Asymptotics of the QMLE for a Class of ARCH(q) Models

Author(s): Dennis Kristensen and Anders Rahbek

Source: *Econometric Theory*, Oct., 2005, Vol. 21, No. 5 (Oct., 2005), pp. 946-961

Published by: Cambridge University Press

Stable URL: https://www.jstor.org/stable/3533519

Econometric Theory, **21**, 2005, 946–961. Printed in the United States of America.
DOI: 10.1017/S0266466605050474

ASYMPTOTICS OF THE QMLE FOR A CLASS OF ARCH(q) MODELS

DENNIS KRISTENSEN
University of Wisconsin

ANDERS RAHBEK
University of Copenhagen

Strong consistency and asymptotic normality are established for the quasi-maximum likelihood estimator for a class of ARCH(q) models. The conditions are that the ARCH process is geometrically ergodic with a moment of arbitrarily small order. Furthermore for consistency, we assume that the second-order moment exists for the nondegenerate rescaled errors and, similarly, that the fourth-order moment exists for asymptotic normality to hold. Contrary to existing literature on (G)ARCH models the parameter space is not assumed to be compact; we only impose a lower bound for the constant term in our parameterization of the conditional variance. It is demonstrated that the general conditions are satisfied for a range of specific models.

1. INTRODUCTION

Since the introduction of the autoregressive conditional heteroskedastic (ARCH) model by Engle (1982), the basic model has been extended and generalized in various ways; see, e.g., Bera and Higgins (1993). One of the most popular estimators for this class of models is the quasi-maximum likelihood estimator (QMLE) where the Gaussian likelihood is used for the true but possibly unknown likelihood. We derive the asymptotic properties of the QMLE for a class of parametric univariate ARCH(q) models, which includes many of the models in the literature; for example, the linear ARCH model (Engle, 1982) and various asymmetric ARCH models (Guegan and Diebolt, 1994; Glosten, Jegannathan, and Runke, 1993; Gourieroux and Monfort, 1992). Except for the linear ARCH model, none of these models has been treated explicitly before.

We are grateful to the editor and the referees for their very helpful and detailed suggestions, which, we believe, improved the paper substantially. We thank Søren T. Jensen for stimulating discussions and Jonathan Dennis for helpful research assistance. Rahbek acknowledges continuing financial support from the Danish Social Sciences Research Council. Kristensen received funding from the Danish Research Agency and the Financial Markets Group, LSE, during this research. Address correspondence to Anders Rahbek, Department of Applied Mathematics and Statistics, University of Copenhagen, Universitetsparken 5, DK-2100 Copenhagen Ø, Denmark; e-mail: rahbek@stat.ku.dk.

We show that the QMLE for this class is consistent and asymptotically normally distributed under weak conditions: we require the existence of an arbitrarily small (log) order moment of the ARCH process itself; only the second-order moment of the rescaled errors is required for consistency, whereas the fourth moment is needed for asymptotic normality. Furthermore, in contrast to most of the existing work, the parameter space is not assumed to be compact. Also, we allow for no ARCH in the proof of consistency; to establish asymptotic normality we assume that ARCH is present, however, because the proof is based on a standard Taylor expansion.

As a further condition we assume that the ARCH process is geometrically ergodic, which is verified for the previously mentioned models under weak conditions. See, e.g., An, Chen, and Huang (1997); Basrak, Davis, and Mikosch (2002); Carrasco and Chen (2002); Lu and Jiang (2001); and Masry and Tjøstheim (1995) for further results on geometric ergodicity in ARCH models. In general, a Markov process $\{X_t\}$, $X_t \in \mathbb{R}^k$ is geometrically ergodic with associated invariant probability measure π, if, for some $\rho < 1$ and all $x \in \mathbb{R}^k$, $\|Q^n(\cdot|x) - \pi(\cdot)\| \leq M_x \rho^n$ for some $M_x > 0$. Here $\|\cdot\|$ denotes the total variation norm and Q^n the n-step transition probability (for details, see Meyn and Tweedie, 1993). In particular, if $X_1 \sim \pi$ then $\{X_t\}$ is stationary and ergodic. Geometric ergodicity implies that the stationary version is β-mixing with exponentially decaying mixing coefficients (see Davydov, 1973; Doukhan, 1994). Importantly, the strong law of large numbers (LLN) holds for any given initial distribution of the geometrically ergodic process (cf. Meyn and Tweedie, 1993, Thm. 17.0.1), which is used repeatedly in the proofs.

Finally, it should be noted that we consider a global QMLE, in contrast to much of the existing literature where local properties are derived. Our consistency proof is based on Huber (1967) and Pfanzagl (1969). Jeantheau (1998) applies Pfanzagl (1969) also to give conditions for consistency of the QMLE in a broad class of (G)ARCH models but assumes that the stationary version of the process is observed and that the parameter space is compact, contrary to here.

Although the QMLE is often used in empirical studies, its asymptotic properties have mainly been derived for the linear (G)ARCH model. Weiss (1986) proved consistency and asymptotic normality for the QMLE in the linear ARCH(q) model requiring the existence of the fourth-order moment of the ARCH process, which contradicts empirical work. The moment conditions were weakened in Lee and Hansen (1994) and Lumsdaine (1996); both of these studies considered the linear GARCH(1,1) model. Lee and Hansen (1994) require a stationary solution to the model and the existence of fourth-order moment of the rescaled errors, whereas the parameter space is assumed to be compact and restricted further. Lumsdaine (1996) considers GARCH processes with second-order moments and the I-GARCH process and assumes that the rescaled errors have 32nd-order moments in addition to restrictions on the parameter space. In Berkes, Horvarth, and Kokoszka (2003), the results are extended to hold for the linear stationary GARCH(p,q) model under weaker conditions on the rescaled errors. Jensen and Rahbek (2004a, 2004b) establish that the asymptotic results hold for the entire parameter region, including the nontationary explosive case, for ARCH(1) and GARCH(1,1) models. Finally, Ling and Li (1998) investigate asymptotics for unit-root ARMA-GARCH models. All of the aforementioned assumes that the parameter space is compact or addresses local consistency.

In the following section we present the model and the assumptions applied. In Section 3 we state the consistency result, and the asymptotic distribution of the QMLE is in Section 4. In Section 5, we go through various popular parametrizations, and show that they satisfy the assumptions made in the general model. We conclude in Section 6. All proofs together with technical lemmas are found in Appendixes A and B, respectively. Appendix C contains two propositions.

Throughout the paper we will use the following notation. For any function $f: \mathbb{R}^m \times \Theta \mapsto \mathbb{R}^n$, we shall write the partial derivative w.r.t. θ as $Df(x,\theta) = (D_j(f_i x,\theta))_{i,j}$ where $D_j f_i = (\partial f_i(x,\theta)/\partial\theta_j)$; the second-order partial derivatives are written as $D^2 f(x,\theta) = (D_{jk} f_i(x,\theta))_{i,j,k}$ where $D_{jk} f_i = (\partial^2 f_i(x,\theta)/\partial\theta_j\partial\theta_k)$. Also, C and C_i, $i = 1,2,\ldots$, will denote generic constants. For any process X_t, use $\{X_t\}$ to denote the infinite sequence $X_1, X_2, \ldots$.

2. THE ARCH MODEL

Assume that we have observed $(y_t, x_{t-1})_{t=1,\ldots,n}$ with $y_t \in \mathbb{R}$ and $x_{t-1} \in \mathbb{R}^m$ given by the ARCH model,

$$y_t = h_t^{1/2} z_t, \tag{1}$$

$$h_t = \omega + \alpha' x_{t-1}, \tag{2}$$

with parameter $\theta = (\omega, \alpha) \in \Theta \subseteq \mathbb{R}^{1+m}$ and where x_0 is fixed.[1] Conditions on the explanatory variables x_{t-1}, the independent and identically distributed (i.i.d.) innovations z_t, and the parameter space Θ are specified subsequently.

Some of the parametrizations that fall within our framework are given in Table 1: linear ARCH (Engle, 1982), β-ARCH (Guegan and Diebolt, 1994) for known $\beta > 0$, μ-ARCH (Engle and Bollerslev, 1986) for known $\mu > 0$, threshold or GJR-ARCH (Glosten et al., 1993; Li and Li, 1996; Zakoian, 1994), SD-ARCH (Schwert, 1990), and TARCH (Gourieroux and Monfort, 1992).

For the asymptotic analysis, we assume that for θ at the true value θ_0 there exists a stationary solution to equations (1)–(2) and denote this $\{(y_t^*, x_{t-1}^*)\}$. However, the process $\{(y_t, x_{t-1})\}$ may be nonstationary as x_0, and hence y_1, need not be initiated from the stationary distribution. To emphasize this, in the following discussion let $h_{0t} = \omega_0 + \alpha_0' x_{t-1}$ and $h_{0t}^* = \omega_0 + \alpha_0' x_{t-1}^*$ denote the true conditional variance process corresponding to $\{(y_t, x_{t-1})\}$ and $\{(y_t^*, x_{t-1}^*)\}$, respectively. Correspondingly, write $h_t = \omega + \alpha' x_{t-1}$ and h_t^* for any $\theta \neq \theta_0$.

The following conditions are assumed to hold.

TABLE 1. Examples of parametrizations

Name	$\alpha' x_{t-1}$
Linear ARCH	$\sum_{i=1}^{q} \alpha_i y_{t-i}^2$
β-ARCH	$\sum_{i=1}^{q} \{\alpha_{1i} 1_{(y_{t-i}<0)} \lvert y_{t-i}\rvert^{\beta} + \alpha_{2i} 1_{(y_{t-i}\geq 0)} \lvert y_{t-i}\rvert^{\beta}\}$
μ-ARCH	$\sum_{i=1}^{q} \alpha_i \lvert y_{t-i}\rvert^{\mu}$
GJR-ARCH	$\sum_{i=1}^{q} \{\alpha_{1i} 1_{(y_{t-i}<0)} y_{t-i}^2 + \alpha_{2i} 1_{(y_{t-i}\geq 0)} y_{t-i}^2\}$
SD-ARCH	$\sum_{i=1}^{q} \alpha_{1i} \lvert y_{t-i}\rvert + \sum_{i=1}^{q} \sum_{j=i}^{q} \alpha_{2ij} \lvert y_{t-i} y_{t-j}\rvert$
TARCH	$\sum_{i=1}^{q} \sum_{j=1}^{k} \alpha_{ij} 1_{A_j}(y_{t-i})$

Note: $1_A(\cdot)$ denotes the indicator function for the set A.

C.1. The innovations $\{z_t\}$ are i.i.d.(0,1) with $E[|\log(z_t^2)|] < \infty$.

C.2. The parameter space is given by $\Theta = \{\theta = (\omega, \alpha) \in \mathbb{R}^{m+1} | \underline{\omega} \leq \omega, 0 \leq \alpha_i\}$ for some fixed $\underline{\omega} > 0$.

C.3. For all $t \geq 1$, (i) $x_{i,t-1} \geq 0$, $1 \leq i \leq m$, (ii) x_{t-1} is measurable w.r.t. $\mathcal{F}_{t-1} = \mathcal{F}(y_{t-1}, \dots, y_{t-q})$ for some finite $q \geq 1$, and (iii) $P(\beta' x_{t-1}^* \neq c) > 0$ for any $\beta \in \mathbb{R}^m \backslash \{0\}$ and $c \in \mathbb{R}$.

C.4. (i) $\{(y_t, x_{t-1})\}$ is geometrically ergodic such that for the true parameter $\theta_0 \in \Theta$ a stationary solution $\{(y_t^*, x_{t-1}^*)\}$ exists with (ii) $E[|\log(x_{i,t}^*)|] < \infty$, $1 \leq i \leq m$.

The assumptions made in C.1 about the rescaled errors are fairly standard, where we note that the i.i.d. assumption is needed to obtain the Markov property of $\{(y_t, x_{t-1})\}$. Lee and Hansen (1994) assumed that $\{z_t\}$ was a stationary martingale difference sequence. It is possible here to introduce limited dependence in the $\{z_t\}$ process without removing the Markov property, e.g., by assuming that $\{(z_t, \dots, z_{t-p})\}$ is a geometrically ergodic Markov chain for some finite $p \geq 0$ and satisfying $E[z_t | \mathcal{F}_{t-1}] = 0$ and $E[z_t^2 | \mathcal{F}_{t-1}] = 1$ where $\mathcal{F}_{t-1}$ includes the lagged z_t's too. For simplicity, we maintain the i.i.d. assumption however.

Condition C.2 defines the noncompact parameter space Θ for the statistical analysis; i.e., the quasi-likelihood function defined subsequently is maximized over $\theta \in \Theta$. Condition C.2 implies—together with C.3(i) and (ii)—that $h_t > 0$ for all $\theta \in \Theta$.

Note that the only real restriction on the parameter space Θ is the requirement that the constant term, ω, is bounded from below by some $\underline{\omega} > 0$. It is used when deriving a uniform lower bound for the quasi-likelihood function in the proof of Lemma 4 and appears difficult to avoid.

For the asymptotics, or the probability analysis, condition C.4(i) of geometric ergodicity restricts the possible values of the true value θ_0 in Θ. As noted geometric ergodicity implies that the strong LLN holds for $\{(y_t, x_{t-1})\}$ irrespective of initial values. Explicit conditions for which C.4 holds in some leading examples are given in Section 5, Theorem 3 and Corollary 1. The assumption of geometric ergodicity can be exchanged with the weaker condition that there exists a version of $\{(y_t, x_{t-1})\}$ that is stationary and ergodic. But then one needs to either (i) assume that it is the stationary version that has been observed (see Jeantheau, 1998) or (ii) show that the observed quasi-likelihood function converges uniformly toward the stationary version of it (see Lee and Hansen, 1994).

Condition C.3(iii) is a nondegeneracy condition that is used in the proof of identification of θ_0. A sufficient, but stronger, condition for C.3(iii) would be that $E[x^*_{t-1} x^{*\prime}_{t-1}]$ is nonsingular, which however requires that the moment exists. For example, in the linear ARCH model we would have to require that $E[y_t^{*4}] < \infty$, contradicting empirical findings.

Finally note that it can be shown that a simple condition for the moment in C.4(ii) to exist is $E[\|x_t^*\|^p] < \infty$ for some $p > 0$. And if one assumes that Θ is compact, the log-moment condition in C.1 can be removed and C.4(ii) weakened to $E[|\log(h^*_{0t})|] < \infty$; cf. the proof of Lemma 4.

3. CONSISTENCY OF THE QMLE

Our first main result is that the QMLE $\hat{\theta}_n$ is strongly consistent, where

$$\hat{\theta}_n = \arg\inf_{\theta\in\Theta} L_n(\theta). \tag{3}$$

Here

$$L_n(\theta) = \frac{1}{n}\sum_{t=1}^{n} l(y_t|x_{t-1};\theta), \qquad l(y_t|x_{t-1};\theta) = \log(h_t) + \frac{y_t^2}{h_t}. \tag{4}$$

We furthermore define the moment function

$$L(\theta) = E[l(y_t^*|x_{t-1}^*;\theta)],$$

which takes values on the extended real line $\mathbb{R} \cup \{-\infty, +\infty\}$. This is well defined by Lemma 1. To allow for noncompact Θ when proving consistency, we use the strategy of Huber (1967). The idea is to find a compact set $\mathcal{C}_0 \subseteq \Theta$, as defined in (B.2) in Appendix B, for which $\theta_0 \in \mathcal{C}_0$ and such that $\inf_{\theta\in\Theta\setminus\mathcal{C}_0} L_n(\theta) \geq L(\theta_0) + \varepsilon$ a.s. as $n \to \infty$ for some $\varepsilon > 0$. It then follows that $\hat{\theta}_n \in \mathcal{C}_0$ a.s. as $n \to \infty$, and we can apply a consistency result of Pfanzagl (1969) on the compact set $\mathcal{C}_0$. This involves showing that $L(\theta)$ is well defined, that θ_0 is identified by $L(\theta)$, and that $\theta \mapsto L(\theta)$ is continuous.

THEOREM 1. *Under C.1–C.4, the QMLE defined by (3) satisfies* $\hat{\theta}_n \xrightarrow{a.s.} \theta_0$.

4. ASYMPTOTIC NORMALITY OF THE QMLE

In this section we show that the QMLE is asymptotically Gaussian. To prove this result, we impose the following assumptions in addition to C.1–C.4.

N.1. For $\theta_0 = (\omega_0, \alpha_0)$, $\omega_0 > \underline{\omega}$ and $\alpha_{0,i} > 0$ for $i = 1,\dots,m$.

N.2. $\kappa_4 = E[(z_t^2 - 1)^2] \in (0,\infty)$.

Assumption N.1, or equivalently $\theta_0 \in \text{int}\Theta$, is needed for a Taylor expansion of the quasi-likelihood function around θ_0. The case of no ARCH, $\alpha_0 = 0$, is not straightforward (see, e.g., Andrews, 1999) and will be left for future work. The second condition, N.2, ensures that the asymptotic variance of the QMLE exists.

The proof of asymptotic normality is standard (see, e.g., Basawa, Feigin, and Heyde, 1976), based on a Taylor expansion of the score,

$$S_n(\theta) = \frac{1}{n}\sum_{i=1}^{n} s(y_t|x_{t-1};\theta), \qquad s(y_t|x_{t-1};\theta) = Dl(y_t|x_{t-1};\theta), \tag{5}$$

around θ_0. In particular, we show in Lemma 9 that $\sqrt{n}S_n(\theta_0) \xrightarrow{\mathcal{D}} N(0,V(\theta_0))$ where

$$V(\theta_0) = E[s(y_t^*|x_{t-1}^*;\theta_0)s(y_t^*|x_{t-1}^*;\theta_0)'] = \kappa_4 E\left[\frac{Dh_{0t}^* Dh_{0t}^{*\prime}}{h_{0t}^{*2}}\right] \tag{6}$$

and furthermore in Lemma 8 that uniformly over θ in a compact set containing θ_0,

$$H_n(\theta) = DS_n(\theta) \xrightarrow{\text{a.s.}} H(\theta) = E\left[\frac{Dh_t^* Dh_t^{*\prime}}{h_t^{*2}}\right]. \tag{7}$$

THEOREM 2. *Under C.1–C.4 and N.1 and N.2,*

$$\sqrt{n}(\hat{\theta}_n - \theta_0) \xrightarrow{\mathcal{D}} N(0, H^{-1}(\theta_0)V(\theta_0)H^{-1}(\theta_0)),$$

where $V(\theta_0)$ and $H(\theta_0)$ are given in (6) and (7), both are positive definite. Furthermore,

$$\hat{V}_n = \frac{1}{n}\sum_{t=1}^{n} s(y_t|x_{t-1};\hat{\theta}_n)s(y_t|x_{t-1};\hat{\theta}_n)', \tag{8}$$

$$\hat{H}_n = \frac{1}{n}\sum_{t=1}^{n} Ds(y_t|x_{t-1};\hat{\theta}_n) \tag{9}$$

are strongly consistent estimators of $V(\theta_0)$ and $H(\theta_0)$, respectively.

Remark 1. In the case where $z_t \sim N(0,1)$, the result collapses to the standard maximum likelihood estimator (MLE) result with $\kappa_4 = 2$ and $H(\theta_0)$ being the information matrix multiplied by 2.

5. EXAMPLES

In this section we consider the models shown in Table 1 and demonstrate that under suitable conditions they satisfy C.2–C.4, which are the only assumptions made on the actual parametrization. The remaining assumptions concern the error structure (C.1 and N.2) and the location of θ_0 in the parameter space (C.2 and N.1).

To do so we consider a specific subclass of models of (1)–(2) given by

$$\alpha' x_{t-1} = \sum_{i=1}^{q} \alpha_i' g(y_{t-i}) = \sum_{i=1}^{q} \sum_{j=1}^{p} \alpha_{ij} g_j(y_{t-i}), \tag{10}$$

which contains the models in Table 1 as special cases, except the SD-ARCH because of the cross-product terms. This is treated separately as part of Corollary 1, which follows. In (10),

$$x_{t-1} = [g(y_{t-1})', \ldots, g(y_{t-q})']', \qquad g(y) = (g_1(y), \ldots, g_p(y))', \tag{11}$$

such that the dimension of x_{t-1} is $m = qp$. For this class, we address the two questions, identification, C.3, and geometric ergodicity, C.4, in turn. Specifically, we make the following assumptions about g.

G.1. $g_j : \mathbb{R} \mapsto \mathbb{R}_+ \cup \{0\}$ is measurable, $j = 1, \ldots, p$.

G.2. For any $h > 0$, $\beta \in \mathbb{R}^p \backslash \{0\}$ and $c \in \mathbb{R}$: $P(\beta' g(h z_t) \neq c) > 0$.

G.3. There exist constants $a_0^{(i)} > 0$, $a_1^{(i)} \geq 0$, $i = 1, \ldots, q$, with $\sum_{i=1}^{q} a_1^{(i)} < 1$, such that for $\alpha = \alpha_0$, $\sum_{j=1}^{p} \alpha_{0,ij} g_j(y) = a_0^{(i)} + a_1^{(i)} y^2 + o(y^2)$ as $|y| \to \infty$.

Note that by Lemma 10, G.1 and G.2 imply C.3. The proof exploits that as $y_{t-i} = h_{t-i}^{1/2} z_{t-i}$, G.2 can be seen as a conditional (on $h_{t-i}^{1/2} = h$) version of C.3(iii) when x_{t-1} takes the form (11). The condition G.3 implies that C.4 holds by using the so-called drift criterion from Markov chain theory.

THEOREM 3. *Assume (i) that C.1, C.2, N.1, and N.2 hold with $z_t \sim f_z \cdot \lambda$, where $f_z > 0$ is a lower semicontinuous density w.r.t. the Lebesgue measure λ. Assume further (ii) that $\alpha' x_{t-1}$ is given by (10) and that G.1–G.3 hold. Then the QMLE is strongly consistent and asymptotically Gaussian.*

As a corollary we state the results specifically for the models in Table 1.

COROLLARY 1. *Assume that the assumptions of Theorem 3 hold. Assume further at the true parameter value for the linear* ARCH(q), $\Sigma_{i=1}^{q}\alpha_{0,i} < 1$*; for the* β-ARCH, $\beta \in (0,2)$*; for the* μ-ARCH, $\mu \in (0,2)$*; for the* GJR-ARCH, $\Sigma_{i=1}^{q}\{\alpha_{0,1i} + \alpha_{0,2i}\} < 1$*; for the* SD-ARCH, $\Sigma_{i=1}^{q}\alpha_{0,2ii} + 2\Sigma_{i=1}^{q}\Sigma_{j=i+1}^{q}\alpha_{0,2ij} < 1$*; and for the* TARCH, $\lambda(A_i) > 0$, $\lambda(A_i \cap A_j) = 0$, $i \neq j$, *and* $\lambda(\mathbb{R}\setminus \cap A_i) > 0$. *Then the* QMLE *for each model is strongly consistent and asymptotically Gaussian.*

Remark 2. Apart from the TARCH case, the last conditions in Corollary 1 are only needed to establish geometric ergodicity. So if weaker conditions for this property can be found, then under these conditions the results remain valid. For example, in the linear ARCH(1) model, we may weaken the preceding condition, $\alpha_0 = \alpha_{0,1} < 1$, to $\log(\alpha_0) < -E[\log(z_t^2)]$ (see Lemma 11). Note that this condition is necessary and sufficient for stationarity and ergodicity (see Nelson, 1990).

Remark 3. The condition given for the TARCH model is used to establish identification. It might be possible to find weaker conditions for this property to hold.

6. CONCLUSION

We have shown consistency and asymptotic normality for the QMLE in a general class of ARCH models under relatively weak conditions. Several extensions are of interest: GARCH models other than the linear one, more general ARCH(q) models that are not linear in the parameters, and multivariate models. The latter has already been investigated by Comte and Lieberman (2003) and Ling and McAleer (2003), but it could be interesting to consider other parametrizations than the ones they work with and perhaps weaken the restrictions they impose.

NOTE

1. The following results can be extended to allow for x_0 being random as long as its distribution does not depend on θ.

REFERENCES

An, H., M. Chen, & F. Huang (1997) The geometric ergodicity and existence of moments for a class of nonlinear time series model. *Statistics and Probability Letters* 31, 213–224.

Andrews, D. (1999) Estimation when a parameter is on a boundary. *Econometrica* 67, 1341–1384.

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.

Basrak, B., R. Davis, & T. Mikosch (2002) Regular variation of GARCH processes. *Stochastic Processes and Their Applications* 99, 95–115.

Bera, A.K. & M. Higgins (1993) ARCH models: Properties and estimation. *Journal of Economic Surveys* 7, 305–366.

Berkes, I., L. Horvath, & P. Kokoszka (2003) GARCH processes: Structure and estimation. *Bernoulli* 9(2), 201–227.

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

Carrasco, M. & X. Chen (2002) Mixing and moment properties of various GARCH and stochastic volatility models. *Econometric Theory* 18, 17–39.

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

Davydov, Y. (1973) Mixing conditions for Markov chains. *Theory of Probability and Its Applications* 18, 312–328.

Doukhan, P. (1994) *Mixing: Properties and Examples.* Lecture Notes in Statistics 85. Springer-Verlag.

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

Engle, R.F. & T. Bollerslev (1986) Modelling the persistence of conditional variances. *Econometric Review* 5, 1–50.

Glosten, L., R. Jegannathan, & D. Runke (1993) Relationship between the expected value and the volatility of the nominal excess return on stocks. *Journal of Finance* 48, 1779–1801.

Gourieroux, C. & A. Monfort (1992) Qualitative threshold ARCH models. *Journal of Econometrics* 52, 159–200.

Guegan, D. & J. Diebolt (1994) Probabilistic properties of the β-ARCH model. *Statistica Sinica* 4, 71–87.

Huber, P.J. (1967) The behavior of maximum likelihood estimates under nonstandard conditions. In L.M. Cam, J. Neyman, & E.L. Scott (eds.), *Proceedings of the Fifth Berkeley Symposium on Mathematical Statistics and Probability*, vol. 1, pp. 221–233. University of California Press.

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

Jensen, S.T. & A. Rahbek (2004a) Asymptotic normality of the QMLE estimator of ARCH in the nonstationary case. *Econometrica* 72, 641–646.

Jensen, S.T. & A. Rahbek (2004b) Asymptotic inference for nonstationary GARCH. *Econometric Theory* 20, 1203–1226.

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

Li, C.W. & W.K. Li (1996) On a double-threshold autoregressive heteroscedastic time series model. *Journal of Applied Econometrics* 11, 253–274.

Ling, S. & W.K. Li (1998) Limiting distributions of maximum likelihood estimators for unstable autoregressive moving-average time series with general autoregressive heteroscedastic errors. *Annals of Statistics* 26, 84–125.

Ling, S. & M. McAleer (2003) Asymptotic theory for a new vector ARMA-GARCH model. *Econometric Theory* 19, 278–308.

Lu, Z. (1998) Two general theorems on geometric ergodicity of a general ARCH type model with applications to some typical models. In J. Gu, G. Fan, S. Wang, & B. Wei (eds.), *Advances in Operations Research and Systems Engineering.* Global-Link Informatics Ltd., Hong Kong.

Lu, Z. & Z. Jiang (2001) L_1 geometric ergodicity of a multivariate nonlinear AR model with an ARCH term. *Statistics and Probability Letters* 51, 121–130.

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.

Masry, E. & D. Tjøstheim (1995) Nonparametric estimation and identification of nonlinear ARCH time series. *Econometric Theory* 11, 258–289.

Meyn, S.P. & R.L. Tweedie (1993) *Markov Chains and Stochastic Stability.* Communications and Control Engineering Series. Springer-Verlag.

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

Pfanzagl, J. (1969) On the measurability and consistency of minimum contrast estimates. *Metrika* 14, 107–131.
Schwert, G.W. (1990) Stock volatility and the crash of 87. *Review of Financial Studies* 3, 77–102.
Tauchen, G.E. (1985) Diagnostic testing and evaluation of maximum likelihood models. *Journal of Econometrics* 30, 415–443.
Weiss, A. (1986) Asymptotic theory for ARCH models. *Econometric Theory* 2, 107–131.
Zakoian, J.-M. (1994) Threshold heteroscedastic models. *Journal of Economic Dynamics and Control* 18, 931–955.

APPENDIX A: Proofs

Proof of Theorem 1. We first establish that $\lim_{n\to\infty}\hat{\theta}_n \in \mathcal{C}_0$ a.s. where $\mathcal{C}_0$ is the compact set given in (B.2) in Lemma 4 with $\theta_0 \in \mathcal{C}_0$. Observe first that by the strong LLN,

$$\lim_{n\to\infty} L_n(\hat{\theta}_n) = \lim_{n\to\infty}\left[\inf_{\theta\in\Theta} L_n(\theta)\right] \geq \lim_{n\to\infty} L_n(\theta_0) = L(\theta_0) \quad \text{a.s.,} \tag{A.1}$$

where $L(\theta_0) < \infty$ by Lemma 3. Next, applying Lemma 4 and the definition of $\underline{l}(y_t|x_{t-1})$ therein, together with the strong LLN gives

$$\lim_{n\to\infty}\left[\inf_{\theta\in\Theta\setminus\mathcal{C}_0} L_n(\theta)\right] \geq \lim_{n\to\infty}\left[\frac{1}{n}\sum_{t=1}^{n}\underline{l}(y_t|x_{t-1})\right] = L(\theta_0) + \varepsilon \quad \text{a.s.} \tag{A.2}$$

Combining (A.1) and (A.2), we obtain that $\lim_{n\to\infty}\hat{\theta}_n \in \mathcal{C}_0$ a.s. Hence we can restrict attention to $\mathcal{C}_0$ and redefine $\hat{\theta}_n$ as $\hat{\theta}_n = \arg\min_{\theta\in\mathcal{C}_0} L_n(\theta)$. The set $\mathcal{C}_0$ is compact, and we can apply Proposition 2: $l(y_t|x_{t-1};\theta) = \log(h_t) + y_t^2/h_t$ is continuous in θ for all (y_t, x_{t-1}); by Lemma 1, $L(\theta) > -\infty$, so the moment is well defined; identification of θ_0 holds by Lemma 3. Finally, for any compact subset $\mathcal{C} \subseteq \mathcal{C}_0$, $E[\inf_{\theta\in\mathcal{C}} l(y_t^*|x_{t-1}^*;\theta)] > -\infty$ by Lemma 1. We can therefore conclude that $P(\lim_{n\to\infty}\hat{\theta}_n = \theta_0) = 1$ as desired. ■

Proof of Theorem 2. The result is obtained by the following standard Taylor expansion of $S_n(\hat{\theta}_n)$ in (5). Note first that by Lemma 5, $S_n(\theta)$ is well defined for all $\theta \in \Theta$. By N.1, there exists a ball $B(\theta_0,\delta) \subseteq \Theta$. We then choose the compact $\mathcal{N}_0$ such that $B(\theta_0,\delta/2) \subseteq \mathcal{N}_0 \subseteq B(\theta_0,\delta)$,

$$\mathcal{N}_0 = [\omega^l, \omega^u] \times [\alpha_1^l, \alpha_1^u] \times \cdots \times [\alpha_m^l, \alpha_m^u], \tag{A.3}$$

where $\underline{\omega} \leq \omega^l < \omega^u$ and $0 < \alpha_i^l < \alpha_i^u$, $i = 1,\dots,m$. As $n \to \infty$, $\hat{\theta}_n \in \operatorname{int}(\mathcal{N}_0)$ a.s. by Theorem 1, such that $S_n(\hat{\theta}_n) = 0$ by the definition of $\hat{\theta}_n$ and hence,

$$0 = S_n(\hat{\theta}_n) = S_n(\theta_0) + H_n(\bar{\theta}_n)(\hat{\theta}_n - \theta_0),$$

where $H_n(\theta)$ is given in (7) and $\bar{\theta}_n = (\bar{\theta}_{n,1},\dots,\bar{\theta}_{n,m+1})' \in \mathcal{N}_0$ because $\bar{\theta}_{n,i} \in [\hat{\theta}_{n,i}, \theta_{0,i}]$. Lemmas 8 and 9 then imply

$$\sqrt{n}(\hat{\theta}_n - \theta_0) = H_n^{-1}(\bar{\theta}_n)\sqrt{n}S_n(\theta_0) \xrightarrow{\mathcal{D}} N(0, H^{-1}(\theta_0)V(\theta_0)H^{-1}(\theta_0)).$$

To prove that $\hat{V}_n \xrightarrow{\text{a.s.}} V(\theta_0)$, define $V_n(\theta) = 1/n\sum_{t=1}^n s(y_t|x_{t-1};\theta)s(y_t|x_{t-1};\theta)'$ and use the following inequality:

$$\|\hat{V}_n - V(\theta_0)\| \le \|\hat{V}_n - V(\hat{\theta}_n)\| + \|V(\hat{\theta}_n) - V(\theta_0)\|$$
$$\le \sup_{\theta\in\mathcal{N}_0} \|V_n(\theta) - V(\theta)\| + \|V(\hat{\theta}_n) - V(\theta_0)\|.$$

By Lemma 7(ii) $\sup_{\theta\in\mathcal{N}_0}\|V_n(\theta) - V(\theta)\| \xrightarrow{\text{a.s.}} 0$ with $V(\theta)$ continuous. Thus, $\hat{\theta}_n \xrightarrow{\text{a.s.}} \theta_0$ implies $\hat{V}(\hat{\theta}_n) \xrightarrow{\text{a.s.}} V(\theta_0)$. Similarly, $\hat{H}_n \xrightarrow{\text{a.s.}} H(\theta_0)$, by Lemma 8. ■

Proof of Theorem 3. The result follows by showing that C.3 and C.4 hold. C.3(i) and (ii) hold by G.1 and G.2; C.3(iii) holds by Lemma 10; and finally, C.4 holds by Lu (1998, Thm. 3.1 with $s(y) = |y|$). ■

Proof of Corollary 1. The result follows for all models but the SD-ARCH by establishing G.1–G.3. Assumption G.1 trivially holds, and G.3 is simple to verify for each model in question under the assumptions of the corollary. Next, turn to G.2.

For the linear and μ-ARCH cases, note that $h|z_t|^\mu \neq b$, a.s. for $\mu > 0$, and hence G.2 holds. Regarding the β and the GJR-ARCH cases, note likewise that

$$P(\alpha_1 1_{(z_t<0)}|z_t|^\beta + \alpha_2 1_{(z_t\ge 0)}|z_t|^\beta \neq b)$$
$$= P(z_t < 0)P(\alpha_1|z_t|^\beta \neq b|z_t < 0) + P(z_t \ge 0)P(\alpha_2|z_t|^\beta \neq b|z_t \ge 0) > 0.$$

For the TARCH case note that the variance of the discrete variable $\sum_{i=1}^p \beta_i 1_{A_i}(hz_t)$ is positive as for any $h > 0$, $0 < P_z(hz_t \in A_i) < 1$, $i = 1,\dots,k$, $\lambda(\mathbb{R}\backslash \cup A_i) > 0$ and, finally, that z_t has a positive density. For the SD-ARCH case, note first that geometric ergodicity follows by Lu (1998, Thm. 3.1). This implies that the Markov chain has the Lebesgue measure, λ, as an irreducibility measure. This again implies, by the Lebesgue decomposition, that the invariant measure has a component that has a strictly positive density w.r.t. λ. Thus, $P((y_t^*,\dots,y_{t-q+1}^*) \in A) > 0$ for any set with $\lambda(A) > 0$. In particular,

$$P(\beta' x_{t-1}^* \neq c) = P((y_{t-1}^*,\dots,y_{t-q}^*) \in f_\beta^{-1}(R\backslash\{c\})) > 0$$

because $f_\beta(y_{t-1}^*,\dots,y_{t-q}^*) = \sum_{i=1}^q \beta_{0i}|y_{t-i}^*| + \sum_{i=1}^q\sum_{j=i}^q \beta_{ij}|y_{t-i}^* y_{t-j}^*|$ is a continuous nonconstant function. ■

APPENDIX B: Lemmas

LEMMA 1. *Under C.1–C.4, $L(\theta) = E[l(y_t^*|x_{t-1}^*;\theta)] > -\infty$. Furthermore, it holds that for any set $\mathcal{K} \subseteq \Theta$, $E[\inf_{\theta\in\mathcal{K}} l(y_t^*|x_{t-1}^*;\theta)] > -\infty$.*

Proof. By C.2, $l(y_t^*|x_{t-1}^*;\theta) \geq \log(h_t^*) \geq \log(\omega) \geq \log(\underline{\omega})$, for any $\theta \in \Theta$, implying

$$l^-(y_t^*|x_{t-1}^*;\theta) := -\min\{0, l(y_t^*|x_{t-1}^*;\theta)\} \leq -\min\{0, \log(\underline{\omega})\} < \infty.$$

Hence, $L(\theta)$ is well defined and bounded from below. The first inequality also implies that $E[\inf_{\theta \in \mathcal{K}} l(y_t^*|x_{t-1}^*;\theta)] > -\infty$. ∎

LEMMA 2. *Under C.1–C.4, $h_t^* = h_{0t}^*$ a.s. $\Rightarrow \theta = \theta_0$.*

Proof. Assume that $h_t^* = h_{0t}^*$ a.s., or equivalently,

$$(\omega - \omega_0) + (\alpha - \alpha_0)' x_{t-1}^* = 0. \quad \textbf{(B.1)}$$

To prove that $\theta = \theta_0$, assume first $\alpha = \alpha_0$, which implies $\omega = \omega_0$. If $\alpha \neq \alpha_0$, $(\alpha - \alpha_0)' x_{t-1}^* = -(\omega - \omega_0)$, which contradicts C.3(iii). ∎

LEMMA 3. *Under C.1–C.4, $L(\theta_0) < \infty$ and $L(\theta_0) \leq L(\theta)$ with equality if and only if $h_t^* = h_{0t}^*$* a.s. *In particular, $L(\theta_0) < L(\theta)$ for all $\theta \neq \theta_0$.*

Proof. First observe that $L(\theta_0) < \infty$. Because

$$l(y_t^*|x_{t-1}^*;\theta_0) = \log(h_{0t}^*) + \frac{h_{0t}^* z_t^2}{h_{0t}^*} = \log(h_{0t}^*) + z_t^2,$$

it follows by C.1 and C.4(ii) that

$$E[|l(y_t^*|x_{t-1}^*;\theta_0)|] \leq E[|\log(h_{0t}^*)|] + 1 < \infty.$$

Let $\theta \neq \theta_0$ be given. Then by Lemma 1 either (i) $L(\theta) = \infty$ or (ii) $L(\theta) \in (-\infty, \infty)$. If (i) holds, $L(\theta_0) < \infty = L(\theta)$. If (ii) holds, the following calculations are allowed:

$$L(\theta) = E\left[\log(h_t^*) + \frac{y_t^{*2}}{h_t^*}\right] = E\left[\log(h_t^*) + \frac{h_{0t}^*}{h_t^*} z_t^2\right] = E\left[\log(h_t^*) + \frac{h_{0t}^*}{h_t^*}\right],$$

where we made use of C.1. From the last equality,

$$L(\theta) - L(\theta_0) = E\left[\log\left(\frac{h_t^*}{h_{0t}^*}\right) + \frac{h_{0t}^*}{h_t^*}\right] - 1 \geq 0$$

with equality if and only if $h_{0t}^* = h_t^*$ a.s. Then by Lemma 2, $L(\theta) = L(\theta_0)$ if and only if $\theta = \theta_0$. ∎

LEMMA 4. *Under C.1–C.4, there exist a compact set $\mathcal{C}_0 \subseteq \Theta$, $\theta_0 \in \mathcal{C}_0$, and a function $\underline{l}$ such that $\inf_{\theta \in \Theta \setminus \mathcal{C}_0} l(y_t|x_{t-1};\theta) \geq \underline{l}(y_t|x_{t-1})$ a.s. where $E[\underline{l}(y_t^*|x_{t-1}^*)] = L(\theta_0) + \varepsilon$ for some $\varepsilon > 0$.*

Proof. Choose $\mathcal{C}_0$ as

$$\mathcal{C}_0 \equiv \{\theta \in \Theta | \underline{\omega} \leq \omega \leq \omega_0 + \delta, 0 \leq \alpha_i \leq \alpha_{0,i} + \delta\} \quad \textbf{(B.2)}$$

for some $\delta > 0$ specified subsequently. Let $\theta \in \Theta \backslash C_0$ be given; then either (i) $\omega_0 + \delta < \omega$ or (ii) $\alpha_{0,i} + \delta < \alpha_i$, for at least one $i \in \{1, \dots, m\}$. In the first case,

$$l(y_t | x_{t-1};\theta) \geq \log(h_t) \geq \log(\omega) > \log(\omega_0 + \delta) > \log(\delta),$$

whereas in the second case

$$\begin{aligned} l(y_t | x_{t-1};\theta) &\geq \log(h_t) \geq \log(\alpha_i x_{i,t-1}) > \log(\alpha_{0,i} x_{i,t-1} + \delta x_{i,t-1}) \\ &\geq \log(\delta) + \log(x_{i,t-1}). \end{aligned}$$

Define $\underline{l}(y_t | x_{t-1}) = \log(\delta) + m(x_{t-1})$ where

$$m(x_{t-1}) = \min\{0, \log(x_{1,t-1}), \dots, \log(x_{m,t-1})\},$$

$$\log(\delta) = L(\theta_0) + \varepsilon - E[m(x^*_{t-1})],$$

for some $\varepsilon > 0$. From the preceding inequalities, $l(y_t | x_{t-1};\theta) \geq \underline{l}(y_t | x_{t-1})$ a.s. Finally, by C.4, (iii) $E[|m(x^*_{t-1})|] < \infty$ such that $\delta \in \mathbb{R}$, and

$$E[\underline{l}(y^*_t | x^*_{t-1})] = L(\theta_0) + \varepsilon - E[m(x^*_{t-1})] + E[m(x^*_{t-1})] = L(\theta_0) + \varepsilon. \quad \blacksquare$$

LEMMA 5. *Under C.1–C.4, $\theta \mapsto s(y|x;\theta)$ and $\theta \mapsto Ds(y|x;\theta)$ exist and are continuous for all $(x,y) \in \mathbb{R}^{m+1}$ and $\theta \in \operatorname{int}\Theta$.*

Proof. First observe that $\theta \mapsto h_t$ is twice continuously differentiable, so by the chain rule so is $\theta \mapsto l(y_t | x_{t-1};\theta)$. Next,

$$s_i(y_t | x_{t-1};\theta) = \frac{D_i h_t}{h_t^2}\left(1 - \frac{y_t^2}{h_1}\right) \quad \text{and} \quad D_j s_i(y_t | x_{t-1};\theta) = \frac{D_i h_t D_j h_t}{h_t^2}\left(2\frac{y_t^2}{h_t} - 1\right), \tag{B.3}$$

where $Dh_t = [1, x'_{t-1}]'$ and it has been used that $D^2 h_t = O_{(m+1)\times(m+1)}$. $\blacksquare$

LEMMA 6. *Under C.1–C.4 and N.1 and N.2, there exists a constant $C < \infty$ that does not depend on θ such that for any $\theta \in \mathcal{N}_0$, where $\mathcal{N}_0$ is defined in (A.3),*

(i) $1/h_t \leq C$,
(ii) $y_t^2/h_t \leq C(1 + z_t^2)$, *and*
(iii) $|D_i h_t|/h_t \leq C, i = 1, \dots, m+1$.

Proof. The fact that $h_t \geq \omega \geq \underline{\omega}$ establishes (i). To prove (ii), use that for any $\theta \in \mathcal{N}_0$, $\alpha_i > 0$, $i = 1, \dots, m$, such that

$$\frac{y_t^2}{h_t} = \frac{h_{0t} z_t^2}{h_t} = \frac{\omega_0 + \alpha_0' x_{t-1}}{\omega + \alpha' x_{t-1}} z_t^2 \leq \left[\frac{\omega_0}{\omega} + \sum_{i=1}^m \frac{\alpha_{i,0}}{\alpha_i}\right] z_t^2 \leq \left[\frac{\omega_0}{\omega^l} + \sum_{i=1}^m \frac{\alpha_{i,0}}{\alpha_i^l}\right] z_t^2,$$

$$\frac{|D_1 h_t|}{h_t} = \frac{1}{\omega + \alpha' x_{t-1}} \leq \frac{1}{\omega^l}, \quad \text{and} \quad \frac{|D_i h_t|}{h_t} = \frac{x_{i,t-1}}{\omega + \alpha' x_{t-1}} \leq \frac{1}{\alpha_i^l}, \qquad i = 1, \dots, m. \quad \blacksquare$$

LEMMA 7. *The following conditions hold under C.1–C.4 and N.1 and N.2, with $\mathcal{N}_0$ defined in (A.3):*

(i) $\|s(y_t|x_{t-1};\theta)\| \le C(1+z_t^2)$, $\theta \in \mathcal{N}_0$, *where* $C < \infty$ *does not depend on* θ.
(ii) The moment function $\theta \mapsto V(\theta) = E[s(y_t^*|x_{t-1}^*;\theta)s(y_t^*|x_{t-1}^*;\theta)']$ *is well defined and continuous for* $\theta \in \mathcal{N}_0$, *and* $\sup_{\theta\in\mathcal{N}_0}\|V_n(\theta) - V(\theta)\| \xrightarrow{P} 0$.

Proof. Using Lemma 6,

$$|s_i(y_t|x_{t-1};\theta)| \le \frac{|D_i h_t|}{h_t}\left[1 + \frac{y_t^2}{h_t}\right] \le C_1[1 + C_2(1+z_t^2)] = C(1+z_t^2).$$

To prove (ii), apply the uniform LLN stated in Proposition 1 with $\mathcal{K} = \mathcal{N}_0$ and $f = s$. The conditions stated there hold by the definition of $\mathcal{N}_0$, Lemma 5, and the bound just derived together with N.2. ■

LEMMA 8. *The following conditions hold under C.1–C.4 and N.1 and N.2, with* $\mathcal{N}_0$ *defined in (A.3):*

(i) $\|Ds(y_t|x_{t-1};\theta)\| \le C(1+z_t^2)$, $\theta \in \mathcal{N}_0$, *where* $C < \infty$ *does not depend on* θ.
(ii) The moment function $\theta \mapsto H(\theta) = E[Ds(y_t^*|x_{t-1}^*;\theta)]$ *is well defined and continuous for* $\theta \in \mathcal{N}_0$, *and* $\sup_{\theta\in\mathcal{N}_0}\|H_n(\theta) - H(\theta)\| \xrightarrow{P} 0$.
(iii) The matrix $H(\theta_0) = E[Dh_{0t}^*(Dh_{0t}^*)'/h_{0t}^{*2}]$ *is nonsingular.*

Proof. Applying the bounds established in Lemma 6,

$$|D_j s_i(y_t|x_{t-1};\theta)| \le \frac{|D_i h_t||D_j h_t|}{h_t^2}\left(2\frac{y_t^2}{h_t} + 1\right) \le C(1+z_t^2)$$

for $C < \infty$. Because $\|Ds(y_t|x_{t-1};\theta)\|^2 = \sum_{ij} D_j s_i(y_t|x_{t-1};\theta)^2$, the first part has been shown to hold. The second part follows from Proposition 1 on $f = Ds$ with $\mathcal{K} = \mathcal{N}_0$. Finally, from (B.3),

$$H(\theta_0) = E[Ds(y_t^*|x_{t-1}^*;\theta_0)] = E\left[\frac{Dh_{0t}^*(Dh_{0t}^*)'}{h_{0t}^{*2}}\right],$$

by C.1. Now, assume that $H(\theta_0)$ is not positive definite; then there exists a vector $\lambda = (c,\beta) \neq 0$ such that $\lambda' Dh_{0t}^* = 0$ a.s. because $h_{0t}^* > 0$. That is, $\beta' x_{t-1}^* = -c$ a.s., which is ruled out by C.4. ■

LEMMA 9. *Under C.1–C.4 and N.1 and N.2,* $\sqrt{n}S_n(\theta_0) \xrightarrow{\mathcal{D}} N(0, V(\theta_0))$.

Proof. First observe that $S_n(\theta_0)$ is a martingale w.r.t. $\mathcal{F}_n$ defined in C.3. To see this, first note that by Lemma 7, $s_{0t} \equiv s(y_t|x_{t-1};\theta_0)$ satisfies $E[s_{0t}|\mathcal{F}_{t-1}] = 0$ by C.1, whereas $E[\|s_{0t}^*\|^2] = E[\|s(y_t^*|x_{t-1}^*;\theta_0)\|^2] < \infty$. Hence by the martingale central limit theorem in, e.g., Brown (1971), the result holds if

$$\forall\delta > 0: n^{-1}\sum_{i=1}^{n} E[\|s_{0t}\|^2 1_{\{\|s_{0t}\| > \sqrt{n}\delta\}}] \to 0, \tag{B.4}$$

$$n^{-1}\sum_{i=1}^{n} E[s_{0t}s_{0t}'|\mathcal{F}_{t-1}] \xrightarrow{P} E[s_{0t}^* s_{0t}^{*\prime}], \tag{B.5}$$

where $E[s_{0t}^* s_{0t}^{*\prime}]$ is a positive definite matrix. By Lemma 7, $\|s_{0t}\| \le \bar{s}_t \equiv C(1+z_t^2)$, and hence $\|s_{0t}\|^2 1_{\{\|s_{0t}\|>\sqrt{n}\delta\}} \le \bar{s}_t^2 1_{\{\bar{s}_t>\sqrt{n}\delta\}}$ where the sequence $\{\bar{s}_t\}$ is i.i.d. with a finite second moment by C.1 and N.2. Hence,

$$n^{-1}\sum_{i=1}^{n} E[\|s_{0t}\|^2 1_{\{\|s_{0t}\|>\sqrt{n}\delta\}}] \le n^{-1}\sum_{i=1}^{n} E[\bar{s}_{i,t}^2 1_{\{\bar{s}_{i,t}>\sqrt{n}\delta\}}] = E[\bar{s}_{i,t}^2 1_{\{\bar{s}_{i,t}>\sqrt{n}\delta\}}] \to 0.$$

Next, as $E[s_{0t} s_{0t}' | \mathcal{F}_{t-1}]$ is a function of (y_t, x_{t-1}) with an existing first-order moment (cf. Lemma 7), the strong LLN can be applied, yielding that the averaged sum in (B.5) converges toward

$$V(\theta_0) = E[s_{0t}^* s_{0t}^{*\prime}] = \kappa_4 E\left[\frac{D_\theta h_{0t}^*(D_\theta h_{0t}^*)'}{h_{0t}^{*2}}\right] = \kappa_4 H(\theta_0).$$

Note that this holds irrespective of the initial distribution. Lemma 8 together with the fact that $\kappa_4 > 0$ implies that $V(\theta_0)$ is a positive definite matrix. ■

LEMMA 10. *Assume that C.1, C.2, and C.4 hold. Then G.1 and G.2 imply C.3.*

Proof. Assumption G.1 directly implies C.3(i) and (ii). To prove (iii), let $\beta = (\beta_1', \dots, \beta_q')' \ne 0$, $\beta_i \in \mathbb{R}^p$, $i = 1, \dots, q$, and $\tilde{c} \in \mathbb{R}$ be given. Without loss of generality $\beta_1 \ne 0$, and hence by G.2,

$$P(\beta_1' g(y_t^*) = \tilde{c} | h_{0t}^* = h) = P(\beta_1' g(h^{1/2} z_t) = \tilde{c}) < 1.$$

As $h_{0t}^* = \omega_0 + \sum_{i=1}^q \alpha_{0,i}' g(y_{t-i}^*)$, then for any $Y \in \mathbb{R}^q$ there exists an $h > 0$ such that

$$P(\beta_1' g(y_t^*) = \tilde{c} | (y_{t-1}^*, \dots, y_{t-q}^*) = Y) = P(\beta_1' g(y_t^*) = \tilde{c} | h_{0t}^* = h) < 1. \qquad \textbf{(B.6)}$$

By definition, $\beta' x_t^* = \beta_1' g(y_t^*) + \sum_{i=2}^q \beta_i' g(y_{t-i+1}^*)$, and choosing $\tilde{c} = c - \sum_{i=2}^q \beta_i' g(y_{t-i+1}^*)$, we obtain from (B.6) that for any $c \in \mathbb{R}$,

$$P\left(\beta_1' g(y_t^*) + \sum_{i=2}^q \beta_i' g(y_{t-i+1}^*) = c | (y_{t-1}^*, \dots, y_{t-q}^*) = Y\right) < 1$$

for any $Y \in \mathbb{R}^q$. Hence, as $P(\beta' x_t^* = c) = \int P(\beta' x_t^* = c | (y_{t-1}^*, \dots, y_{t-q}^*)) \times dP(y_{t-1}^*, \dots, y_{t-q}^*)$, $P(\beta' x_t^* = c) < 1$. ■

LEMMA 11. *The* ARCH(1) *process is geometrically ergodic if* $\log(\alpha_1) < -E\log z_t^2$.

Proof. This follows by Carrasco and Chen (2002, Thm. 1 with $A(e) = \alpha_1 e^2$ and $B(e) = \omega$), using Nelson (1990, proof of Thm. 2) together with Nelson (1990, Thms. 3 and 4). ■

APPENDIX C: Auxiliary Results

Here we state a uniform LLN and a consistency result, which are modifications of Tauchen (1985, Lem. 1) and Pfanzagl (1969, Thm. 1.12), respectively. Proofs can be obtained from the authors. In the following discussion $\{X_t\}$ is a geometrically ergodic Markov chain sequence such that a stationary version $\{X_t^*\}$ exists. Observe that all moments are stated in terms of $\{X_t^*\}$, whereas convergence takes place w.r.t. the possibly nonstationary version of $\{X_t\}$. All functions in the following exposition are assumed to be Borel-measurable.

PROPOSITION 1. *Assume that $\mathcal{K} \subseteq \Theta$ is a compact set and that the function $f: \mathbb{R}^d \times \mathbb{R}^d \times \mathcal{K} \mapsto \mathbb{R}^k$ satisfies the following conditions.*

(i) $\theta \mapsto f(y|x;\theta)$ is continuous on $\mathcal{K}$, $\forall x, y \in \mathbb{R}^d$.
(ii) There exists a function $\bar{F}(X_t|X_{t-1})$ not depending on θ such that $\|f(X_t|X_{t-1};\theta)\| \leq \bar{F}(X_t|X_{t-1})$ a.s., $\forall \theta \in \mathcal{K}$, and $E[\bar{F}(X_t^*|X_{t-1}^*;\theta)] < \infty$.

Then, $\sup_{\theta\in\mathcal{K}} \|1/n \sum_t f(X_t|X_{t-1};\theta) - E[f(X_t^*|X_{t-1}^*;\theta)]\| \xrightarrow{a.s.} 0$ *where the mapping $\theta \mapsto E[f(X_t^*|X_{t-1}^*;\theta)]$ is continuous.*

PROPOSITION 2. *Make the following assumptions.*

(i) The parameter space $\mathcal{K}$ is a compact euclidean space with $\theta_0 \in \mathcal{K}$.
(ii) $\theta \mapsto l(y|x;\theta)$ is continuous, $\forall x, y \in \mathbb{R}^d$.
(iii) $L(\theta) = E[l(X_t^*|X_{t-1}^*;\theta)]$ exists, $\forall \theta \in \mathcal{K}$.
(iv) $L(\theta_0) < L(\theta)$, $\forall \theta \neq \theta_0$.
(v) $E[\inf_{\theta\in\mathcal{D}} l(X_t^*|X_{t-1}^*;\theta)] > -\infty$, for any compact set $\mathcal{D} \subset \mathcal{K}$ with $\theta_0 \notin \mathcal{D}$.

Then $\hat{\theta}_n = \arg\min_{\theta\in\mathcal{K}} n^{-1} \sum_{t=1}^n l(X_t|X_{t-1};\theta)$ *is strongly consistent: $\hat{\theta}_n \xrightarrow{a.s.} \theta_0$ as $n \to \infty$.*