\\ &\le B\left[\frac{\omega_0}{\omega^2}\sum_{k=0}^{\infty}\beta^kEh_{t-k-1}(\theta) + \frac{\alpha_0}{\alpha\omega}\sum_{k=0}^{\infty}k\beta_0^kEh_{t-k-1}(\theta)\right.\\ &\qquad\left. + \frac{1}{\omega}\sum_{k=0}^{\infty}\beta_0^kEh_{0t-1-k}\right]\\ &\le \frac{B\bar{h}}{\omega_l}\left(\frac{\omega_0}{\omega_l(1-\beta_u)} + \frac{\alpha_0\beta_0}{\alpha_l(1-\beta_0)^2} + \frac{1}{(1-\beta_0)}\right),\end{aligned}$$

where the third inequality uses the fact that $\beta^kh_{t-k-1}(\theta)/h_t(\theta) \le 1$. This completes the demonstration that $E|\partial l_t(\theta)/\partial\beta| < \infty$. No other part of the proof of Theorem 1 depends on the restriction of the parameter space to Θ_2, and we conclude that $\tilde{\theta}_n \to_p \theta_0$. ■

Proof of Lemma 9. (1) Lemma 8(2) established the existence of the first moment of the gradient. A review of that proof reveals that the derivatives with respect to ω, α, and γ will have finite second moments if $E(\epsilon_t^2/h_t(\theta))^2 < \infty$. Indeed, using (A.2), Lemma 1, and Lemma 5(2), uniformly in $\theta \in \Theta_4$,

$$E\left(\frac{\epsilon_t^2}{h_t(\theta)}\right)^2 = E\left(\frac{h_{0t}z_t^2}{h_t(\theta)}\right)^2$$

$$= E\left(E(z_t^4 \mid \mathfrak{F}_{t-1})\left(\frac{h_{0t}}{h_t(\theta)}\right)^2\right) \le \mathcal{K}E\left(\frac{h_{0t}}{h_t(\theta)}\right)^2 \le \mathcal{K}(BH_c)^2 < \infty.$$

It remains to show that the derivative with respect to β has a finite second moment. Recall,

$$\frac{\partial l_t(\theta)}{\partial \beta} = \left(\frac{e_t^2}{h_t(\theta)} - 1\right)\frac{\partial h_t(\theta)}{\partial \beta}\,\frac{1}{h_t(\theta)}$$

whose second moment is clearly dominated by

$$E\left(\frac{\epsilon_t^2}{h_t(\theta)}\,h_{\beta t}(\theta)\right)^2 = E\left(E(z_t^4 \mid \mathfrak{F}_{t-1})\,\frac{h_{0t}}{h_t(\theta)}\,h_{\beta t}(\theta)\right)^2 \le \mathcal{K}\left\|\frac{h_{0t}}{h_t(\theta)}\,h_{\beta t}(\theta)\right\|_2^2$$

$$\le \mathcal{K}\left\|\frac{h_{0t}}{h_t(\theta)}\right\|_4^2 \|h_{\beta t}(\theta)\|_4^2 \le \mathcal{K}(BH_c)^2 H_\beta^2 < \infty,$$

where the first inequality uses (A.2) and the final uses Lemma 1, Lemma 5(2), and Lemma 8(1). We have shown that the second moment of each element of the gradient $\nabla l_t(\theta)$ is finite for $\theta \in \Theta_4$, and hence $E|\nabla l_t(\theta)\,\nabla l_t(\theta)'| < \infty$ as stated.

(2) First, observe that $\{\nabla l_t(\theta_0), \mathfrak{F}_t\}$ is a stationary and ergodic martingale difference sequence with finite variance A_0 (as shown in Lemma 9(1)). Thus, by the invariance principle for stationary martingale differences (e.g. Billingsley [3], Theorem 23.1), we have

$$A_0^{-1/2}\,\frac{1}{\sqrt{n}}\sum_{t=1}^{[nr]} \nabla l_t(\theta_0) \Rightarrow W(r).$$

To complete the result, we need to show that

$$\frac{1}{\sqrt{n}}\max_{i \le n}\left|\sum_{t=1}^{i} (\nabla l_t(\theta_0) - \nabla l_t^*(\theta_0))\right| \to_p 0,$$

which will follow if

$$\sum_{t=1}^{\infty} E|\nabla l_t(\theta_0) - \nabla l_t^*(\theta_0)|^\delta < \infty$$

for some $\delta > 0$, which can be shown for $\delta = p$ using methods analogous to the proof of Lemma 3. ∎

Proof of Lemma 10. (1) Differentiation yields

$$\frac{\partial^2 h_t}{\partial \beta^2} = \beta\,\frac{\partial^2 h_{t-1}}{\partial \beta^2} + 2\,\frac{\partial h_{t-1}}{\partial \beta}$$

or

$$h_{\beta\beta t}(\theta) = 2\sum_{k=0}^{\infty} \beta^k \frac{\partial h_{t-k-1}(\theta)}{\partial \beta} \frac{1}{h_t(\theta)} = 2\sum_{k=0}^{\infty} \beta^k h_{\beta t-k-1}(\theta) \frac{h_{t-k-1}(\theta)}{h_t(\theta)}.$$

Using Minkowski's inequality, Hölder's inequality, Lemma 8(1), and similar arguments yield uniformly in $\theta \in \Theta_{2r}$,

$$\begin{aligned} \| h_{\beta\beta t}(\theta) \|_r &\le 2\sum_{k=0}^{\infty} \beta^k \| h_{\beta t-k-1}(\theta) \|_{2r} \left\| \frac{h_{t-k-1}(\theta)}{h_t(\theta)} \right\|_{2r} \\ &\le 2H_\beta \sum_{k=0}^{\infty} \beta^k \left\| \frac{h_{t-k-1}(\theta)}{h_t(\theta)} \right\|_{2r} \le 2H_\beta H_\beta. \end{aligned}$$

(2) Similarly, we find

$$h_{\beta\beta\beta t}(\theta) = 3\sum_{k=0}^{\infty} \beta^k \frac{\partial^2 h_{t-k-1}(\theta)}{\partial \beta^2} \frac{1}{h_t(\theta)} = 3\sum_{k=0}^{\infty} \beta^k h_{\beta\beta t-k-1}(\theta) \frac{h_{t-k-1}(\theta)}{h_t(\theta)},$$

and uniformly in $\theta \in \Theta_{3r}$,

$$\begin{aligned} \| h_{\beta\beta\beta t}(\theta) \|_r &\le 3\sum_{k=0}^{\infty} \beta^k \| h_{\beta\beta t-k-1}(\theta) \|_{3r/2} \left\| \frac{h_{t-k-1}(\theta)}{h_t(\theta)} \right\|_{3r} \\ &\le 6H_\beta^2 \sum_{k=0}^{\infty} \beta^k \left\| \frac{h_{t-k-1}(\theta)}{h_t(\theta)} \right\|_{3r} \le 6H_\beta^3. \end{aligned}$$

■

Proof of Lemma 11. There are 10 distinct second derivatives of the likelihood and 16 distinct third derivatives, each of whose expectations needs to be bounded. All of these derivatives have closed-form solutions. Finding bounds in all cases consists of repeated application of Minkowski's and Hölder's inequalities and the bounds of the previous lemmata. Since this full derivation would take too much space, we provide the full derivation only for the derivatives with respect to β. These derivatives are the most demanding and require the most restrictions.

(1) A straightforward calculation yields

$$\frac{\partial^2 l_t(\theta)}{\partial \beta^2} = \frac{e_t^2}{h_t(\theta)} \left(h_{\beta\beta t}(\theta) - 2h_{\beta t}(\theta)^2 \right) + h_{\beta t}(\theta)^2 - h_{\beta\beta t}(\theta).$$

Thus using Lemma 1, Lemma 8(1), and Lemma 10(1),

$$\begin{aligned} E\left| \frac{\partial^2 l_t(\theta)}{\partial \beta^2} \right| &\le E\left(\frac{e_t^2}{h_t(\theta)} [h_{\beta\beta t}(\theta) + 2h_{\beta t}(\theta)^2] \right) + Eh_{\beta t}(\theta)^2 + Eh_{\beta\beta t}(\theta) \\ &\le E\left(\left(\frac{\epsilon_t^2 + 2\epsilon_t g + g^2}{h_t(\theta)} \right) [h_{\beta\beta t}(\theta) + 2h_{\beta t}(\theta)^2] \right) + H_\beta^2 + 2H_\beta^2 \\ &= E\left(\left(\frac{h_{0t} + g^2}{h_t(\theta)} \right) [h_{\beta\beta t}(\theta) + 2h_{\beta t}(\theta)^2] \right) + 3H_\beta^2 \\ &\le \left(\left\| \frac{h_{0t}}{h_t(\theta)} \right\|_2 + \frac{g^2}{\omega} \right) \left(\| h_{\beta\beta t}(\theta) \|_2 + 2\| h_{\beta t}(\theta) \|_4^2 \right) + 3H_\beta^2 \\ &\le \left(BH_c + \frac{(\gamma_u - \gamma_l)^2}{\omega_l} \right) (2H_\beta^2 + 8H_\beta^4) + 3H_\beta^2 < \infty. \end{aligned}$$

(2) A straightforward calculation yields

$$\frac{\partial^3 l_t(\theta)}{\partial\beta^3} = \frac{e_t^2}{h_t(\theta)}\left(h_{\beta\beta\beta t}(\theta) - 6h_{\beta\beta t}(\theta)h_{\beta t}(\theta) + 2h_{\beta t}(\theta)^3\right)$$
$$- h_{\beta\beta\beta t}(\theta) + 3h_{\beta\beta t}(\theta)h_{\beta t}(\theta) - 3h_{\beta t}(\theta)^3.$$

We find uniformly in $\theta \in \Theta_4$,

$$\begin{aligned}
E\left|\frac{\partial^3 l_t(\theta)}{\partial\beta^3}\right| &\le E\left(\frac{e_t^2}{h_t(\theta)}\left(h_{\beta\beta\beta t}(\theta) + 6h_{\beta\beta t}(\theta)h_{\beta t}(\theta) + 2h_{\beta t}(\theta)^3\right)\right) \\
&\quad + Eh_{\beta\beta\beta t}(\theta) + 3Eh_{\beta\beta t}(\theta)h_{\beta t}(\theta) + 3Eh_{\beta t}(\theta)^3 \\
&\le \left(\left\|\frac{h_{0t}}{h_t(\theta)}\right\|_2 + \frac{g^2}{\omega}\right) \\
&\quad \times \left(\|h_{\beta\beta\beta t}(\theta)\|_2 + 6\|h_{\beta\beta t}(\theta)\|_4\|h_{\beta t}(\theta)\|_4 + 2\|h_{\beta t}(\theta)\|_6^3\right) \\
&\quad + \|h_{\beta\beta\beta t}(\theta)\|_1 + 3\|h_{\beta\beta t}(\theta)\|_2\|h_{\beta t}(\theta)\|_2 + 3\|h_{\beta t}(\theta)\|_3^3 \\
&\le \left(BH_c + \frac{(\gamma_u - \gamma_l)^2}{\omega_l}\right)\left(6H_\beta^3 + 12H_\beta^3 + 2H_\beta^3\right) + \left(6H_\beta^3 + 6H_\beta^3 + 3H_\beta^3\right) \\
&= \left(20\left(BH_c + \frac{(\gamma_u - \gamma_l)^2}{\omega_l}\right) + 15\right)H_\beta^3 < \infty.
\end{aligned}$$

(3) Since for all $\theta \in \Theta_4$, $\nabla^2 l_t(\theta)$ is stationary and ergodic with finite expectation (as shown in part (1) of this lemma), the weak law of large numbers yields

$$-\frac{1}{n}\sum_{t=1}^{n}\nabla^2 l_t(\theta) \to_p B(\theta)$$

for all $\theta \in \Theta_4$. The boundedness of the third derivatives (part (2) of this lemma) implies that $\nabla^2 l_t(\theta)$ is stochastically equicontinuous, so the convergence is uniform in Θ_4 and the limit function $B(\theta)$ is continuous in Θ_4. See, for example, Theorem 3 of Andrews [2]. The desired result therefore follows if

$$\sup_{\theta\in\Theta_4}\frac{1}{n}\sum_{t=1}^{n}\left(\nabla^2 l_t(\theta) - \nabla^2 l_t^*(\theta)\right) \to_p 0,$$

which can be shown using the same methods as in Lemma 3. ■

Proof of Theorem 3. The previous lemmata have established the standard conditions for asymptotic normality in nonlinear estimation (as for example Amemiya [1], Theorem 4.1.3). First, $\hat{\theta}_n$ is consistent (Theorem 1). Second,

$$\frac{1}{\sqrt{n}}\sum_{t=1}^{n}\nabla l_t^*(\theta_0) \to_D N(0,A_0)$$

(Lemma 9). Third,

$$\frac{1}{n}\sum_{t=1}^{n}\nabla^2 l_t^*(\theta) \to_p B(\theta)$$

uniformly in Θ_4 (Lemma 11(3)). Fourth, $B(\theta)$ is continuous in Θ_4 (also Lemma 11 (3)). Fifth and finally, $B_0 > 0$, as shown by the proof of identifiability in Theorem 1.

■

Proof of Lemma 12. (1) Pointwise convergence of $(1/n)\sum_{t=1}^{n} \nabla l_t(\theta)\,\nabla l_t(\theta)'$ to $A(\theta)$ is the result of the weak law of large numbers for stationary sequences, under the moment result of Lemma 9(1). Uniform convergence (and continuity of $A(\theta)$) will obtain if we can establish stochastic equicontinuity, which is a consequence of uniform boundedness of the expectation of the first derivatives. Again we focus on the derivative with respect to β for brevity. Differentiation yields

$$\frac{1}{2}\frac{\partial}{\partial\beta}\left(\frac{\partial l_t(\theta)}{\partial\beta}\right)^2 = \frac{\partial l_t(\theta)}{\partial\beta}\,\frac{\partial^2 l_t(\theta)}{\partial\beta^2}$$

$$= \frac{e_t^4}{h_t^2(\theta)}\,(h_{\beta t}h_{\beta\beta t} - 2h_{\beta t}^3) + \frac{e_t^2}{h_t(\theta)}\,(3h_{\beta t}^3 - 2h_{\beta t}h_{\beta\beta t}) - h_{\beta t}^3 + h_{\beta t}h_{\beta\beta t}.$$

Thus, uniformly in $\theta \in \Theta_6$,

$$\frac{1}{2}E\left|\frac{\partial}{\partial\beta}\left(\frac{\partial l_t(\theta)}{\partial\beta}\right)^2\right| \le 8E\left(\frac{\epsilon_t^4 + g^4}{h_t^2(\theta)}\,|h_{\beta t}h_{\beta\beta t} + 2h_{\beta t}^3|\right)$$

$$+ E\left(\frac{h_{0t} + g^2}{h_t(\theta)}\,|3h_{\beta t}^3 + 2h_{\beta t}h_{\beta\beta t}|\right) + Eh_{\beta t}^3 + Eh_{\beta t}h_{\beta\beta t}$$

$$\le 8\mathcal{K}\left\|\frac{h_{0t}}{h_t(\theta)}\right\|_4^2 (\|h_{\beta t}\|_4\|h_{\beta\beta t}\|_4 + 2\|h_{\beta t}\|_6^3)$$

$$+ \frac{8g^4}{\omega^2}(\|h_{\beta t}\|_2\|h_{\beta\beta t}\|_2 + 2\|h_{\beta t}\|_3^3)$$

$$+ \left\|\frac{h_{0t}}{h_t(\theta)}\right\|_2 (3\|h_{\beta t}\|_6^3 + 2\|h_{\beta t}\|_4\|h_{\beta\beta t}\|_4)$$

$$+ \frac{g^2}{\omega}(3\|h_{\beta t}\|_3^3 + 2\|h_{\beta t}\|_2\|h_{\beta\beta t}\|_2) + \|h_{\beta t}\|_3^3 + \|h_{\beta t}\|_2\|h_{\beta\beta t}\|_2$$

$$\le \left(32\mathcal{K}(BH_c)^2 + \frac{32(\gamma_u - \gamma_l)^4}{\omega_l^2} + 7BH_c + \frac{7(\gamma_u - \gamma_l)^2}{\omega_l} + 3\right)$$

$$\times H_\beta^3 < \infty,$$

where the first inequality uses c_r inequality.

This bound implies that the sequence $(1/n)\sum_{t=1}^{n} \nabla l_t(\theta)\,\nabla l_t(\theta)' - A(\theta)$ is stochastically equicontinuous in $\theta \in \Theta_6$ as desired. The desired result therefore follows if

$$\sup_{\theta\in\Theta_6}\frac{1}{n}\left|\sum_{t=1}^{n}(\nabla l_t(\theta)\,\nabla l_t(\theta)' - \nabla l_t^*(\theta)\,\nabla l_t^*(\theta)')\right| \to_p 0,$$

which can be shown using the same methods as in Lemma 3.

(2) The result given in part (1) of this lemma and the consistency of $\hat\theta_n$ for θ_0 implies that $\hat A_n = \hat A_n(\hat\theta_n) \to_p A(\theta_0) = A_0$. Similarly, Lemma 11(3) yields $\hat B_n = \hat B_n(\hat\theta_n) \to_p B(\theta_0) = B_0$. The continuous mapping theorem completes the proof. ■