European University Institute

Department of Economics

EUI Working Papers

ECO 2008/25

Parameter Estimation in Nonlinear AR-GARCH Models

Mika Meitz and Pentti Saikkonen

EUROPEAN UNIVERSITY INSTITUTE
DEPARTMENT OF ECONOMICS

Parameter Estimation in Nonlinear AR-GARCH Models

MIKA MEITZ

and

PENTTI SAIKKONEN

EUI Working Paper **ECO** 2008/25

ISSN 1725-6704

Printed in Italy
European University Institute
Badia Fiesolana
I – 50014 San Domenico di Fiesole (FI)
Italy

http://www.eui.eu/
http://cadmus.eui.eu/

Parameter estimation in nonlinear AR–GARCH models*

Mika Meitz
University of Oxford

Pentti Saikkonen
University of Helsinki

May 31, 2008

Abstract

This paper develops an asymptotic estimation theory for nonlinear autoregressive models with conditionally heteroskedastic errors. We consider a functional coefficient autoregression of order p (AR(p)) with the conditional variance specified as a general nonlinear first order generalized autoregressive conditional heteroskedasticity (GARCH(1,1)) model. Strong consistency and asymptotic normality of the global Gaussian quasi maximum likelihood (QML) estimator are established under conditions comparable to those recently used in the corresponding linear case. To the best of our knowledge, this paper provides the first results on consistency and asymptotic normality of the QML estimator in nonlinear autoregressive models with GARCH errors.

Keywords: AR-GARCH, asymptotic normality, consistency, nonlinear time series, quasi maximum likelihood estimation.
JEL codes: C13 and C22.

*We acknowledge financial support from the Academy of Finland (PS), the Finnish Foundation for the Advancement of Securities Markets (MM), OP-Pohjola Group Research Foundation (MM and PS), and the Yrjö Jahnsson Foundation (MM and PS). Parts of this research were carried out while the first author was visiting the Center for Research in Econometric Analysis of Time Series (CREATES) at University of Aarhus (funded by the Danish National Research Foundation) and during the second author's Fernand Braudel Fellowship at the European University Institute. Both institutions are thanked for their hospitality. Address correspondence to: Mika Meitz, Department of Economics, University of Oxford, Manor Road Building, Manor Road, Oxford, OX1 3UQ, United Kingdom; e-mail: mika.meitz@economics.ox.ac.uk; or to: Pentti Saikkonen, Department of Mathematics and Statistics, University of Helsinki, P. O. Box 68, FIN–00014 University of Helsinki, Finland; e-mail: pentti.saikkonen@helsinki.fi.

1 Introduction

This paper studies asymptotic estimation theory for nonlinear autoregressive models with conditionally heteroskedastic errors. Such models have been widely used to analyze financial time series ever since the introduction of generalized autoregressive conditionally heteroskedastic (GARCH) models by Engle (1982) and Bollerslev (1986). In addition to 'pure' GARCH models, where the conditional mean is set to zero (or a constant), specifications combining autoregressive moving average (ARMA) type models with errors following a GARCH process (ARMA–GARCH models) have been applied. Furthermore, a variety of nonlinear specifications have been used instead of the linear one (see, e.g., the early survey article by Bollerslev, Engle, and Nelson (1994)).

Asymptotic properties of the (Gaussian) quasi maximum likelihood (QML) estimator in GARCH-type models have been investigated in a number of papers. Contributions in the case of linear pure GARCH models include Lee and Hansen (1994), Lumsdaine (1996), Boussama (2000), Berkes, Horváth, and Kokoszka (2003), Hall and Yao (2003), Jensen and Rahbek (2004), and Francq and Zakoïan (2004, 2007). These papers also contain further references. The linear ARMA–GARCH case has been studied in Weiss (1986), Pantula (1988), Ling and Li (1997, 1998), Ling and McAleer (2003), Francq and Zakoïan (2004), Lange, Rahbek, and Jensen (2006), and Ling (2007a).[1] Of these papers, Weiss (1986), Pantula (1988), and Lange, Rahbek, and Jensen (2006) only deal with ARCH, but not GARCH, errors. Ling and Li (1997, 1998) allow for GARCH errors and establish weak consistency and asymptotic normality of a local, but not global, QML estimator. Their results were extended to the global QML estimator by Ling and McAleer (2003) who proved weak consistency and asymptotic normality under second and sixth order moment conditions, respectively (in the case of ARCH errors, they only needed fourth order moments for asymptotic normality). Strong consistency and asymptotic normality of the global QML estimator were proved by Francq and Zakoïan (2004) under conditions that appear to be the weakest so far. Their consistency result only requires a fractional order moment condition for the observed process and, in the pure GARCH case, they showed that weak moment conditions also suffice for asymptotic normality. However, in the ARMA–GARCH case they still needed finite fourth order moments for the observed process to obtain asymptotic normality. Finally, Lange, Rahbek, and Jensen (2006) and Ling (2007a) consider weighted QML estimators instead of the usual one. As these previous papers indicate, the inclusion of an autoregressive conditional mean entails non-trivial complications for the development of asymptotic estimation theory.

The aforementioned papers are all confined to the linear case. Estimation in nonlinear pure

[1] Estimation theory for related 'double autoregressive' models is developed, among others, by Ling (2007b), where further references can also be found.

ARCH, but not GARCH, models is considered by Kristensen and Rahbek (2005a,b). To the best of our knowledge, Straumann and Mikosch (2006) are the only ones to consider asymptotic estimation theory in nonlinear GARCH models. These authors study QML estimation in a rather general nonlinear pure GARCH model. The examples explicitly treated in their paper are the so-called AGARCH model and EGARCH model. They prove consistency and asymptotic normality of the QML estimator in the case of the AGARCH model but in the EGARCH model only consistency is established. As their work indicates, allowing for nonlinearities in GARCH models considerably complicates the development of asymptotic estimation theory.

In this paper, we consider QML estimation in autoregressive models with GARCH errors and allow both the conditional mean and conditional variance to take general nonlinear forms. Specifically, the conditional mean is modeled as a functional-coefficient autoregression of order p (AR(p)) similar to that in Chen and Tsay (1993) and the conditional variance is specified as a general nonlinear first order GARCH model (GARCH(1,1)). As far as we know, this paper provides the first results on consistency and asymptotic normality of the QML estimator in nonlinear autoregressive models with GARCH errors. Obtaining such results has until recently been hindered by the lack of conditions guaranteeing stationarity and ergodicity for nonlinear AR–GARCH models. Such conditions were recently obtained by Cline (2007) and Meitz and Saikkonen (2008b) whose work opened up the way for the developments of this paper. Based on this previous work, we can only present concrete examples in the case where the conditional heteroskedasticity is modeled by first order GARCH models. This is a major reason why we have decided to leave the extension to higher order GARCH models for future research. Another reason is that the technical difficulties are considerable already in the first order case. An instance of such difficulties is that in one of our examples we have been forced to resort to Markov chain theory to verify identification conditions needed to establish consistency of the QML estimator and positive definiteness of its asymptotic covariance matrix. As far as we know, the only previous reference using a similar approach is Chan and Tong (1986) where Markov chain methods are used to show the positive definiteness of the asymptotic covariance matrix of a QML estimator in a homoskedastic smooth transition autoregressive model. Because our treatment of these issues may also be useful in other nonlinear time series models, this part of the paper may be of independent interest.

In order to relate our paper to previous literature, we note that our results can also be viewed as extensions to those developing asymptotic estimation theory in homoskedastic nonlinear autoregressions. Above we already mentioned the paper by Chan and Tong (1986) which studies a homoskedastic special case of the general model considered in this paper. Another paper related to ours is Tjøstheim (1986) which derives asymptotic properties of least squares and

QML estimators in general nonlinear autoregressions. Although conditional heteroskedasticity is also allowed for, the focus is mainly in homoskedastic models and GARCH type models are not considered. These two papers differ from ours in that they obtain consistency of a local, not global, optimizer of the objective function. There also exists an extensive literature on the estimation theory in general nonlinear dynamic econometric models; for an excellent review and synthesis, see Pötscher and Prucha (1991a,b). However, we have found it difficult to directly apply the general results in this literature, although our proofs are based on the same underlying ideas. A major reason is that, under the assumptions to be used in this paper, a uniform law of large numbers cannot be directly applied to prove the consistency of the QML estimator.

We establish strong consistency and asymptotic normality of the QML estimator under conditions which, when specialized to the linear AR–GARCH model, coincide with the conditions used by Francq and Zakoïan (2004). For consistency, only a mild moment condition is required, whereas existence of fourth order moments of the observed process is needed for asymptotic normality. Thus, the use of our more general nonlinear framework does not come at the cost of more restrictive assumptions. Our results are also closely related to those obtained by Straumann and Mikosch (2006) in the pure GARCH case. As far as the treatment of the conditional variance is concerned, we use ideas similar to theirs in our more general model. Further comparisons to previous work are provided in the subsequent sections.

The rest of this paper is organized as follows. The model considered is introduced in Section 2, and the consistency result is given in Section 3. Differentiability of certain components of the Gaussian likelihood function is treated in Section 4. These results are needed for the asymptotic normality of the QML estimator which is presented in Section 5. Concrete examples are discussed in Section 6, and Section 7 concludes. All proofs are given in Appendices.

Finally, a word on notation and terminology used in this paper. Unless otherwise indicated, all vectors will be treated as column vectors. For the sake of uncluttered notation, we shall write $x = (x_1, ..., x_n)$ for the (column) vector x where the components x_i may be either scalars or vectors (or both). An open interval of the real line will also be denoted as (a, b), but the context will make the meaning clear. For example, we denote $\mathbb{R}_+ = (0, \infty)$. For any scalar, vector, or matrix x, the Euclidean norm is denoted by $|x|$. For a random variable (scalar, vector, or matrix), the L_p–norm is denoted by $\|X\|_p = (E[|X|^p])^{1/p}$, where $p > 0$ (note that this is a vector norm only when $p \geq 1$). If $\|X_n\|_p < \infty$ for all n, $\|X\|_p < \infty$, and $\lim_{n\to\infty} \|X_n - X\|_p = 0$, X_n is said to converge in L_p–norm to X. A random function $X_n(\theta)$ is said to be L_p–dominated in Θ if there exists a positive random variable D_n such that $|X_n(\theta)| \leq D_n$ for all $\theta \in \Theta$ and $\|D_n\|_p < \infty$ uniformly in n. Finally, 'a.s.' stands for 'almost surely'.

2 Model

2.1 Data generation process

We consider a fairly general (univariate) nonlinear autoregressive model with GARCH(1,1) errors. The model is defined by

$$y_t = \sum_{j=1}^{p} a_j \left(y_{t-1}, \ldots, y_{t-p}; \mu_0\right) y_{t-j} + b\left(y_{t-1}, \ldots, y_{t-p}; \mu_0\right) + \sigma_t \varepsilon_t, \quad t = 1, 2, \ldots, \tag{1}$$

where a_j and b are nonlinear functions of p lagged values of y_t and the $m \times 1$ parameter vector μ_0, σ_t is a positive function of y_s, $s < t$, and ε_t is a sequence of independent and identically distributed random variables with $E[\varepsilon_t] = 0$ and $E[\varepsilon_t^2] = 1$ such that ε_t is independent of $\{y_s, \ s < t\}$. Thus, the first two terms on the right hand side specify the conditional mean of y_t whereas σ_t^2, the squared volatility, is the conditional variance. The specification of the conditional variance is assumed to be of the general parametric form

$$\sigma_t^2 = g(u_{0,t-1}, \sigma_{t-1}^2; \theta_0), \tag{2}$$

where $\theta_0 = (\mu_0, \lambda_0)$ with λ_0 an $l \times 1$ parameter vector specific to the conditional variance, and

$$u_{0,t} = y_t - f\left(y_{t-1}, \ldots, y_{t-p}; \mu_0\right) \tag{3}$$

with

$$f\left(z; \mu_0\right) = a\left(z; \mu_0\right)' z + b\left(z; \mu_0\right) \tag{4}$$

and $a\left(z; \mu_0\right) = \left(a_1\left(z; \mu_0\right), \ldots, a_p\left(z; \mu_0\right)\right)$ $(z \in \mathbb{R}^p)$.

We use the subscript '0' to signify true parameter values. Thus, θ_0 is a fixed but unknown and arbitrary point in a parameter space to be specified subsequently and equations (1)–(4) define the generation process of the observed time series used to estimate θ_0. We assume that the data are generated by a stationary and ergodic process with finite moments of some order. Specifically, we make the following assumption.

Assumption DGP. *The process (y_t, σ_t^2) defined by equations (1)–(4) is stationary and ergodic with $E[|y_t|^{2r}] < \infty$ and $E[\sigma_t^{2r}] < \infty$ for some $r > 0$.*

Sufficient conditions for Assumption DGP to hold were recently obtained by Meitz and Saikkonen (2008b). Using theory developed for Markov chains, they give conditions for geometric ergodicity in general nonlinear AR–GARCH models. For their results to hold, they have to assume (in addition to a number of technical assumptions) that the error term ε_t has a positive and lower semicontinuous (Lebesgue) density on $\mathbb{R}$. This is more than needed in some recent work on the estimation of GARCH and ARMA–GARCH models (see Berkes, Horváth, and Kokoszka (2003), Francq and Zakoïan (2004), and Straumann and Mikosch (2006)). Meitz and Saikkonen (2008b) also need rather stringent smoothness conditions on the nonlinear functions in (1) and (2) as well as boundedness of a_j and b. Such conditions are not needed by Cline (2007) who also establishes geometric ergodicity in nonlinear AR–GARCH models. Cline (2007) considers a very general model but his assumptions are not easy to check. Indeed, Cline (2007) only verifies all his assumptions for a threshold model and, as is well-known, a discontinuity in the (Gaussian) likelihood function makes the estimation theory of threshold models with an unknown threshold location nonstandard (see, e.g., Chan (1993)). However, we are able to obtain partial results for a model with a known threshold location in the conditional variance.

As shown in Meitz and Saikkonen (2008b), Assumption DGP can be justified for several widely used models. The conditional mean can be as in a smooth version of the general functional-coefficient autoregressive model of Chen and Tsay (1993) which includes as special cases the exponential autoregressive model of Haggan and Ozaki (1981) and the smooth transition autoregressive models discussed by Teräsvirta (1994) and van Dijk, Teräsvirta, and Franses (2002) among others. In addition to the standard linear GARCH model of Bollerslev (1986) the conditional variance can be a smooth transition GARCH model proposed González-Rivera (1998) and further discussed by Lundbergh and Teräsvirta (2002), Lanne and Saikkonen (2005), and Meitz and Saikkonen (2008a).

Assumption DGP may of course be verified without relying on the results of Meitz and Saikkonen (2008b), although this may be difficult in the case of a general nonlinear model. However, in Section 6 we exemplify this possibility with a model in which the conditional mean is linear and the conditional variance can either be an asymmetric GARCH model (see Ding, Granger, and Engle (1993)) or a threshold GARCH model (see Glosten, Jaganathan, and Runkle (1993) or Zakoïan (1994)).

Regarding the moment conditions in Assumption DGP, they are mild and not stronger than needed in the linear case studied by Francq and Zakoïan (2004). They suffice to prove the consistency of the QML estimator but not asymptotic normality for which more stringent moment conditions, similar to those in Francq and Zakoïan (2004), are needed.

Finally, although Assumption DGP applies to a variety of well-known models it imposes the rather strong requirement that the data are generated by a stationary process, by which we mean that the initial values in (1) and (2) have the stationary distribution. In this respect, our approach is similar to that in Berkes, Horváth, and Kokoszka (2003), Francq and Zakoïan (2004), and Straumann and Mikosch (2006). The possibility to allow for nonstationary initial values in the pure GARCH case is discussed by Straumann and Mikosch (2006, Section 9) but the situation seems quite complicated in our context. We shall say more about this later. In ARCH models the situation is different, for it becomes possible to use limit theorems developed for Markov chains and avoid the assumption of stationary initial values (see Kristensen and Rahbek (2005a)).

2.2 Approximating the conditional variance process

A difficulty with developing estimation theory for the model introduced in the previous section (and even for a pure GARCH model) is that the conditional variance process is not observable and its stationary distribution is, in general, unknown. Thus, even if the value of the true parameter vector θ_0 were known it is not possible to compute the value of the conditional variance σ_t^2 from an observed time series. For that, an initial value with the stationary distribution of σ_t^2 would be needed (see equation (2)) and such an initial value is not available in practice. Thus, because the Gaussian likelihood function depends on the conditional variance we have to use an approximation.

Motivated by the preceding discussion we introduce the process

$$h_t(\theta) = \begin{cases} \varsigma_0, & t = 0, \\ g(u_{t-1}, h_{t-1}(\theta); \theta), & t = 1, 2, \ldots, \end{cases} \tag{5}$$

where $\theta = (\mu, \lambda)$ is an $(m + l) \times 1$ parameter vector with true value $\theta_0 = (\mu_0, \lambda_0)$ and $u_t = y_t - f(y_{t-1}, \ldots, y_{t-p}; \mu)$. Once the initial value ς_0 has been specified one can use equation (5) to compute $h_t(\theta)$, $t = 1, 2, ...$, recursively for any chosen value of the parameter vector θ. For simplicity, we assume the initial value ς_0 to be a positive constant independent of θ, which is also the choice most common in practice.[2] When there is no need to make the dependence of $h_t(\theta)$ explicit about the parameter vector θ we use the notation h_t. Similarly, the short-hand notation $f_t = f_t(\mu) = f(y_{t-1}, \ldots, y_{t-p}; \mu)$ will sometimes be used.

If the results of Meitz and Saikkonen (2008b) are used to justify the ergodicity assumed in Assumption DGP then, given any initial value, the conditional distribution of $h_t(\theta_0)$ approaches the stationary distribution of the true conditional variance σ_t^2 as $t \to \infty$. Furthermore, limit theorems developed for Markov chains apply to realizations of the process $(y_t, h_t(\theta_0))$. Unfortunately, however, this is not sufficient to prove consistency and asymptotic normality of the QML estimator of the parameter vector θ_0. The reason is that in these proofs one has to consider the

[2]The results in this paper could be generalized to the case of a stochastic initial value $\varsigma_0(\theta)$ depending on θ, but, to avoid additional technical complications, we have decided not to pursue this matter.

process $h_t(\theta)$ for parameter values different from the true value θ_0 but the results of Meitz and Saikkonen (2008b) only apply to the process $h_t(\theta_0)$ and say nothing about properties of $h_t(\theta)$ when $\theta \neq \theta_0$. Another point to note is that the process $h_t(\theta)$ depends on the entire past history of the observed process y_t. If $h_t(\theta)$ were a function of a fixed finite number of lagged values of y_t the aforementioned difficulty could be overcome, for the stationarity and ergodicity of y_t would make it possible to apply well-known limit theorems to statistics involving the process $h_t(\theta)$. In ARCH models this is the case and explains why the development of asymptotic estimation theory is not hampered by nonstationary initial values (see Kristensen and Rahbek (2005a)).

The preceding discussion means that we have to study properties of the process $h_t(\theta)$ for all $\theta = (\mu, \lambda)$ in a permissible parameter space. Due to the relatively simple structure of the standard GARCH model this is quite straightforward in the linear ARMA–GARCH model considered by Francq and Zakoïan (2004). However, nonlinear GARCH models are considerably more difficult, as the recent work of Straumann and Mikosch (2006) shows. Our approach is to follow these authors and extend some of their arguments to a model with a nonlinear conditional mean. To this end, we impose the following assumptions which are central in proving the consistency of the QML estimator. The permissible parameter spaces of μ and λ are denoted by M and Λ, respectively, so that their product $\Theta = \mathsf{M} \times \Lambda$ defines the permissible space of θ.

Assumption C1. *The true parameter value $\theta_0 \in \Theta = \mathsf{M} \times \Lambda$, where M and Λ are compact subsets of $\mathbb{R}^m$ and $\mathbb{R}^l$, respectively.*

Assumption C2. *The function $g : \mathbb{R} \times \mathbb{R}_+ \times \Theta \to \mathbb{R}_+$ is continuous with respect to all its arguments and satisfies the following two conditions.*

(i) *For some $0 < \varrho < 1$ and $0 < \varkappa, \varpi < \infty$, $g(u, x; \theta) \leq \varrho x + \varkappa u^2 + \varpi$ for all $\theta \in \Theta$, $u \in \mathbb{R}$, and $x \in \mathbb{R}_+$.*

(ii) *For some $0 < \kappa < 1$, $|g(u, x_1; \theta) - g(u, x_2; \theta)| \leq \kappa |x_1 - x_2|$ for all $\theta \in \Theta$, $u \in \mathbb{R}$, and $x_1, x_2 \in \mathbb{R}_+$.*

Assumption C3. *The functions $a : \mathbb{R}^p \times \mathsf{M} \to \mathbb{R}^p$ and $b : \mathbb{R}^p \times \mathsf{M} \to \mathbb{R}$ are such that $a(\cdot; \mu)$ and $b(\cdot; \mu)$ are bounded uniformly in μ and Borel measurable for every μ.*

As usual in nonlinear estimation problems, Assumption C1 requires the parameter space to be compact. From a mathematical point of view this assumption provides a convenient simplification although it may not be easy to justify in practice. Assumption C2 is more stringent than needed to justify Assumption DGP (see Assumption 4 in Meitz and Saikkonen (2008b)). This particularly holds for the Lipschitz condition in Assumption C2(ii). It would be possible to relax this condition along the lines in Straumann and Mikosch (2006) but we prefer not to pursue this matter because it would complicate the exposition and we have no example where such a relaxed condition would be necessary. From a mathematical point of view, the boundedness assumption in Assumption C3 is rather stringent but still satisfied by several functional-coefficient autoregressive models including the exponential autoregressive model and various smooth transition autoregressive models (see Haggan and Ozaki (1981), Teräsvirta (1994), and van Dijk, Teräsvirta, and Franses (2002)).

Using Assumptions C1–C3 we can prove the following result.

Proposition 1. *Suppose Assumptions DGP and C1–C3 hold. Then, for all $\theta \in \Theta$ there exists a stationary and ergodic solution $h_t^*(\theta)$ to the equation*

$$h_t(\theta) = g(u_{t-1}, h_{t-1}(\theta); \theta), \quad t = 1, 2, \ldots. \tag{6}$$

This solution is continuous in θ, measurable with respect to the σ–algebra generated by $(y_{t-1}, y_{t-2}, ...)$, and it is unique when (6) is extended to all $t \in \mathbb{Z}$. Furthermore, the solution $h_t^*(\theta)$ has the properties $h_t^*(\theta_0) = \sigma_t^2$ and $E[\sup_{\theta\in\Theta} h_t^{*r}(\theta)] < \infty$, and, if $h_t(\theta)$, $\theta \in \Theta$, are any other solutions to the equation (6), then for some $\gamma > 1$, $\gamma^t \sup_{\theta\in\Theta}|h_t^*(\theta) - h_t(\theta)| \to 0$ in L_r–norm as $t \to \infty$.

Proposition 1 is proved in Appendix B by using an analogous more general lemma given in Appendix A. This lemma is similar to Theorem 3.1 of Bougerol (1993) and Theorem 2.8 of Straumann and Mikosch (2006) although more specific. Proposition 1 shows that the stationary solution $h_t^*(\theta_0)$ to equation (6) with $\theta = \theta_0$ coincides with the true conditional variance of the data generation process and that any other solution obtained with $\theta = \theta_0$ converges to the true conditional variance exponentially fast. Note, however, that the mode of convergence is different from that in the aforementioned result of Meitz and Saikkonen (2008b). Also, the convergence to the stationary solution does not only hold for the true parameter value θ_0 but uniformly over the parameter space Θ. This last fact and the existence of the stationary and ergodic solution $h_t^*(\theta)$ will be of importance in our subsequent developments. Indeed, with Proposition 1 (and assumptions to be imposed later) we can prove the consistency and asymptotic normality of the QML estimator of the parameter vector θ_0. As already mentioned, this requires more stringent conditions about the function g than needed to establish the geometric ergodicity of the data generation process. It is worth noting that no similar strengthening is needed for the function f (i.e., the functions a_j $(j = 1, ..., p)$ and b) that specifies the conditional mean of the model. This is due to the fact that the technique used to prove Proposition 1 (and the aforementioned theorems of Bougerol (1993) and Straumann and Mikosch (2006)) is only needed for the conditional variance process, and not for the conditional mean. Had we needed a similar method for the conditional mean, this might have lead to Lipschitz (contraction) conditions also for the function f, which could have considerably restricted the type of permitted nonlinearity.

3 Consistency of the QML estimator

Suppose we have an observed time series $y_{-p}, \ldots, y_0, y_1, \ldots, y_T$ generated by the stationary and ergodic process defined by equations (1)–(4) (cf. Assumption DGP). We shall estimate the unknown parameter vector θ_0 by minimizing the objective function

$$L_T(\theta) = T^{-1} \sum_{t=1}^{T} l_t(\theta), \quad \text{where } l_t(\theta) = \log(h_t) + \frac{u_t^2}{h_t}$$

and $u_t = y_t - f(y_{t-1}, \ldots, y_{t-p}; \mu)$ and h_t are as in (3) and (5) with dependence on the parameter vectors μ and θ suppressed. Clearly, $L_T(\theta)$ is an approximation to the conditional Gaussian log-likelihood multiplied by $-2/T$. We do not assume Gaussianity, however, so that the resulting estimator is a QML estimator. Conditioning is on the first $p+1$ observations and the initial value ς_0 needed to compute the approximate conditional variances $h_t(\theta)$ $(t = 1, ..., T)$. It follows from Proposition 1 that $h_t(\theta)$ approximates the stationary solution to equation (6) which for $\theta = \theta_0$ coincides with the true conditional variance σ_t^2.

We also define

$$L_T^*(\theta) = T^{-1} \sum_{t=1}^{T} l_t^*(\theta), \quad \text{where } l_t^*(\theta) = \log(h_t^*) + \frac{u_t^2}{h_t^*}$$

and $h_t^* = h_t^*(\theta)$ is the stationary and ergodic solution to equation (6) (see Proposition 1). Due to stationarity, the function $L_T^*(\theta)$ is easier to work with than $L_T(\theta)$ and, using assumptions to be made below, it turns out that minimizers of $L_T^*(\theta)$ and $L_T(\theta)$ are asymptotically equivalent.

In addition to the assumptions already made we need further assumptions about the nonlinear functions used to model the conditional mean and conditional variance. Regarding the conditional mean, we impose the following assumption.

Assumption C4. *The functions* $a : \mathbb{R}^p \times \mathsf{M} \to \mathbb{R}^p$ *and* $b : \mathbb{R}^p \times \mathsf{M} \to \mathbb{R}$ *are such that* $a(z; \cdot)$ *and* $b(z; \cdot)$ *are continuous for every* $z \in \mathbb{R}^p$.

The continuity of the functions a and b combined with the continuity of the function g imposed in Assumption C2 ensures that the Gaussian log-likelihood function $L_T(\theta)$ is continuous. This is a common requirement in nonlinear estimation problems and, in conjunction with the assumed compactness of the parameter space Θ, it implies the existence of a measurable minimizer $\hat{\theta}_T = (\hat{\mu}_T, \hat{\lambda}_T)$ of $L_T(\theta)$ (see, e.g., Pötscher and Prucha (1991a), Lemma 3.4). In view of the continuity of $h_t^*(\theta)$ established in Proposition 1 the same is true for a minimizer of $L_T^*(\theta)$.

As for the conditional variance, we have to supplement Assumption C2 by the following technical condition.

Assumption C5. *The function* $g : \mathbb{R} \times \mathbb{R}_+ \times \Theta \to \mathbb{R}_+$ *is bounded away from zero in the sense that* $\inf_{(u,x,\theta)\in\mathbb{R}\times\mathbb{R}_+\times\Theta} g(u, x; \theta) = \underline{g}$ *for some* $\underline{g} > 0$.

This condition bounds the function g away from zero in the same way as, for example, Assumption C.3 of Straumann and Mikosch (2006). This assumption is somewhat unnatural but appears difficult to avoid even in pure ARCH models (cf. condition C.2 in Kristensen and Rahbek (2005a)).

Our final assumption for the consistency of the QML estimator $\hat{\theta}_T$ is the following identification condition.

Assumption C6.

(i) $f(y_{t-1}, ..., y_{t-p}; \mu) = f(y_{t-1}, ..., y_{t-p}; \mu_0)$ *a.s. only if* $\mu = \mu_0$.[3]

(ii) $h_t^*(\mu_0, \lambda) = \sigma_t^2$ *a.s. only if* $\lambda = \lambda_0$.

As will be seen in the proof of Theorem 1 (Appendix B), given the assumptions so far, Assumption C6 is equivalent to $E[L_T^*(\theta)]$ being uniquely minimized at θ_0. In the present context, this is essentially equivalent to θ_0 being an identifiably unique minimizer of $L_T^*(\theta)$ in the sense of Pötscher and Prucha (1991a, Definition 3.1) and White (1980, Definition 2.1).[4] Although more explicit than an identifiable uniqueness condition, the conditions in Assumption C6 are still of a general nature, and in particular cases they have to be verified by using more basic assumptions about the functional forms of the specified conditional mean and conditional variance. In nonlinear cases this turns out to be difficult, and we next provide some comments on this.

So far, there appears to be rather limited previous work available on the verification of an identification condition such as C6(i) in nonlinear autoregressive models of the type considered in this paper. Although Chan and Tong (1986) and Tjøstheim (1986) consider estimation in homoskedastic nonlinear autoregressions with structures similar to ours, their results concern a local, not global, minimizer of the objective function, and therefore they need not verify an identification condition corresponding to C6(i). Lai (1994) considers (global) least squares estimation in nonlinear regression models, and his identification condition (2.2) is related to ours. However, he does not verify this condition in any examples similar to ours. It appears

[3]This condition could also be expressed by using the functions a and b as in (4).

[4]'Essentially' equivalent because in our situation $E[L_T^*(\theta)]$ takes values in $\mathbb{R} \cup \{+\infty\}$ instead of $\mathbb{R}$; if $E[L_T^*(\theta)]$ is finite in Θ, compactness of Θ and lower semi-continuity of $E[L_T^*(\theta)]$ (to be shown in the proof of Theorem 1) suffice for this equivalence.

challenging to verify condition C6(i) in a nonlinear autoregression with a nonlinear structure sufficiently general for the results to be applicable in practice. For instance, general results such as those provided by Pötscher and Prucha (1991a) do not consider verifying conditions of this kind. In one of our examples we have found it difficult to verify condition C6(i) without resorting to rather complicated derivations that involve the application of Markov chain theory. The basic idea is to impose suitable assumptions on the function f so that, for every $\mu \neq \mu_0$, there exists a (Borel) measurable set $A \subset \mathbb{R}^p$ such that $f(z;\mu) \neq f(z;\mu_0)$ for all $z \in A$. Then condition C6(i) clearly holds if the event $\{(y_{t-1},...,y_{t-p}) \in A\}$ has positive probability. Using Markov chain theory it is possible to show that events of this kind indeed have positive probability even though the precise form of the stationary distribution of the process y_t is unknown.

Regarding condition C6(ii), it agrees with the identification condition used by Straumann and Mikosch (2006) in their nonlinear GARCH model (we are not aware of any other papers dealing with identification in nonlinear GARCH models). However, in their examples they do not consider nonlinearities as complicated as we do, and, therefore, they do not need to rely on Markov chain theory to verify the identification condition (although even in their case the verification is quite complicated). One of our examples is again rather difficult and we have been forced to resort to Markov chain theory to verify condition C6(ii).

As a final remark we note that in the verification of Assumption C6, it may also be necessary to make assumptions about the distribution of the error term ε_t. For instance, in order to prove consistency in a linear ARMA–GARCH model, Francq and Zakoïan (2004) assume that the distribution of ε_t^2 is non-degenerate and a similar condition also appears in Straumann and Mikosch (2006, Theorems 5.1 and 5.5). However, in nonlinear cases much more may need to be assumed, as one of our examples suggests.

Now we can state our consistency result which is proved in Appendix B.

Theorem 1. *Suppose Assumptions DGP and C1–C6 hold. Then the QML estimator $\hat{\theta}_T$ is strongly consistent, that is, $\hat{\theta}_T \to \theta_0$ a.s.*

The proof of this theorem makes use of the relation between the Gaussian log-likelihood function $L_T(\theta)$ and its stationary and ergodic counterpart $L_T^*(\theta)$. Instead of the QML estimator $\hat{\theta}_T$ the proof is reduced to its infeasible analog obtained by minimizing $L_T^*(\theta)$ (for details, see Appendix B). The same approach has also been used in the related previous work of Berkes, Horváth, and Kokoszka (2003), Francq and Zakoïan (2004), and Straumann and Mikosch (2006). Similarly to these authors, we can prove consistency with very mild moment conditions (see Assumption DGP). As a final remark we note that, with our assumptions, a 'classical' consistency proof relying on an application of a uniform law of large numbers (see, e.g., Pötscher and Prucha

(1991a)) is not directly applicable. Therefore, our proof relies on alternative (though well-known) arguments similar to those also used by Straumann and Mikosch (2006) in part 2 of their proof of Theorem 4.1 (for details, see Appendix B).

4 Derivatives of the approximate conditional variance process

For the asymptotic normality of the QML estimator of the parameter vector θ_0 we subsequently need to consider the first and second derivatives of the objective function $L_T(\theta)$ as well as its stationary ergodic counterpart $L_T^*(\theta)$. A complication that arises is the differentiability of the processes h_t and h_t^*. In this section we give conditions under which both of these processes are twice continuously (partially) differentiable and the derivatives of h_t converge to those of h_t^*. Similarly to Subsection 2.2, the differentiability of h_t and h_t^* is more straightforward in the case of a linear ARMA–GARCH model considered by Francq and Zakoïan (2004). In nonlinear GARCH models the situation is rather complex, and again our approach is to follow the arguments in Straumann and Mikosch (2006) and extend them to our case with a nonlinear conditional mean.

We begin with some assumptions.

Assumption N1. *The true parameter value θ_0 is an interior point of Θ.*

Assumption N1 is necessary for the asymptotic normality of the QML estimator. Together with the differentiability assumptions to be imposed shortly it allows us to use a conventional Taylor series expansion of the score. Estimation in linear GARCH models when θ_0 is allowed to be on the boundary of the parameter space has only recently been considered by Francq and Zakoïan (2007) (see also Andrews (1999)). In this case, the resulting asymptotic distribution is no longer normal. We leave this for future research.

Assumption N1 together with the consistency of the QML estimator implies that in the subsequent analysis we (without loss of generality) only need to consider parameter values in an arbitrarily small open ball centered at θ_0. For concreteness, let Θ_0 be a compact convex set contained in the interior of Θ that has θ_0 as an interior point. This gives us a suitable set Θ_0 on which to investigate the differentiability and the validity of the Taylor expansions of the objective functions $L_T(\theta)$ and $L_T^*(\theta)$ and their components. The assumed compactness will be convenient because we will apply Lemma A.3 (in Appendix A) to examine the differentiability of the processes h_t and h_t^* on Θ_0. On the other hand, convexity ensures that all intermediate points obtained from Taylor expansions will also be in Θ_0.

To present the next assumption, we partition the set Θ_0 as $\Theta_0 = \mathsf{M}_0 \times \Lambda_0$.

Assumption N2. *The function $g(\cdot,\cdot;\cdot)$ is twice continuously partially differentiable on $\mathbb{R}\times\mathbb{R}_+\times\Theta_0$ and the functions $a(z;\cdot)$ and $b(z;\cdot)$ are twice continuously partially differentiable on M_0 for every $z\in\mathbb{R}^p$.*

Assumption N2 is necessary for the differentiability of the objective function $L_T(\theta)$ on the set Θ_0, and is similar to (parts of) Assumptions D.1 and D.3 of Straumann and Mikosch (2006). A difference to these assumptions is that due to the presence of the conditional mean, the function g is required to be differentiable also with respect to its first argument (we will see in Section 6, Example 2, that this additional requirement turns out to be restrictive).

We next impose restrictions on the derivatives of the functions g, a, and b. Denote the first and second partial derivatives of g with $g_{\upsilon_1}=\partial g(u,h;\theta)/\partial\upsilon_1$ and $g_{\upsilon_1\upsilon_2}=\partial^2 g(u,h;\theta)/\partial\upsilon_1\partial\upsilon_2'$, where υ_1 and υ_2 can be any of u, h, or θ. Define a_μ, $a_{\mu\mu}$, b_μ, and $b_{\mu\mu}$ similarly (e.g. $a_\mu=\partial a(z;\mu)/\partial\mu$).

Assumption N3.

(i) *For some $C<\infty$ and all $\mu\in\mathsf{M}_0$ and $z\in\mathbb{R}^p$, the quantities $|a_\mu|$, $|a_{\mu\mu}|$, $|b_\mu|$, and $|b_{\mu\mu}|$ are bounded by C.*

(ii) *For some $C<\infty$ and all $\theta\in\Theta_0$, $u\in\mathbb{R}$, and $x\in\mathbb{R}_+$, the quantities $|g_\theta|$, $|g_u|$, $|g_{\theta\theta}|$, $|g_{uu}|$, $|g_{\theta u}|$, and $|g_{u\theta}|$ (evaluated at $(u,x;\theta)$) are bounded by $C(1+u^2+x)$.*

(iii) *For some $\kappa'<\infty$ and all $u\in\mathbb{R}$ and $x_1,x_2\in\mathbb{R}_+$,*

$$\begin{aligned}|g_\upsilon(u,x_1;\theta)-g_\upsilon(u,x_2;\theta)| &\le \kappa'|x_1-x_2|, \qquad \upsilon=u,h,\theta,\\ |g_{\upsilon_1\upsilon_2}(u,x_1;\theta)-g_{\upsilon_1\upsilon_2}(u,x_2;\theta)| &\le \kappa'|x_1-x_2|, \qquad \upsilon_1,\upsilon_2=u,h,\theta.\end{aligned}$$

The first condition in Assumption N3 places further restrictions on the behaviour of the functions a and b in the conditional mean function. Like the boundedness conditions already imposed on them in Assumption C3, these conditions may be stringent from a mathematical point of view but are typically satisfied in applications. The second and third parts of Assumption N3 are related to conditions C2(i) and (ii) already imposed on the function g. The condition in N3(ii) is used to ensure the existence of certain moments involving the partial derivatives of g (a less stringent condition would also suffice, but this one is used for its simplicity). Condition N3(iii) is a Lipschitz continuity requirement for the partial derivatives of g but, unlike the condition on the function g itself in C2(ii), the partial derivatives need not be contractions (i.e., κ' does not need to be less than one).

We now introduce further notation that is needed to present the derivatives of h_t and h_t^* in a reasonably concise form. Denote the first and second partial derivatives of the function $h_t(\theta)$ with $h_{\theta,t} = \partial h_t(\theta)/\partial\theta$ and $h_{\theta\theta,t} = \partial^2 h_t(\theta)/\partial\theta\partial\theta'$, respectively. Similarly, denote $f_{\theta,t} = \partial f_t(\theta)/\partial\theta$ and $f_{\theta\theta,t} = \partial^2 f_t(\theta)/\partial\theta\partial\theta'$ (note that $f_{\theta,t} = -\partial u_t(\theta)/\partial\theta$ and $f_{\theta\theta,t} = -\partial^2 u_t(\theta)/\partial\theta\partial\theta'$, and also that although both f_t and u_t depend only on μ and not on λ, we will often use the argument θ for simplicity). Furthermore, let $g_{\upsilon_1,t} = [g_{\upsilon_1}]_{u=u_{t-1}(\theta),h=h_{t-1}(\theta)} = \partial g(u_{t-1}(\theta), h_{t-1}(\theta);\theta)/\partial\upsilon_1$ denote the first partial derivative of g evaluated at $u = u_{t-1}(\theta)$ and $h = h_{t-1}(\theta)$, and define $g_{\upsilon_1\upsilon_2,t}$ similarly (υ_1 and υ_2 can be any of u, h, or θ). Finally, all the derivatives may be partitioned conformably with the partition $\theta = (\mu, \lambda)$, and θ is replaced with either μ or λ when denoting these blocks (for example, $h_{\theta,t} = (h_{\mu,t}, h_{\lambda,t})$; note also that $f_{\lambda,t}$, $f_{\lambda\lambda,t}$, $f_{\mu\lambda,t}$, and $f_{\lambda\mu,t}$ are zero vectors or matrices).

The first and second derivatives of the difference equation $h_t = g(u_{t-1}, h_{t-1}; \theta)$, $t = 1, 2, ...,$ can now be derived by straightforward but tedious differentiation. We have

$$\begin{aligned}
h_{\theta,t} &= g_{\theta,t} - g_{u,t} f_{\theta,t-1} + g_{h,t} h_{\theta,t-1}, \qquad t = 1, 2, ..., \\
h_{\theta\theta,t} &= g_{\theta\theta,t} + g_{uu,t} f_{\theta,t-1} f'_{\theta,t-1} - f_{\theta,t-1} g_{u\theta,t} - g_{\theta u,t} f'_{\theta,t-1} - g_{u,t} f_{\theta\theta,t-1} \\
&\quad + (g_{\theta h,t} - g_{uh,t} f_{\theta,t-1}) h'_{\theta,t-1} + h_{\theta,t-1} \left(g_{h\theta,t} - g_{hu,t} f'_{\theta,t-1}\right) \\
&\quad + g_{hh,t} h_{\theta,t-1} h'_{\theta,t-1} + g_{h,t} h_{\theta\theta,t-1}, \qquad t = 1, 2, ...,
\end{aligned}$$

where the recursions are initialized from a zero vector and matrix, respectively. For further conciseness we denote

$$\alpha_{\theta,t} = g_{\theta,t} - g_{u,t} f_{\theta,t-1}, \qquad \beta_t = g_{h,t}, \qquad \gamma_{\theta,t} = g_{\theta h,t} - g_{uh,t} f_{\theta,t-1}, \qquad \delta_t = g_{hh,t}, \tag{7}$$

$$\alpha_{\theta\theta,t} = g_{\theta\theta,t} + g_{uu,t} f_{\theta,t-1} f'_{\theta,t-1} - f_{\theta,t-1} g_{u\theta,t} - g_{\theta u,t} f'_{\theta,t-1} - g_{u,t} f_{\theta\theta,t-1}, \tag{8}$$

and with this notation the derivatives of h_t satisfy the difference equations

$$h_{\theta,t} = \alpha_{\theta,t} + \beta_t h_{\theta,t-1}, \quad t = 1, 2, ..., \tag{9}$$

$$h_{\theta\theta,t} = \alpha_{\theta\theta,t} + \beta_t h_{\theta\theta,t-1} + \gamma_{\theta,t} h'_{\theta,t-1} + h_{\theta,t-1} \gamma'_{\theta,t} + \delta_t h_{\theta,t-1} h'_{\theta,t-1}, \quad t = 1, 2, ... \tag{10}$$

We also define stationary ergodic counterparts of the quantities appearing in (7)–(8). To this end, let $g^*_{\upsilon_1,t} = [g_{\upsilon_1}]_{u=u_{t-1}(\theta),h=h^*_{t-1}(\theta)} = \partial g\left(u_{t-1}(\theta), h^*_{t-1}(\theta);\theta\right)/\partial\upsilon_1$ denote this partial derivative evaluated at $u = u_{t-1}(\theta)$ and $h = h^*_{t-1}(\theta)$, where $h^*_t(\theta)$ is the stationary ergodic solution obtained from Proposition 1, and define $g^*_{\upsilon_1\upsilon_2,t}$ similarly (υ_1 and υ_2 can be any of u, h, or θ). Furthermore, let $\alpha^*_{\theta,t}$, β^*_t, $\gamma^*_{\theta,t}$, δ^*_t, and $\alpha^*_{\theta\theta,t}$ denote the analogously defined counterparts of the quantities in (7)–(8) (for example, $\beta^*_t = g^*_{h,t} = \partial g(u_{t-1}(\theta), h^*_{t-1}(\theta);\theta)/\partial h$).

Given these assumptions and notation, we obtain the following result.

Proposition 2. *Suppose Assumptions DGP, C1–C6, and N1–N3 hold.*
(a) *For all $\theta \in \Theta_0$ there exists a stationary ergodic solution $h^*_{\theta,t}(\theta)$ to the equation*

$$h_{\theta,t}(\theta) = \alpha^*_{\theta,t} + \beta^*_t h_{\theta,t-1}(\theta), \quad t = 1, 2, ... \tag{11}$$

This solution is measurable with respect to the σ–algebra generated by $(y_{t-1}, y_{t-2}, ...)$, it is unique when (11) is extended to all $t \in \mathbb{Z}$, and $E\big[\sup_{\theta\in\Theta_0} |h^*_{\theta,t}(\theta)|^{r/2}\big] < \infty$. Furthermore, the stationary ergodic solution $h^*_t(\theta)$ obtained from Proposition 1 is continuously partially differentiable on Θ_0 for every $t \in \mathbb{Z}$ and $\partial h^*_t(\theta)/\partial\theta = h^*_{\theta,t}(\theta)$.
(b) *If $h_t(\theta)$ and $h_{\theta,t}(\theta)$, $\theta \in \Theta_0$, are any solutions to the difference equations (6) and (9), respectively, then for some $\gamma > 1$, $\gamma^t \sup_{\theta\in\Theta_0} |h^*_{\theta,t}(\theta) - h_{\theta,t}(\theta)| \to 0$ in $L_{r/4}$–norm as $t \to \infty$.*

Proposition 2(a) shows that $h^*_t(\theta)$ is continuously differentiable and that its derivative coincides with $h^*_{\theta,t}(\theta)$, the stationary ergodic solution to (11). Part (b) of the proposition shows that for any other solution $h_t(\theta)$ to equation (6), its derivative $h_{\theta,t}(\theta)$ converges to $h^*_{\theta,t}(\theta)$ exponentially fast and uniformly over Θ_0. These facts will be of importance when we subsequently consider the first derivatives of the objective function $L_T(\theta)$ and its stationary ergodic counterpart $L^*_T(\theta)$. In particular, using part (a) we can show that $L^*_T(\theta)$ is continuously differentiable with a stationary and ergodic derivative, whereas using part (b) we can establish that this derivative provides an approximation to the first derivative of $L_T(\theta)$.

Our next proposition gives an analogous result for the second derivatives.

Proposition 3. *Suppose Assumptions DGP, C1–C6, and N1–N3 hold.*
(a) *For all $\theta \in \Theta_0$ there exists a stationary ergodic solution $h^*_{\theta\theta,t}(\theta)$ to the equation*

$$h_{\theta\theta,t}(\theta) = \alpha^*_{\theta\theta,t} + \beta^*_t h_{\theta\theta,t-1}(\theta) + \gamma^*_{\theta,t} h^{*\prime}_{\theta,t-1}(\theta) + h^*_{\theta,t-1}(\theta)\gamma^{*\prime}_{\theta,t} + \delta^*_t h^*_{\theta,t-1}(\theta) h^{*\prime}_{\theta,t-1}(\theta), \;\; t = 1, 2, ... \tag{12}$$

This solution is measurable with respect to the σ–algebra generated by $(y_{t-1}, y_{t-2}, ...)$, it is unique when (12) is extended to all $t \in \mathbb{Z}$, and $E\big[\sup_{\theta\in\Theta_0} |h^*_{\theta\theta,t}(\theta)|^{r/4}\big] < \infty$. Furthermore, the stationary ergodic solution $h^*_t(\theta)$ obtained from Proposition 1 is twice continuously partially differentiable on Θ_0 for every $t \in \mathbb{Z}$ and $\partial^2 h^*_t(\theta)/\partial\theta\partial\theta' = h^*_{\theta\theta,t}(\theta)$.
(b) *If $h_t(\theta)$, $h_{\theta,t}(\theta)$, and $h_{\theta\theta,t}(\theta)$, $\theta \in \Theta_0$, are any solutions to the difference equations (6), (9), and (10), respectively, then for some $\gamma > 1$, $\gamma^t \sup_{\theta\in\Theta_0} |h^*_{\theta\theta,t}(\theta) - h_{\theta\theta,t}(\theta)| \to 0$ in $L_{r/8}$–norm as $t \to \infty$.*

The results of Proposition 3 are analogous to those of Proposition 2. We note that in the moment and convergence results obtained for $h^*_{\theta,t}$ and $h^*_{\theta\theta,t}$ in Propositions 2 and 3, respectively, the exact orders ($r/2$, $r/4$, or $r/8$) are not crucial as long as these results hold for some positive exponents. Our approach here is somewhat different from the one used by Straumann and Mikosch (2006, Propositions 6.1 and 6.2) in that we obtain moment results for $h^*_{\theta,t}$ and $h^*_{\theta\theta,t}$ and use convergence in L_p–norm instead of the almost sure convergence used by them. As a consequence, the use of these results in subsequent proofs appears to lead to less complex and more transparent derivations.[5]

5 Asymptotic normality of the QML estimator

As already indicated, the moment conditions used to prove strong consistency of the QML estimator are not sufficient to establish asymptotic normality. Further restrictions are needed for the moments of the observed process as well as for the derivatives of the process $h^*_t(\theta)$. We make the following assumption.

Assumption N4. *Assumption DGP holds with $r = 2$, the random variables ε_t satisfy $E[\varepsilon_t^4] < \infty$, and*

$$\left\| \sup_{\theta\in\Theta_0} \frac{|h^*_{\theta,t}(\theta)|}{h^*_t(\theta)} \right\|_4 < \infty \quad \textit{and} \quad \left\| \sup_{\theta\in\Theta_0} \frac{|h^*_{\theta\theta,t}(\theta)|}{h^*_t(\theta)} \right\|_2 < \infty.$$

The first two conditions mean that finiteness of fourth moments is assumed for the observed process y_t, which is much more than needed to prove consistency. As discussed by Francq and Zakoïan (2004) and Ling (2007a) in the linear ARMA–GARCH case, it is quite expected that finiteness of second moments of the observed process is required to make a suitable central limit theorem applicable to the score vector and, even in this linear case, it has proved difficult to do without assuming finite fourth moments. The moment conditions imposed on the derivatives of h^*_t are satisfied when the conditional mean is modeled by a linear function and conditional variance by a standard linear GARCH(1,1) model (see Francq and Zakoïan (2004) and Ling (2007a)). In our general nonlinear model it seems difficult to replace these conditions with something more explicit. However, as will be seen in Section 6, these conditions are satisfied in the nonlinear example we consider.

The assumptions made so far guarantee finiteness of the expectations

$$\mathcal{I}(\theta) \overset{def}{=} E\left[\frac{\partial L^*_T(\theta)}{\partial\theta}\frac{\partial L^*_T(\theta)}{\partial\theta'}\right] \quad \text{and} \quad \mathcal{J}(\theta) \overset{def}{=} E\left[\frac{\partial^2 L^*_T(\theta)}{\partial\theta\partial\theta'}\right]$$

[5] We note that only the moment and convergence results for $h^*_{\theta,t}$, but not those for $h^*_{\theta\theta,t}$, are explicitly used in the proofs that follow. The results for $h^*_{\theta\theta,t}$ are, however, required to justify the twice continuous partial differentiability of h^*_t on Θ_0 and the relation $\partial^2 h^*_t(\theta)/\partial\theta\partial\theta' = h^*_{\theta\theta,t}(\theta)$ although we have omitted the details of this in the proofs.

for $\theta \in \Theta_0$. Explicit expressions for these matrices are derived in Appendix D, Lemmas D.1 and D.2. If the matrices $\mathcal{I}(\theta_0)$ and $\mathcal{J}(\theta_0)$ are positive definite the asymptotic covariance matrix of the QML estimator $\hat{\theta}_T$ is also positive definite, as required for statistical inference. In order to guarantee this, we impose the following three conditions.

Assumption N5.

(i) *The distribution of ε_t is not concentrated at two points.*

(ii) $x_\mu' \frac{\partial f_t(\mu_0)}{\partial \mu} = 0$ *a.s. only if* $x_\mu = 0$ *($x_\mu \in \mathbb{R}^m$).*

(iii) $x_\lambda' \frac{\partial g(u_{0,t}, \sigma_t^2; \theta_0)}{\partial \lambda} = 0$ *a.s. only if* $x_\lambda = 0$ *($x_\lambda \in \mathbb{R}^l$).*

The third condition in Assumption N5 is similar to the one used by Straumann and Mikosch (2006, Assumption N.4) in the pure GARCH case, whereas the second one is its analogue for the conditional mean. These two conditions require the components of both $\partial f_t(\mu_0)/\partial \mu$ and $\partial g(u_{0,t}, \sigma_t^2; \theta_0)/\partial \lambda$ to be linearly independent with probability one. Due to the generality of our model these two conditions seem difficult to replace with more transparent counterparts. However, if the function f used to model the conditional mean is linear, the first condition is automatically satisfied given that N5(i) holds (or as long as ε_t is not degenerate; see Appendix E, Example 1). Moreover, if conditional heteroskedasticity is modeled by a standard linear GARCH(1,1) model and provided that homoskedasticity is ruled out, the second condition also holds given that N5(i) is satisfied (or as long as ε_t^2 is not degenerate; see Appendix E, Example 1). For a model containing both a conditional mean and a conditional variance, condition N5(i) appears to be the minimal requirement on the error term ε_t to ensure the positive definiteness of the asymptotic covariance matrix of the QML estimator $\hat{\theta}_T$. This condition was also used by Francq and Zakoïan (2004) in the context of their linear ARMA–GARCH model and, as they point out, is marginally stronger than the requirement that the random variable ε_t^2 is not degenerate (in the case ε_t has an asymmetric distribution). In the context of a nonlinear GARCH model, a condition at least as strong as N5(i) may often be needed to ensure that condition N5(iii) holds. We will return to this in the concrete examples of the next section, but already note that, for instance, Straumann and Mikosch (2006) need condition N5(i) when verifying their counterpart of N5(iii) (see the example in their Section 8).

Verifying conditions N5(ii) and N5(iii) for particular nonlinear models may be complicated. The technical difficulties are similar to those already discussed in connection with the verification of the identification conditions in Assumption C6, and we only mention that we have resorted to Markov chain techniques in order to be able to verify them. As far as we know, the only previous example of this kind of approach is Chan and Tong (1986, Appendix II) where Markov chain techniques are used to show the positive definiteness of the asymptotic covariance matrix of the nonlinear least squares estimator in a homoskedastic smooth transition autoregressive model.[6]

Now we can state the main result of this section.

Theorem 2. *Suppose Assumptions DGP, C1–C6, and N1–N5 hold. Then*

$$T^{1/2}(\hat{\theta}_T - \theta_0) \xrightarrow{d} N\left(0, \mathcal{J}(\theta_0)^{-1}\mathcal{I}(\theta_0)\,\mathcal{J}(\theta_0)^{-1}\right),$$

where the matrices $\mathcal{I}(\theta_0)$ and $\mathcal{J}(\theta_0)$ are given in (32) and (35) in Appendix D and are positive definite. Moreover, if the distribution of ε_t is symmetric, $\mathcal{I}(\theta_0)$ and $\mathcal{J}(\theta_0)$ can be expressed as

$$\mathcal{I}(\theta_0) = \begin{bmatrix} 4E\left[\frac{f_{\mu,t}(\mu_0)}{\sigma_t}\frac{f'_{\mu,t}(\mu_0)}{\sigma_t}\right] & 0_{m\times l} \\ 0_{l\times m} & 0_{l\times l} \end{bmatrix} + E\left[\varepsilon_t^4 - 1\right] E\left[\frac{h^*_{\theta,t}(\theta_0)}{\sigma_t^2}\frac{h^{*\prime}_{\theta,t}(\theta_0)}{\sigma_t^2}\right] \tag{13}$$

and

$$\mathcal{J}(\theta_0) = \begin{bmatrix} 2E\left[\frac{f_{\mu,t}(\mu_0)}{\sigma_t}\frac{f'_{\mu,t}(\mu_0)}{\sigma_t}\right] & 0_{m\times l} \\ 0_{l\times m} & 0_{l\times l} \end{bmatrix} + E\left[\frac{h^*_{\theta,t}(\theta_0)}{\sigma_t^2}\frac{h^{*\prime}_{\theta,t}(\theta_0)}{\sigma_t^2}\right]. \tag{14}$$

As in the consistency proof, we shall follow Berkes, Horváth, and Kokoszka (2003), Francq and Zakoïan (2004), and Straumann and Mikosch (2006) and first show that the infeasible QML estimator obtained by minimizing the function $L_T^*(\theta)$ has the limiting distribution stated in the theorem. After this intermediate step, the proof is completed by showing that the same limiting distribution applies to the corresponding feasible estimator $\hat{\theta}_T$ (for details, see Appendix D).

To compute approximate standard errors for the components of $\hat{\theta}_T$ and construct asymptotically valid Wald tests we need consistent estimators for the matrices $\mathcal{I}(\theta_0)$ and $\mathcal{J}(\theta_0)$. The expressions of these matrices in (32) and (35) in Appendix D reveal that it suffices to find consistent estimators for

$$E\left[\frac{f_{\mu,t}(\mu_0)}{\sigma_t}\frac{f'_{\mu,t}(\mu_0)}{\sigma_t}\right],\ E\left[\varepsilon_t^4 - 1\right],\ E\left[\frac{h^*_{\theta,t}(\theta_0)}{\sigma_t^2}\frac{h^{*\prime}_{\theta,t}(\theta_0)}{\sigma_t^2}\right],\ \text{and}\ E\left[\frac{f_{\mu,t}(\mu_0)}{\sigma_t}\frac{h^{*\prime}_{\theta,t}(\theta_0)}{\sigma_t^2}\right] \tag{15}$$

(in the case of a symmetric error distribution the fourth one is not required, as expressions (13) and (14) reveal). The obvious choices for these quantities are

$$T^{-1}\sum_{t=1}^T \frac{\hat{f}_{\mu,t}}{\hat{h}_t^{1/2}}\frac{\hat{f}'_{\mu,t}}{\hat{h}_t^{1/2}},\ \ T^{-1}\sum_{t=1}^T\left(\frac{\hat{u}_t^4}{\hat{h}_t^2} - 1\right),\ \ T^{-1}\sum_{t=1}^T \frac{\hat{h}_{\theta,t}}{\hat{h}_t}\frac{\hat{h}'_{\theta,t}}{\hat{h}_t},\ \text{and}\ T^{-1}\sum_{t=1}^T \frac{\hat{f}_{\mu,t}}{\hat{h}_t^{1/2}}\frac{\hat{h}'_{\theta,t}}{\hat{h}_t}, \tag{16}$$

[6] Tjøstheim (1986, Section 4.1) is able to verify his counterpart of condition N5(ii) in a very simple manner in a homoskedastic first order exponential autoregressive model.

respectively, where "$\hat{\ }$" signifies evaluation at the QML estimator $\hat{\theta}_T$. The obvious estimators of $\mathcal{I}(\theta_0)$ and $\mathcal{J}(\theta_0)$ obtained in this way are denoted by $\hat{\mathcal{I}}_T$ and $\hat{\mathcal{J}}_T$, respectively. It is shown in Appendix D that, under the conditions of Theorem 2,

$$\hat{\mathcal{I}}_T \to \mathcal{I}(\theta_0) \text{ a.s.} \quad \text{and} \quad \hat{\mathcal{J}}_T \to \mathcal{J}(\theta_0) \text{ a.s.} \tag{17}$$

Thus, a consistent estimator of the asymptotic covariance matrix $\mathcal{J}(\theta_0)^{-1}\mathcal{I}(\theta_0)\mathcal{J}(\theta_0)^{-1}$ in Theorem 2 is given by $\hat{\mathcal{J}}_T^{-1}\hat{\mathcal{I}}_T\hat{\mathcal{J}}_T^{-1}$. Finally, note that if ε_t is normally distributed (or, more generally, if $E\left[\varepsilon_t^4\right] = 3$ and $E\left[\varepsilon_t^3\right] = 0$), the relation $\mathcal{I}(\theta_0) = 2\mathcal{J}(\theta_0)$ obviously holds. Then the limiting distribution of $\hat{\theta}_T$ simplifies (see Theorem 2) which can accordingly be taken into account in the computation of standard errors and Wald test statistics.

6 Examples

We shall now consider concrete examples to which our general theory applies. In each case we give a set of low-level conditions that guarantee the validity of Assumptions DGP, C1–C6, and N1–N5. Verifying that the stated conditions imply these assumptions is postponed to Appendix E.

Example 1: Linear AR–GARCH. Consider the linear AR(p)–GARCH(1,1) model in which the conditional mean and conditional variance are given by

$$f(y_{t-1}, \ldots, y_{t-p}; \mu_0) = \phi_{0,0} + \sum_{j=1}^{p} \phi_{0,j} y_{t-j} \quad \text{and} \quad \sigma_t^2 = g\left(u_{0,t-1}, \sigma_{t-1}^2; \theta_0\right) = \omega_0 + \alpha_0 u_{0,t-1}^2 + \beta_0 \sigma_{t-1}^2,$$

respectively, where $u_{0,t} = y_t - (\phi_{0,0} + \sum_{j=1}^{p} \phi_{0,j} y_{t-j}) = \sigma_t \varepsilon_t$ and ε_t is a sequence of independent and identically distributed random variables with $E[\varepsilon_t] = 0$ and $E[\varepsilon_t^2] = 1$. The parameter vectors μ and λ are given by $\mu = (\phi_0, \ldots, \phi_p)$ and $\lambda = (\omega, \alpha, \beta)$, respectively. These parameters take values in the permissible parameter spaces M and Λ that are compact subsets of $\mathbb{R}^{p+1}$ and $(0, \infty) \times [0, \infty) \times [0, 1)$ containing the true parameter vectors μ_0 and λ_0, respectively. Note that our definition of the parameter space includes the restriction that $\beta < 1$ over Θ.

Consider the following set of conditions.

(a) **(i)** $E\left[\ln(\beta_0 + \alpha_0 \varepsilon_t^2)\right] < 0$

(ii) $1 - \sum_{j=1}^{p} \phi_{0,j} z^j \neq 0$, $|z| \leq 1$

(b) **(i)** ε_t^2 has a non-degenerate distribution

(ii) $\alpha_0 > 0$

(c) **(i)** The true parameter value θ_0 is an interior point of Θ

(ii) $E\big[(\beta_0 + \alpha_0 \varepsilon_t^2)^2\big] = \beta_0^2 + 2\alpha_0\beta_0 + \alpha_0^2 E[\varepsilon_t^4] < 1$

(iii) The distribution of ε_t is not concentrated at two points

The two conditions in part (a) imply the validity of Assumption DGP for the linear AR(p)–GARCH(1,1) model as defined above (for details of this and the following statements, see Appendix E). The former condition agrees with the necessary and sufficient condition for the (strict) stationarity and geometric ergodicity of the conditional variance process obtained in Nelson (1990) and Francq and Zakoïan (2006), respectively. The latter is necessary and sufficient for the existence of a strictly stationary causal solution to a conventional linear AR(p) model. If the conditions in part (b) are also assumed, Assumptions C1–C6 hold. The conditions in (b) are needed to ensure the identifiability of the parameters in the conditional variance part. Finally, conditions in (a)–(c) (where (b.i) becomes unnecessary) suffice for Assumptions N1–N5 to hold. Condition (c.i) is obviously required for asymptotic normality of the parameter estimator to hold. The second condition, which implicitly includes the requirement that $E[\varepsilon_t^4] < \infty$, ensures that the conditional variance process, and hence also y_t, has finite fourth moments. This is much more than is needed for asymptotic normality of the QML estimator in the pure GARCH case but, as already discussed, appears difficult to avoid in the AR–GARCH case. Finally, condition (c.iii), which is slightly stronger than (b.i), is needed for the identification condition N5 to hold.

We note that our conditions (a)–(c) (almost) coincide with those required in Francq and Zakoïan (2004) for strong consistency and asymptotic normality of the QML estimator in the case of a linear AR(p)–GARCH(1,1) model.[7] Therefore, although our framework allows for rather general forms of nonlinearity, it does not come at the cost of assumptions that would be stronger than those required in the linear case in earlier literature. We refer to Francq and Zakoïan (2004) for a discussion of previous, more stringent, assumptions used in QML estimation of linear GARCH and ARMA–GARCH models.

Example 2: AR–AGARCH. As a second example, we consider a model in which a linear AR(p) model is combined with the Asymmetric GARCH (AGARCH) model of Ding, Granger, and Engle (1993). For this model we are able to show strong consistency, but not asymptotic

[7]There appears to be only one small difference. In their condition A8 restricting the conditional mean, Francq and Zakoïan (2004) assume that the roots of the autoregressive polynomial are outside the unit circle for all $\theta \in \Theta$, whereas our condition (a.ii) requires this only for the true parameter value θ_0. However, inspecting their proofs it would seem that this stronger requirement is actually not used. In this sense, our conditions appear to coincide with theirs.

normality, of the QML estimator. The set-up is otherwise exactly the same as in Example 1, except that now the conditional variance process is defined as

$$\sigma_t^2 = g\left(u_{0,t-1}, \sigma_{t-1}^2; \theta_0\right) = \omega_0 + \alpha_0(|u_{0,t-1}| - \gamma_0 u_{0,t-1})^2 + \beta_0\sigma_{t-1}^2, \tag{18}$$

and the parameter vector λ defined as $\lambda = (\omega, \alpha, \beta, \gamma)$ with the permissible parameter space Λ a compact subset of $(0, \infty) \times [0, \infty) \times [0, 1) \times [-1, 1]$ containing the true parameter vector λ_0. Note that, letting $1(\cdot)$ stand for the indicator function, (18) can be rewritten as

$$\sigma_t^2 = \omega_0 + \alpha_0(1-\gamma_0)^2 u_{0,t-1}^2 1(u_{0,t-1} \geq 0) + \alpha_0(1+\gamma_0)^2 u_{0,t-1}^2 1(u_{0,t-1} < 0) + \beta_0\sigma_{t-1}^2,$$

so that the threshold GARCH formulations of Glosten, Jaganathan, and Runkle (1993) and Zakoïan (1994) are included in the AGARCH model.

Consider the following set of conditions.

(a) **(i)** $E\left[\ln\left(\beta_0 + \alpha_0(|\varepsilon_t| - \gamma_0\varepsilon_t)^2\right)\right] < 0$

(ii) $1 - \sum_{j=1}^{p} \phi_{0,j} z^j \neq 0$, $|z| \leq 1$

(b) **(i)** The distribution of ε_t is not concentrated at two points

(ii) $\alpha_0 > 0$

Conditions (a.i) and (a.ii) ensure the validity of Assumption DGP for the AR–AGARCH model. Condition (a.i) agrees with the necessary and sufficient condition for the (strict) stationarity and geometric ergodicity of the conditional variance process obtained in Straumann and Mikosch (2006, Theorem 3.5) and Meitz and Saikkonen (2008a, Example 1), respectively. Altogether the conditions in (a) and (b) ensure that Assumptions C1–C6 hold. Note that the restriction $-1 \leq \gamma \leq 1$ imposed on the parameter γ and the slightly stronger condition (b.i) compared to Example 1 are needed to verify the identification condition in C6(ii).

In this example, we are unable to show the asymptotic normality of the QML estimator. This is due to the appearance of $|u_{0,t}|$ in the equation defining the conditional variance, which, as can readily be verified, invalidates Assumption N2 requiring the function g to be twice continuously differentiable with respect to all its arguments. A similar complication occurs in several other nonlinear GARCH models that involve absolute values. In the pure AGARCH model the situation simplifies because $u_{0,t} = y_t$ contains no parameters and therefore differentiability of g with respect to u is not required. In this case the asymptotic normality of the QML estimator is proved by Straumann and Mikosch (2006).

Example 3: Nonlinear AR–GARCH. As a third example we consider a model in which both the conditional mean and the conditional variance are nonlinear. We model the conditional mean by a fairly general subclass of the functional-coefficient autoregressive models of Chen and Tsay (1993). The best known special case to which our results apply is the logistic smooth transition autoregressive specification considered by Teräsvirta (1994). For the conditional variance, we consider a smooth transition GARCH model similar to those discussed by González-Rivera (1998) and Lundbergh and Teräsvirta (2002). The resulting nonlinear AR–GARCH model is a special case of the one considered by Meitz and Saikkonen (2008b) whose results on geometric ergodicity we can apply. Using similar arguments other models of interest could also be considered. For instance, the nonlinearity in the conditional expectation might be of the exponential autoregressive type of Haggan and Ozaki (1981) or the smooth transition in the conditional variance might be of the type considered by Lanne and Saikkonen (2005).

In the nonlinear AR(p)–GARCH(1,1) model we consider the conditional mean and conditional variance are given by

$$f\left(y_{t-1},\ldots,y_{t-p};\mu_0\right)=\phi_{0,0}+\psi_{0,0}F(y_{t-d};\varphi_0)+\sum_{j=1}^{p}\left(\phi_{0,j}+\psi_{0,j}F(y_{t-d};\varphi_0)\right)y_{t-j},$$

and

$$\sigma_t^2=g\left(u_{0,t-1},\sigma_{t-1}^2;\theta_0\right)=\omega_0+\left(\alpha_{0,1}+\alpha_{0,2}G(u_{0,t-1};\gamma_0)\right)u_{0,t-1}^2+\beta_0\sigma_{t-1}^2, \tag{19}$$

respectively, where $u_{0,t}=y_t-f\left(y_{t-1},\ldots,y_{t-p};\mu_0\right)=\sigma_t\varepsilon_t$, ε_t is a sequence of independent and identically distributed random variables with $E[\varepsilon_t]=0$ and $E[\varepsilon_t^2]=1$, and $\varphi_0=(\varphi_{0,1},\varphi_{0,2})$ and $\gamma_0=(\gamma_{0,1},\gamma_{0,2})$. The parameter vectors μ and λ are $\mu=(\phi_0,\ldots,\phi_p,\psi_0,\ldots,\psi_p,\varphi_1,\varphi_2)$ and $\lambda=(\omega,\alpha_1,\alpha_2,\beta,\gamma_1,\gamma_2)$ and the permissible parameter spaces M and Λ are compact subsets of $\mathbb{R}^{2p+3}\times\mathbb{R}_+$ and $\mathbb{R}_+\times[0,\infty)^2\times[0,1)\times\mathbb{R}\times\mathbb{R}_+$ containing the true parameter vectors μ_0 and λ_0, respectively. In both $\varphi=(\varphi_1,\varphi_2)$ and $\gamma=(\gamma_1,\gamma_2)$, the first parameter is supposed to have the role of a location parameter so that it takes values in $\mathbb{R}$, whereas the latter parameter is a scale parameter and hence is restricted to be positive (these restrictions and interpretations are done only for concreteness and are not necessary for the development of the theory). The nonlinear functions F and G are assumed to take values in $[0,1]$. The former depends on the lagged observable y_{t-d}, where d is a fixed known integer between 1 and p (which is not estimated), whereas the latter depends on u_{t-1}.

For clarity of exposition, we concentrate on the case of F and G being cumulative distribution functions of the logistic distribution, that is,

$$F\left(y;\varphi_1,\varphi_2\right)=\left[1+\exp\left(-\varphi_2(y-\varphi_1)\right)\right]^{-1}\quad\text{and}\quad G\left(u;\gamma_1,\gamma_2\right)=\left[1+\exp\left(-\gamma_2(u-\gamma_1)\right)\right]^{-1},$$

although our results also hold much more generally. This is also one of the most common choices in practice. In Appendix E we give a set of conditions for the functions F and G that suffice for our results to hold. It is straigtforward to verify that these conditions are satisfied with the choice of logistic functions (or, for example, the cumulative distribution functions of the normal distribution). In the following we assume that the functions F and G satisfy the additional conditions given in Appendix E.

To present the conditions for this model we require some additional notation. For $p = 1$, define $A_{01} = \phi_{0,1}$ and $A_{02} = \phi_{0,1} + \psi_{0,1}$, and for $p > 1$ define A_{01} and A_{02} as the $p \times p$ matrices

$$A_{01} = \begin{bmatrix} \phi_{0,1} \;\cdots\; \phi_{0,p-1} & \phi_{0,p} \\ I_{p-1} & 0_{p-1} \end{bmatrix} \quad \text{and} \quad A_{02} = \begin{bmatrix} \phi_{0,1} + \psi_{0,1} \;\cdots\; \phi_{0,p-1} + \psi_{0,p-1} & \phi_{0,p} + \psi_{0,p} \\ I_{p-1} & 0_{p-1} \end{bmatrix},$$

where I_{p-1} and 0_{p-1} denote the $(p-1) \times (p-1)$ identity matrix and a $(p-1) \times 1$ vector of zeros, respectively. We also need the concept of joint spectral radius defined for a set of bounded square matrices $\mathcal{A}$ by

$$\rho(\mathcal{A}) = \limsup_{k\to\infty} \left(\sup_{A \in \mathcal{A}^k} \|A\| \right)^{1/k},$$

where $\mathcal{A}^k = \{A_1 A_2 \cdots A_k : A_i \in \mathcal{A},\ i = 1, \ldots, k\}$ and $\|\cdot\|$ can be any matrix norm (the value of $\rho(\mathcal{A})$ does not depend on the choice of this norm). If the set $\mathcal{A}$ only contains a single matrix A then the joint spectral radius of $\mathcal{A}$ coincides with $\rho(A)$, the spectral radius of A. Several useful results about the joint spectral radius are given in the recent paper by Liebscher (2005) where further references can also be found; see also Meitz and Saikkonen (2008b).

Now consider the following set of conditions.

(a) **(i)** The ε_t have a (Lebesgue) density which is positive and lower semicontinuous on $\mathbb{R}$

(ii) Either $\sum_{j=1}^{p} \max\{|\phi_{0,j}|, |\phi_{0,j} + \psi_{0,j}|\} < 1$ or $\rho(\{A_{01}, A_{02}\}) < 1$

(iii) $E\big[\log\big(\beta_0 + (\alpha_{0,1} + \alpha_{0,2})\varepsilon_t^2\big)\big] < 0$

(iv) $\alpha_{0,1} > 0$ and $\beta_0 > 0$

(b) **(i)** At least one of the $\psi_{0,j}$, $j = 0, \ldots, p$, is nonzero

(ii) $\alpha_{0,2} > 0$

(c) **(i)** The true parameter value θ_0 is an interior point of Θ

(ii) $E\big[\big(\beta_0 + (\alpha_{0,1} + \alpha_{0,2})\varepsilon_t^2\big)^2\big] = \beta_0^2 + 2(\alpha_{0,1} + \alpha_{0,2})\beta_0 + (\alpha_{0,1} + \alpha_{0,2})^2 E[\varepsilon_t^4] < 1$

Conditions (a.i)–(a.iv) ensure the validity of Assumption DGP in the case of the considered nonlinear AR–GARCH model. Condition (a.i) restricts the error term more than required in

Examples 1 and 2, but this is needed to verify Assumption DGP with the results of Meitz and Saikkonen (2008b). Condition (a.i) also facilitates the verification of the identification conditions in Assumptions C6 and N5. As our discussion following Assumption C6 indicated, this is now a considerably more complicated task than in the preceding examples and involves using Markov chain techniques to show that the events $\{(y_{t-1}, ..., y_{t-p}) \in A\}$ have a positive probability with suitably defined (Borel) measurable sets $A \subset \mathbb{R}^p$. Condition (a.i) will be critical in establishing this. The two alternative conditions in (a.ii) are both sufficient restrictions on the conditional mean needed to show the validity of Assumption DGP. They are used in Meitz and Saikkonen (2008b, Section 4) and, as discussed by Liebscher (2005, p. 682), the latter condition is strictly weaker than the former one. Condition (a.iii) is an analogue of the moment conditions (a.i) in the previous two examples, and it also coincides with the sufficient condition for geometric ergodicity of a pure smooth transition GARCH model given in Example 4 of Meitz and Saikkonen (2008a). Condition (a.iv) excludes the ARCH case, but is required for the results in Meitz and Saikkonen (2008b) to hold. In many applications the estimate of β would typically be rather large (and close to unity), and hence condition (a.iv) is not very restrictive in practice.[8]

If conditions (b.i) and (b.ii) are also assumed, Assumptions C1–C6 hold. These two conditions are required to identify the parameters of the model. Finally, the additional conditions (c.i) and (c.ii) ensure that Assumptions N1–N5 also hold and are completely analogous to (c.i) and (c.ii) in Example 1.

Above we assumed that the function G is strictly increasing and the value of the parameter $\alpha_{0,2}$ is positive, in which case the coefficient of $u_{0,t-1}^2$ in (19) increases with $u_{0,t-1}$. Often, an empirically interesting case is the one in which the effect is in the opposite direction. This case is obtained by choosing G to be strictly decreasing (in the preceding logistic example the permissible parameter space of γ_2 is then a compact subset of $(-\infty, 0)$ instead of $(0, \infty)$). Our results also apply to this case (with minor changes to the derivations; see Appendix E).

7 Conclusion

In this paper we have developed an asymptotic estimation theory for nonlinear functional-coefficient AR(p) models with conditionally heteroskedastic errors specified as a general nonlinear GARCH(1,1) model. We proved strong consistency and asymptotic normality of the QML estimator under conditions similar to those previously employed in linear ARMA–GARCH mod-

[8]The ARCH case could be treated separately as is also mentioned in Meitz and Saikkonen (2008b, p. 465). For brevity, we do not pursue this further and only mention that in this case many of the required derivations would simplify considerably.

els. In particular, for consistency only a mild moment condition was required, whereas existence of fourth order moments of the observed process was needed for asymptotic normality. To the best of our knowledge, our paper is the first one to derive asymptotic estimation theory for a model allowing for nonlinearity in both the conditional mean and in the GARCH-type conditional variance.

Because our specification for the conditional variance was restricted to a GARCH(1,1) model it would be of interest to replace it by a higher order GARCH model. Relaxing our assumptions is another topic for potential future work. In particular, it would be useful if asymptotic normality could be established without the assumption of finite fourth order moments. As far as QML estimators are concerned, this has turned out to be difficult even in the linear case where weighted QML estimators have been developed as alternatives (see Ling (2007a) and the discussion therein). Another interesting extension would be to relax our assumption about the differentiability of the conditional variance function, and thereby make it possible to obtain asymptotic normality of the QML estimator also for the type of models discussed in our Example 2. Furthermore, our assumptions about permitted nonlinearity in the GARCH-part were more stringent than those needed to obtain stationarity and ergodicity of the data generation process so that relaxing these assumptions would be of interest.

Appendix A: Auxiliary results

We shall first give two simple lemmas which are useful in several subsequent proofs.

Lemma A.1. *For any* $r > 0$, $\left\|\sum_{i=1}^{k} x_i\right\|_r \leq \Delta_{r,k} \sum_{i=1}^{k} \|x_i\|_r$, *where* $\Delta_{r,k} = \max\{1, k^{1/r-1}\}$.

Proof. The case $r \geq 1$ follows from Minkowski's inequality. When $0 < r < 1$, Loève's c_r–inequality (see Davidson (1994), p. 140) first applied with r and then with $1/r$ yields

$$\left(E\left|\sum_{i=1}^{k} x_i\right|^r\right)^{1/r} \leq c_r^{1/r}\left(\sum_{i=1}^{k} E\left|x_i\right|^r\right)^{1/r} \leq c_r^{1/r} c_{1/r} \sum_{i=1}^{k}\left(E\left|x_i\right|^r\right)^{1/r},$$

where $c_r^{1/r} = 1$ and $c_{1/r} = k^{1/r-1}$. Hence the result. ■

Lemma A.2. *Suppose for some* $r > 0$, $\gamma > 1$, *and nonnegative process* x_t, $\gamma^t x_t$ *converges to zero in* L_r*–norm. Then* $\sum_{t=1}^{\infty} x_t < \infty$ *a.s. and* $\left\|\sum_{t=1}^{\infty} x_t\right\|_r < \infty$ *also holds.*

Proof. By the Borel-Cantelli lemma, the first result follows if we show that $\sum_{t=1}^{\infty} P(x_t > \delta^t) < \infty$ for some $\delta \in (0, 1)$. By assumption, $\gamma^t \|x_t\|_r \to 0$, and hence we can find a $C < \infty$ such that $\|x_t\|_r \leq C\gamma^{-t}$ for all $t \in \mathbb{N}$. Hence $E[x_t^r] \leq C^r \gamma^{-tr}$ for all $t \in \mathbb{N}$. Choose a δ such that

$\gamma^{-1} < \delta < 1$. Then $(\delta\gamma)^{-r} < 1$, and

$$\sum_{t=1}^{\infty} P(x_t > \delta^t) \leq \sum_{t=1}^{\infty} \delta^{-tr} E\left[x_t^r\right] \leq C^r \sum_{t=1}^{\infty} (\delta\gamma)^{-tr} < \infty.$$

This proves the former result. When $r \geq 1$ the latter result follows from the aforementioned inequality $\|x_t\|_r \leq C\gamma^{-t}$ by using Minkowski's inequality and monotone convergence. When $r < 1$ the same conclusion is obtained by using Loeve's c_r–inequality (see Davidson (1994), p. 140) instead of Minkowski's inequality (cf. the proof of Lemma A.1). ■

The following lemma presents a result which is similar to Theorem 3.1 of Bougerol (1993) and Theorem 2.8 of Straumann and Mikosch (2006). Its formulation involves a function $G : M_v \times M_z \times K \to M_z$ where M_v, M_z, and K are subsets of Euclidean spaces and K is compact. The function G is assumed to satisfy the following condition.

Condition G (i) For all $\vartheta \in K$, $|G(v, z; \vartheta)| \leq \bar{\varrho}\,|z| + \psi\left(|v|\right)$, where $0 < \bar{\varrho} < 1$ is a constant and $\psi : [0, \infty) \to [0, \infty)$ a measurable function.
(ii) The function $G(\cdot, \cdot; \cdot)$ is continuous and, for all $(v, \vartheta) \in M_v \times K$, $|G(v, z_1; \vartheta) - G(v, z_2; \vartheta)| \leq \bar{\kappa}\,|z_1 - z_2|$ for some $0 < \bar{\kappa} < 1$ and all z_1, $z_2 \in M_z$.

By $\mathbb{C}\left(K, M_z\right)$ we denote the Banach space of continuous functions from K into M_z endowed with the supremum norm $|\cdot|_K$, that is, $|z|_K = \sup_{\vartheta\in K}|z(\vartheta)|$.

Lemma A.3. *Let Condition G hold. Then, for all $\vartheta \in K$, there exists a stationary and ergodic solution $z_t^*\left(\vartheta\right)$ to the equation*

$$z_t\left(\vartheta\right) = G\left(v_{t-1}\left(\vartheta\right), z_{t-1}\left(\vartheta\right); \vartheta\right), \quad t = 1, 2, ..., \tag{20}$$

where z_0 is a random function taking values in $\mathbb{C}\left(K, M_z\right)$ and v_t is a stationary and ergodic process taking values in $\mathbb{C}\left(K, M_v\right)$ and satisfying $E[\sup_{\vartheta\in K} \psi\left(|v_t\left(\vartheta\right)|\right)^r] < \infty$, $r > 0$. The solution $z_t^*\left(\vartheta\right)$ is continuous in ϑ, measurable with respect to the σ–algebra generated by $\left(v_{t-1}\left(\vartheta\right), v_{t-2}\left(\vartheta\right), ...\right)$, and it is unique when (20) is extended to all $t \in \mathbb{Z}$. Moreover, $E[\sup_{\vartheta\in K}|z_t^*(\vartheta)|^r] < \infty$ and, if $z_t(\vartheta)$, $\vartheta \in K$, are any other solutions to (20) with $E[\sup_{\vartheta\in K}|z_0(\vartheta)|^r] < \infty$, then for a finite constant C (depending on r and the distribution of z_0),

$$\left\|\sup_{\vartheta\in K}|z_t^*(\vartheta) - z_t(\vartheta)|\right\|_r \leq C\bar{\kappa}^t.$$

Compared to Bougerol (1993, Theorem 3.1) and Straumann and Mikosch (2006, Theorem 2.8), Lemma A.3 is more specific although sufficient for the purpose of this paper. Due to its specificity its application in subsequent proofs also appears to lead to less complex derivations.

Another difference to the abovementioned theorems is that Lemma A.3 also implies the existence of certain moments, which turns out to be useful. In particular, because the stationary solution z_t^* obtained from Lemma A.3 is an element of $\mathbb{C}(K, M_z)$, Theorem 2.7 of Straumann and Mikosch (2006) immediately gives the result

$$\sup_{\vartheta \in K} \left| T^{-1} \sum_{t=1}^{T} z_t^*(\vartheta) - E[z_t^*(\vartheta)] \right| \to 0 \quad \text{a.s.}$$

when $r \geq 1$.

Proof of Lemma A.3. We apply Theorem 3.1 of Bougerol (1993) (see also Theorem 2.8 of Straumann and Mikosch (2006)). Define the random function $G_t : \mathbb{C}(K, M_z) \to \mathbb{C}(K, M_z)$ as $[G_t(x)](\vartheta) = G(v_{t-1}(\vartheta), x(\vartheta); \vartheta)$ $(x \in \mathbb{C}(K, M_z),\ \vartheta \in K)$. Then G_t, $t \in \mathbb{Z}$, is a stationary and ergodic sequence of mappings. By the continuity assumption in Condition G(ii) and the fact that z_0 belongs to $\mathbb{C}(K, M_z)$, the function $z_t(\cdot)$ defined by equation (20) is in $\mathbb{C}(K, M_z)$ and is a solution to the difference equation $x_t = G_t(x_{t-1})$, $t \geq 1$. Define

$$\rho(G_t) = \sup \left\{ \frac{|G_t(x_1) - G_t(x_2)|_K}{|x_1 - x_2|_K};\ x_1, x_2 \in \mathbb{C}(K, M_z),\ x_1 \neq x_2 \right\}$$

and notice that, due to our Lipschitz condition in Condition G(ii),

$$\begin{aligned} |G_t(x_1) - G_t(x_2)|_K &= \sup_{\vartheta \in K} |G(v_{t-1}(\vartheta), x_1(\vartheta); \vartheta) - G(v_{t-1}(\vartheta), x_2(\vartheta); \vartheta)| \\ &\leq \bar{\kappa} \sup_{\vartheta \in K} |x_1(\vartheta) - x_2(\vartheta)| \\ &= \bar{\kappa} |x_1 - x_2|_K . \end{aligned}$$

Thus, $\rho(G_t)$ is a stationary and ergodic process bounded from above by $\bar{\kappa} < 1$.

Now consider Theorem 3.1 of Bougerol (1993), and note that its assumptions (C1) and (C2) are satisfied due to the assumptions imposed. Specifically, by Condition G(i), the moment condition imposed on $\psi(|v_t|)$, and Lemma A.1, $|G_1(x) - x|_K^r$ has finite expectation for any $x \in \mathbb{C}(K, M_z)$, and thus (C1) holds by Jensen's inequality. Regarding (C2), it holds (with $p = 1$) because $\rho(G_t)$ is bounded from above by $\bar{\kappa} < 1$. The existence of a stationary ergodic solution $z_t^* \in \mathbb{C}(K, M_z)$ to (20) now follows from this theorem whereas the stated uniqueness can be obtained from Remark 2.9(2) of Straumann and Mikosch (2006). Defining $z_{t,n}(x) = (G_t \circ \cdots \circ G_{t-n})(x)$ with $n \geq 0$ and a fixed $x \in \mathbb{C}(K, M_z)$ as the backward iterates obtained by repetitive application of the random function G_t, we also find from the aforementioned papers that z_t^* can be defined as the (almost sure) limit $z_t^* = \lim_{n\to\infty} z_{t,n}(x)$ (with any fixed $x \in \mathbb{C}(K, M_z)$). Hence, $z_t^*(\vartheta)$ is measurable with respect to the σ–algebra generated by $(v_{t-1}(\vartheta), v_{t-2}(\vartheta), ...)$ (cf. Proposition 2.6 of Straumann and Mikosch (2006)).

As for the remaining assertions, fix $x \in \mathbb{C}(K, M_z)$ and use Condition G(i) to obtain

$$\begin{aligned}
|[z_{t,n}(x)](\vartheta)| &= |G(v_{t-1}(\vartheta), [(G_{t-1} \circ \cdots \circ G_{t-n})(x)](\vartheta); \vartheta)| \\
&\leq \bar{\varrho}\, |[(G_{t-1} \circ \cdots \circ G_{t-n})(x)](\vartheta)| + \psi(|v_{t-1}(\vartheta)|) \\
&= \bar{\varrho}\, |[z_{t-1,n-1}(x)](\vartheta)| + \psi(|v_{t-1}(\vartheta)|)
\end{aligned}$$

and, continuing iteratively,

$$|[z_{t,n}(x)](\vartheta)| \leq \bar{\varrho}^n\, |[z_{t-n,0}(x)](\vartheta)| + \sum_{j=0}^{n-1} \bar{\varrho}^j \psi(|v_{t-j-1}(\vartheta)|).$$

Here

$$|[z_{t-n,0}(x)](\vartheta)| = |[G_{t-n}(x)](\vartheta)| = |G(v_{t-n-1}(\vartheta), x(\vartheta); \vartheta)| \leq \bar{\varrho}\, |x(\vartheta)| + \psi(|v_{t-n-1}(\vartheta)|),$$

where the inequality is again due to Condition G(i). Because the preceding inequalities hold for all $\vartheta \in K$, we have

$$\begin{aligned}
|z_{t,n}(x)|_K &\leq \bar{\varrho}^{n+1}\, |x|_K + \sum_{j=0}^{n} \bar{\varrho}^j \sup_{\vartheta \in K} \psi(|v_{t-j-1}(\vartheta)|) \\
&\leq |x|_K + \sum_{j=0}^{\infty} \bar{\varrho}^j \sup_{\vartheta \in K} \psi(|v_{t-j-1}(\vartheta)|).
\end{aligned}$$

Denote the stationary process defined by the last expression by w_t. By Lemma A.2, this process is well defined because the series converges a.s. and, furthermore, $E[|w_t|^r] < \infty$ where Lemma A.1 is also made use of. Hence, we can conclude that the collection of random variables $\{|z_{t,n}(x)|_K^r,\ n = 1, 2, ...\}$ is uniformly integrable (see Billingsley (1995, p. 338)). Thus, because $z_t^* = \lim_{n\to\infty} z_{t,n}(x)$ (in $\mathbb{C}(K, M_z)$) we also have $\lim_{n\to\infty} |z_{t,n}|_K^r = |z_t^*|_K^r$ and the above mentioned uniform integrability allows us to conclude that $E[|z_t^*|_K^r]$ ($= E[\sup_{\vartheta\in K} |z_t^*(\vartheta)|^r]$) is the finite limit of $E[|z_{t,n}(x)|_K^r]$ (see Davidson (1994), Theorem 12.8).

Now consider the last assertion. Using Condition G(ii),

$$\begin{aligned}
|z_t^* - z_t|_K^r &= \sup_{\vartheta \in K} \left|G(v_{t-1}(\vartheta), z_{t-1}^*(\vartheta); \vartheta) - G(v_{t-1}(\vartheta), z_{t-1}(\vartheta); \vartheta)\right|^r \\
&\leq \bar{\kappa}^r \sup_{\vartheta \in K} \left|z_{t-1}^*(\vartheta) - z_{t-1}(\vartheta)\right|^r \\
&= \bar{\kappa}^r \left|z_{t-1}^* - z_{t-1}\right|_K^r.
\end{aligned}$$

Continuing iteratively,

$$|z_t^* - z_t|_K^r \leq \bar{\kappa}^{rt}\, |z_0^* - z_0|_K^r \leq \bar{\kappa}^{rt} \max\{1, 2^{r-1}\}\, (|z_0^*|_K^r + |z_0|_K^r),$$

where the second inequality follows from Lemma A.1. Because the two norms in the last expression have finite expectations the stated inequality follows. ■

Appendix B: Proofs for Sections 2 and 3

Proof of Proposition 1. We apply Lemma A.3. Specifically, choosing $M_v = \mathbb{R}$, $M_z = \mathbb{R}_+$, $K = \Theta$, $G = g$, $v_t = u_t = y_t - f(y_{t-1}, \ldots, y_{t-p}; \mu)$, and $z_t(\theta) = h_t(\theta) = g(u_{t-1}(\theta), h_{t-1}(\theta); \theta)$, it follows from Assumption C2 that Conditions G(i) and (ii) are satisfied with the function $\psi(x) = \varkappa x^2 + \varpi$. Furthermore, by the definition of the function f and Assumption C3,

$$|u_t| = |y_t - f(y_{t-1}, \ldots, y_{t-p}; \mu)| \leq |y_t| + C \sum_{j=1}^{p} |y_{t-j}| + C$$

for some finite constant C. Thus, Assumption DGP and Lemma A.1 give $\left\| \sup_{\theta \in \Theta} |u_t| \right\|_{2r} < \infty$, implying the moment condition $E[\sup_{\vartheta \in K} \psi(|v_t(\vartheta)|)^r] < \infty$. The stated result, except for the equality $h_t^*(\theta_0) = \sigma_t^2$, now follows from Lemma A.3 (note that the solution $h_t^*(\theta)$ is initialized from $h_0^*(\theta)$ having this stationary distribution instead of the constant ς_0). From the proof of this lemma it is also seen that h_t^* can be defined as the (almost sure) limit $h_t^* = \lim_{n \to \infty} h_{t,n}$, where $h_{t,n} = (g_t \circ \cdots \circ g_{t-n})(x)$, $n \geq 0$, are the backward iterates obtained by repetitive application of the random function $[g_t(x)](\theta) = g(u_{t-1}(\theta), x(\theta); \theta)$ with a fixed $x \in \mathbb{C}(\Theta, \mathbb{R}_+)$. To prove that $h_t^*(\theta_0) = \sigma_t^2$ (cf. Propositions 3.7 and 3.12 in Straumann and Mikosch (2006)), note that $h_t^*(\theta_0) = \lim_{n \to \infty} h_{t,n}(\theta_0)$ a.s. where $h_{t,n}(\theta_0) = [(g_t \circ \cdots \circ g_{t-n})(x)](\theta_0)$ and $[g_t(x)](\theta_0) = g(u_{0,t-1}, x(\theta_0); \theta_0)$. By Assumption DGP and the definition of $h_{t,n}(\theta_0)$, $(h_{t,n}(\theta_0), \sigma_t^2)$ is stationary for every fixed n, and hence $h_{t,n}(\theta_0) - \sigma_t^2$ and $h_{n,n}(\theta_0) - \sigma_n^2$ are identically distributed. Regarding the latter, repeated use of Assumption C2(ii) yields $\left| h_{n,n}(\theta_0) - \sigma_n^2 \right| \leq \kappa^n \left| h_{0,0}(\theta_0) - \sigma_0^2 \right|$, where $\left| h_{0,0}(\theta_0) - \sigma_0^2 \right| = \left| g(u_{0,-1}, x(\theta_0); \theta_0) - \sigma_0^2 \right| \leq \varrho x(\theta_0) + \varkappa u_{0,-1}^2 + \varpi + \sigma_0^2$ by Assumption C2(i). Making use of Assumption DGP, the result $\left\| \sup_{\theta \in \Theta} |u_t| \right\|_{2r} < \infty$ obtained above, and Lemma A.1, $\left\| h_{n,n}(\theta_0) - \sigma_n^2 \right\|_r \leq C\kappa^n$ for all $n \geq 0$ and for some finite C. Because $h_{t,n}(\theta_0) - \sigma_t^2$ and $h_{n,n}(\theta_0) - \sigma_n^2$ are identically distributed, $\left\| h_{t,n}(\theta_0) - \sigma_t^2 \right\|_r \leq C\kappa^n$ and, using Lemma A.2, we can conclude that $\lim_{n \to \infty}(h_{t,n}(\theta_0) - \sigma_t^2) = 0$ a.s. As noticed above, $h_t^*(\theta_0) = \lim_{n \to \infty} h_{t,n}(\theta_0)$ a.s., and hence $h_t^*(\theta_0) - \sigma_t^2 = 0$ a.s.

Finally, note also that from Lemma A.3 we obtain the inequality

$$\left\| \sup_{\theta \in \Theta} |h_t^* - h_t| \right\|_r \leq C\kappa^t, \tag{21}$$

for some finite constant C, a result that will repeatedly be used in the proofs. ■

Proof of Theorem 1. For strong consistency of the estimator $\hat{\theta}_T$ it suffices to show that, for every $\delta > 0$,

$$\liminf_{T \to \infty} \inf_{\theta \in B(\theta_0, \delta)^c} \left(L_T(\theta) - L_T(\theta_0) \right) > 0 \quad \text{a.s.},$$

where $B(\theta_0, \delta) = \{\theta \in \Theta : |\theta - \theta_0| < \delta\}$ and $B(\theta_0, \delta)^c$ is the complement of this set in Θ (see, e.g., Pötscher and Prucha (1991a, p. 145)). To this end, first recall that $l_t^*(\theta)$ and $l_t(\theta)$ denote the summands of $L_T^*(\theta)$ and $L_T(\theta)$, respectively. It will be seen below that $E[l_t^*(\theta)]$ is well defined taking values in $\mathbb{R} \cup \{+\infty\}$ but $E[l_t^*(\theta_0)] < \infty$. Next note that

$$\begin{aligned}\liminf_{T\to\infty} \inf_{\theta\in B(\theta_0,\delta)^c} \big(L_T(\theta) - L_T(\theta_0)\big) \geq & -\limsup_{T\to\infty} \sup_{\theta\in\Theta} \big| \big(L_T^*(\theta) - L_T^*(\theta_0)\big) - \big(L_T(\theta) - L_T(\theta_0)\big) \big| \\ & + \liminf_{T\to\infty} \big(E[l_t^*(\theta_0)] - L_T^*(\theta_0)\big) \\ & + \liminf_{T\to\infty} \inf_{\theta\in B(\theta_0,\delta)^c} \big(L_T^*(\theta) - E[l_t^*(\theta_0)]\big). \end{aligned} \tag{22}$$

We shall prove that the first two terms on the minorant side of (22) equal zero a.s. whereas the third term is strictly positive.

We begin by showing that

$$\sup_{\theta\in\Theta} |L_T^*(\theta) - L_T(\theta)| \to 0 \quad \text{a.s. as } T \to \infty, \tag{23}$$

from which it follows that the first term on the minorant side of (22) equals zero a.s. Note that

$$|l_t^*(\theta) - l_t(\theta)| = \big|\log(h_t^*) - \log(h_t) + u_t^2(1/h_t^* - 1/h_t)\big| \leq \underline{g}^{-1} |h_t^* - h_t| + \underline{g}^{-2} u_t^2 |h_t^* - h_t|,$$

where the inequality makes use of the mean value theorem and Assumption C5. Using Lemma A.1 and the Cauchy-Schwartz inequality we obtain

$$\left\| \sup_{\theta\in\Theta} |l_t^*(\theta) - l_t(\theta)| \right\|_{r/2} \leq C_1 \left(1 + \left\| \sup_{\theta\in\Theta} u_t^2 \right\|_r \right) \left\| \sup_{\theta\in\Theta} |h_t^* - h_t| \right\|_r$$

for some finite C_1. As seen in the proof of Proposition 1, the term in the parenthesis is finite, whereas inequality (21) gives the upper bound $C\kappa^t$ for the term $\|\sup_{\theta\in\Theta}|h_t^* - h_t|\|_r$. Hence, there exists a $\gamma > 1$ such that $\gamma^t \sup_{\theta\in\Theta} |l_t^*(\theta) - l_t(\theta)|$ converges to zero in $L_{r/2}$–norm, and thus $\sum_{t=1}^{\infty} \sup_{\theta\in\Theta} |l_t^*(\theta) - l_t(\theta)| < \infty$ a.s. by Lemma A.2. Hence the result in (23) follows.

To handle the remaining two terms, first consider the summands of $L_T^*(\theta)$, $l_t^*(\theta)$. By Proposition 1, h_t^* is stationary and ergodic, and hence the same holds for $\log(h_t^*) + u_t^2/h_t^*$. Because $h_t^* \geq \underline{g}$, $l_t^*(\theta)$ is bounded from below uniformly in Θ, implying that $E[l_t^*(\theta)]$ is well defined and belongs to $\mathbb{R} \cup \{+\infty\}$ (in particular, $E[\inf_{\theta\in\Theta} l_t^*(\theta)] > -\infty$). Also, by Proposition 1, $E[\sup_{\theta\in\Theta} h_t^{*r}] < \infty$ with $r > 0$, and hence $E[\sup_{\theta\in\Theta} \log(h_t^*)] < \infty$ by Jensen's inequality. As for the term u_t^2/h_t^*, notice that

$$u_t^2 = \sigma_t^2 \varepsilon_t^2 - 2\left(f_t(\mu) - f_t(\mu_0)\right) \sigma_t \varepsilon_t + \left(f_t(\mu) - f_t(\mu_0)\right)^2. \tag{24}$$

For $\theta = \theta_0$, $u_t^2(\theta_0) = \sigma_t^2 \varepsilon_t^2$, and therefore $E[l_t^*(\theta_0)] < \infty$ because $E[\varepsilon_t^2] < \infty$. However, for $\theta \neq \theta_0$, we may have $E[u_t^2/h_t^*] = \infty$. (We note that if $E[\sup_{\theta\in\Theta} l_t^*(\theta)] < \infty$, a uniform law of large numbers applies giving $\sup_{\theta\in\Theta}|L_T^*(\theta)-E[l_t^*(\theta)]|\to 0$ a.s. as $T\to\infty$, in which case the proof simplifies; cf. Straumann and Mikosch (2006), part 2 of the proof of Theorem 4.1.) That the second term on the minorant side of (22) equals zero a.s. can now be concluded from the ergodic theorem (because $l_t^*(\theta_0)$ is a stationary ergodic sequence with $E[l_t^*(\theta_0)]<\infty$).

Now consider the third term on the minorant side of (22). As in Pfanzagl (1969), proof of Lemma 3.11, it can be shown that

$$\liminf_{T\to\infty}\inf_{\theta\in B(\theta_0,\delta)^c}L_T^*(\theta)\geq\inf_{\theta\in B(\theta_0,\delta)^c}E[l_t^*(\theta)]\quad\text{a.s.}\tag{25}$$

We give a brief outline of the required steps. Exactly as in the aforementioned proof of Pfanzagl, it can be shown that $E[l_t^*(\theta)]$ is a lower semicontinuous function on Θ and, moreover, for every $\theta\in\Theta$ there exists an open neighborhood $B(\theta)$ of θ such that $E\left[\inf_{\theta\in B(\theta)\cap\Theta}l_t^*(\theta)\right]>l_\bullet$ whenever $E[l_t^*(\theta)]>l_\bullet$ (we note that $E[l_t^*(\theta)]$ can equal ∞, and also that the property $E[\inf_{\theta\in\Theta}l_t^*(\theta)]>-\infty$ is required here so that the monotone convergence theorem applies). Now let $l_\bullet$ be such that $E[l_t^*(\theta)]>l_\bullet$ for all $\theta\in B(\theta_0,\delta)^c$. The open sets $B(\theta)$, $\theta\in B(\theta_0,\delta)^c$, form a cover of the compact set $B(\theta_0,\delta)^c$, and hence we may choose a finite subcover, say $B(\theta_{(1)}),\ldots,B(\theta_{(k)})$. Because $E[\inf_{\theta\in\Theta}l_t^*(\theta)]>-\infty$, the ergodic theorem yields

$$\lim_{T\to\infty}\inf_{\theta\in B(\theta_{(i)})\cap\Theta}L_T^*(\theta)\geq\lim_{T\to\infty}T^{-1}\sum_{t=1}^{T}\inf_{\theta\in B(\theta_{(i)})\cap\Theta}l_t^*(\theta)=E\left[\inf_{\theta\in B(\theta_{(i)})\cap\Theta}l_t^*(\theta)\right]\quad\text{a.s.,}\tag{26}$$

$i=1,\ldots,k$, even when the expected value in (26) equals $+\infty$ (cf. Billingsley (1995), pp. 284 and 495, and Francq and Zakoïan (2004), p. 617). Making use of the inequality $\inf_{\theta\in B(\theta_0,\delta)^c}L_T^*(\theta)\geq\min_{i=1,\ldots,k}\inf_{\theta\in B(\theta_{(i)})\cap\Theta}L_T^*(\theta)$ and (26) we obtain

$$\liminf_{T\to\infty}\inf_{\theta\in B(\theta_0,\delta)^c}L_T^*(\theta)\geq\liminf_{T\to\infty}\min_{i=1,\ldots,k}\inf_{\theta\in B(\theta_{(i)})\cap\Theta}L_T^*(\theta)\geq\min_{i=1,\ldots,k}E\left[\inf_{\theta\in B(\theta_{(i)})\cap\Theta}l_t^*(\theta)\right]>l_\bullet\quad\text{a.s.}$$

Because $l_\bullet$ is arbitrary, we obtain the result in (25).

By (25) and the lower semicontinuity of $E[l_t^*(\theta)]$ the third term on the minorant side of (22) is positive if $E[l_t^*(\theta)]-E[l_t^*(\theta_0)]\geq 0$ with equality if and only if $\theta=\theta_0$. Because $E[l_t^*(\theta_0)]<\infty$ this obviously holds if $E[l_t^*(\theta)]=\infty$. Therefore in the following we assume that $E[l_t^*(\theta)]<\infty$. In (24) both σ_t^2 and $(f_t(\mu)-f_t(\mu_0))$ are functions of $(y_{t-1},y_{t-2},...)$ only, and hence independent of ε_t. Also h_t^* is a function of $(y_{t-1},y_{t-2},...)$ only, and hence we obtain $E\left[u_t^2/h_t^*\right]=E\left[\sigma_t^2/h_t^*\right]+E[(f_t(\mu)-f_t(\mu_0))^2/h_t^*]$ and, furthermore,

$$E[l_t^*(\theta)]-E[l_t^*(\theta_0)]=E\left[\log\left(h_t^*/\sigma_t^2\right)\right]+E\left[\sigma_t^2/h_t^*\right]+E[(f_t(\mu)-f_t(\mu_0))^2/h_t^*]-1.\tag{27}$$

Making use of the inequality $x-\log(x)\geq 1$ ($x\in\mathbb{R}_+$) and the identification conditions in Assumption C6 we conclude that the expression in (27) is nonnegative and equals zero if and only if $\theta=\theta_0$. This completes the proof. ■

Appendix C: Proofs for Section 4

We first present a simple lemma which is used in the proofs of Propositions 2 and 3.

Lemma C.1. *Suppose the assumptions of Propositions 2 and 3 hold. Then (i)* $\alpha^*_{\theta,t}$ *and* $\alpha^*_{\theta\theta,t}$ *are* $L_{r/2}$*–dominated in* Θ_0 *whereas* $\gamma_{\theta,t}$ *and* $\gamma^*_{\theta,t}$ *are* L_{2r}*–dominated in* Θ_0*, (ii)* $|\alpha^*_{\theta,t}-\alpha_{\theta,t}|$, $|\alpha^*_{\theta\theta,t}-\alpha_{\theta\theta,t}|$, $|\beta^*_t-\beta_t|$, $|\gamma^*_{\theta,t}-\gamma_{\theta,t}|$*, and* $|\delta^*_t-\delta_t|$ *are all bounded from above by* $C_{t-1}|h^*_{t-1}-h_{t-1}|$*, where* $C_{t-1}=\kappa'(1+2|f_{\theta,t-1}|+|f_{\theta,t-1}|^2+|f_{\theta\theta,t-1}|)$ *is* L_r*–dominated in* Θ_0*, and (iii)* $\sup_{\theta\in\Theta_0}|\beta_t|\leq\kappa$, $\sup_{\theta\in\Theta_0}|\beta^*_t|\leq\kappa$, $\sup_{\theta\in\Theta_0}|\delta_t|<\kappa'$*, and* $\sup_{\theta\in\Theta_0}|\delta^*_t|<\kappa'$*, where* κ *and* κ' *are as in Assumptions C2(ii) and N3(iii), respectively.*

Proof. To prove part (i), first note that $\left\|\sup_{\theta\in\Theta_0}h^*_t\right\|_r<\infty$ by Proposition 1 and that u_t is L_{2r}–dominated in Θ_0, as seen in the proof of the same proposition. Thus, Assumption N3(ii) and Lemma A.1 imply that $g^*_{\theta,t}$, $g^*_{u,t}$, $g^*_{\theta\theta,t}$, $g^*_{uu,t}$, $g^*_{u\theta,t}$, and $g^*_{\theta u,t}$ are L_r–dominated in Θ_0. The Lipschitz conditions of Assumptions C2(ii) and N3(iii) ensure that $g^*_{h,t}$, $g^*_{\theta h,t}$, $g_{\theta h,t}$, $g^*_{uh,t}$, $g_{uh,t}$, and $g^*_{hh,t}$ are bounded by a finite constant uniformly over Θ_0. Moreover, Assumptions DGP and N3(i) ensure that $f_{\theta,t}$ and $f_{\theta\theta,t}$ are L_{2r}–dominated in Θ_0 (cf. the beginning of the proof of Proposition 1). The result now follows from Lemma A.1, the Cauchy-Schwartz inequality, and the norm inequality (for simplicity, the same order, $r/2$, is used for the first two terms). In (ii), the boundedness of the absolute differences follows directly from the Lipschitz conditions of Assumption N3(iii) (again, for simplicity, the same upper bound is used for all the absolute differences). As was noted above, $f_{\theta,t}$ and $f_{\theta\theta,t}$ are L_{2r}–dominated in Θ_0, and hence C_{t-1} is L_r–dominated in Θ_0 by Lemma A.1. The results in (iii) follow from the Lipschitz conditions of Assumptions C2(ii) and N3(iii). ■

Proof of Proposition 2. To prove part (a), we first apply Lemma A.3. Set $z_t(\theta)=h_{\theta,t}(\theta)$ and $v^*_{t-1}(\theta)=(\alpha^*_{\theta,t},\beta^*_t)$. For all $v\in\mathbb{R}^{m+l+1}$, $z\in\mathbb{R}^{m+l}$, and $\theta\in\Theta_0$, define the function G as $G(v,z;\theta)=(v_1,\ldots,v_{m+l})+v_{m+l+1}z$, where the subscript denotes a particular coordinate of the vector v. Thus $z_t(\theta)=h_{\theta,t}(\theta)$ satisfies the difference equation $z_t(\theta)=G\left(v^*_{t-1}(\theta),z_{t-1}(\theta);\theta\right)$. Condition G, the continuity of $v^*_t(\cdot)$, and the moment condition $E[\sup_{\theta\in\Theta_0}\psi(|v^*_t(\theta)|)^{r/2}]<\infty$ hold with $\psi(x)=x$ due to Assumption N2 and Lemma C.1. The results of part (a), except for the last one concerning differentiability, now follow from Lemma A.3 (note that the solution $h^*_{\theta,t}(\theta)$ is understood to be initialized from $h^*_{\theta,0}(\theta)$ having this stationary distribution).

The continuous differentiability of $h^*_t(\theta)$ and the relation $\partial h^*_t(\theta)/\partial\theta=h^*_{\theta,t}(\theta)$ can be proved in a manner similar to the one used in Straumann and Mikosch (2006, pp. 2483–2484). To this end, let $x\in\mathbb{C}(\Theta,\mathbb{R}_+)$ be twice continuously differentiable on Θ_0 and define the sequence $\tilde{h}_n(\theta)$, $n\geq 0$, with $\tilde{h}_0(\theta)=x(\theta)$ and $\tilde{h}_n(\theta)=h_{n,n-1}(\theta)$, $n\geq 1$, where $h_{t,s}=(g_t\circ\cdots\circ g_{t-s})(x)$,

$s \geq 0$, with $[g_t(x)](\theta) = g(u_{t-1}(\theta), x(\theta); \theta)$ (cf. proof of Proposition 1). Thus $\tilde{h}_n(\theta)$, $n \geq 0$, is a random sequence in $\mathbb{C}(\Theta, \mathbb{R}_+)$ with its elements being twice continuously differentiable on Θ_0 with probability one (the latter fact follows from Assumption N2). Moreover, $\tilde{h}_n(\theta)$ and $\tilde{h}_{\theta,n}(\theta) = \partial \tilde{h}_n(\theta)/\partial\theta$ are solutions to the difference equations (6) and (9), respectively. Hence, by part (b) of this proposition (the proof of which does not rely on the subresult currently being proven), for some $\gamma > 1$,

$$\gamma^n \sup_{\theta\in\Theta_0} |h^*_{\theta,n}(\theta) - \tilde{h}_{\theta,n}(\theta)| \to 0 \quad \text{in } L_{r/4} - \text{norm as } n \to \infty. \tag{28}$$

On the other hand, note that for any fixed $n \geq 1$, $(\partial h_{t,n-1}(\theta)/\partial\theta, h^*_{\theta,t}(\theta))$ is a stationary process. Therefore, $(\partial h_{t,n-1}(\theta)/\partial\theta, h^*_{\theta,t}(\theta))$ and $(\partial h_{n,n-1}(\theta)/\partial\theta, h^*_{\theta,n}(\theta))$ are identically distributed. In the latter, $\partial h_{n,n-1}(\theta)/\partial\theta = \tilde{h}_{\theta,n}(\theta)$, and hence, making use of (28), it also holds that $\gamma^n \sup_{\theta\in\Theta_0} |h^*_{\theta,t}(\theta) - \partial h_{t,n-1}(\theta)/\partial\theta| \to 0$ in $L_{r/4}$–norm as $n \to \infty$. By Lemma A.2, $\sup_{\theta\in\Theta_0} |h^*_{\theta,t}(\theta) - \partial h_{t,n-1}(\theta)/\partial\theta| \to 0$ a.s. as $n \to \infty$. To conclude, we have shown that $h_{t,n-1}(\theta)$ converges to $h^*_t(\theta)$ a.s. as $n \to \infty$ for each $\theta \in \Theta_0$ (see the proof of Proposition 1) and that $\partial h_{t,n-1}(\theta)/\partial\theta$ converges uniformly to $h^*_{\theta,t}(\theta)$ a.s. as $n \to \infty$. Now, by Lang (1993, Theorem XIII.9.1) and the continuity of $h^*_{\theta,t}(\theta)$ (obtained from Lemma A.3), $h^*_t(\theta)$ is continuously differentiable on Θ_0 and $\partial h^*_t(\theta)/\partial\theta = h^*_{\theta,t}(\theta)$.

To prove part (b), note that by the definitions, using Lemma C.1, and denoting $a_{t-1} = C_{t-1}(1 + |h^*_{\theta,t-1}|)$ we have

$$\begin{aligned} \left|h^*_{\theta,t} - h_{\theta,t}\right| &\leq \left|\alpha^*_{\theta,t} - \alpha_{\theta,t}\right| + \left|\beta^*_t - \beta_t\right| \left|h^*_{\theta,t-1}\right| + \left|\beta_t\right| \left|h^*_{\theta,t-1} - h_{\theta,t-1}\right| \\ &\leq a_{t-1} \left|h^*_{t-1} - h_{t-1}\right| + \kappa \left|h^*_{\theta,t-1} - h_{\theta,t-1}\right|. \end{aligned}$$

Repeated substitution now yields

$$\left|h^*_{\theta,t} - h_{\theta,t}\right| \leq \sum_{j=0}^{t-1} \kappa^{t-1-j} a_j \left|h^*_j - h_j\right| + \kappa^t \left|h^*_{\theta,0} - h_{\theta,0}\right|,$$

where $h_{\theta,0} = 0$. Using Lemma A.1 and Hölder's inequality we obtain

$$\Delta^{-1}_{r/4,t+1} \left\|\sup_{\theta\in\Theta_0} \left|h^*_{\theta,t} - h_{\theta,t}\right|\right\|_{r/4} \leq \sum_{j=0}^{t-1} \kappa^{t-1-j} \left\|\sup_{\theta\in\Theta_0} a_j\right\|_{r/3} \left\|\sup_{\theta\in\Theta_0} \left|h^*_j - h_j\right|\right\|_r + \kappa^t \left\|\sup_{\theta\in\Theta_0} \left|h^*_{\theta,0}\right|\right\|_{r/4}.$$

In the former term on the majorant side, $\left\|\sup_{\theta\in\Theta_0} a_j\right\|_{r/3}$ is bounded by a finite constant by Hölder's inequality, part (a), and Lemma C.1, whereas $\left\|\sup_{\theta\in\Theta_0} |h^*_j - h_j|\right\|_r \leq C\kappa^j$ by (21). Thus the former term is bounded by $C' t\kappa^{t-1}$ for some finite C'. In the latter term, the norm is finite by part (a). Therefore, for some finite C'',

$$\left\|\sup_{\theta\in\Theta_0} |h^*_{\theta,t} - h_{\theta,t}|\right\|_{r/4} \leq C'' \max\{t, t^{4/r}\}\kappa^{t-1}, \tag{29}$$

from which the stated result follows. ■

Proof of Proposition 3. To prove part (a), we first apply Lemma A.3. Set $z_t(\theta) = \mathrm{vec}(h_{\theta\theta,t})$ and $v_{t-1}^*(\theta) = (\mathrm{vec}(\alpha_{\theta\theta,t}^*), \beta_t^*, \gamma_{\theta,t}^*, \delta_t^*, h_{\theta,t}^*)$, where $\mathrm{vec}(\cdot)$ signifies the usual columnwise vectorization of a matrix. For all $v \in \mathbb{R}^{(m+l+1)^2+1}$, $z \in \mathbb{R}^{(m+l)^2}$, and $\theta \in \Theta_0$, define the function G as $G(v, z; \theta) = v_1 + v_2 z + \mathrm{vec}(v_3 v_5') + \mathrm{vec}(v_5 v_3') + v_4 \,\mathrm{vec}(v_5 v_5')$, where $v = (v_1, v_2, v_3, v_4, v_5)$ is partitioned conformably with the partition of $v_{t-1}^*(\theta)$ above. Thus $z_t(\theta) = \mathrm{vec}(h_{\theta\theta,t})$ satisfies the difference equation $z_t(\theta) = G\left(v_{t-1}^*(\theta), z_{t-1}(\theta); \theta\right)$. Condition G as well as the moment condition $E[\sup_{\theta\in\Theta_0} \psi(|v_t^*(\theta)|)^{r/4}] < \infty$ hold with $\psi(x) = \bar{\varkappa}x^2 + \bar{\varpi}$ $(0 < \bar{\varkappa}, \bar{\varpi} < \infty)$ due to the Cauchy-Schwarz inequality, Proposition 2, and Lemmas A.1 and C.1. The results of part (a), except for the last one concerning differentiability, now follow from Lemma A.3 (with $h_{\theta\theta,t}^*(\theta)$ being initialized from $h_{\theta\theta,0}^*(\theta)$ having this stationary distribution). Finally, the proof of differentiability and of the relation $\partial^2 h_t^*(\theta)/\partial\theta\partial\theta' = h_{\theta\theta,t}^*(\theta)$ is analogous to that in Proposition 2, cf. Straumann and Mikosch (2006, pp. 2485–2486). We omit the details for brevity, and only note that the result of part (b) is needed to prove this.

To prove part (b), note that by the definitions

$$\begin{aligned} \left|h_{\theta\theta,t}^* - h_{\theta\theta,t}\right| \;\leq\; & \left|\alpha_{\theta\theta,t}^* - \alpha_{\theta\theta,t}\right| + \left|\beta_t^* h_{\theta\theta,t-1}^* - \beta_t h_{\theta\theta,t-1}\right| + \left|\gamma_{\theta,t}^* h_{\theta,t-1}^{*\prime} - \gamma_{\theta,t} h_{\theta,t-1}'\right| \\ & + \left|h_{\theta,t-1}^* \gamma_{\theta,t}^{*\prime} - h_{\theta,t-1}\gamma_{\theta,t}'\right| + \left|\delta_t^* h_{\theta,t-1}^* h_{\theta,t-1}^{*\prime} - \delta_t h_{\theta,t-1} h_{\theta,t-1}'\right|. \end{aligned} \tag{30}$$

The second, third (which equals the fourth), and fifth term on the majorant side of (30) are bounded from above by $|\beta_t^* - \beta_t||h_{\theta\theta,t-1}^*| + |\beta_t||h_{\theta\theta,t-1}^* - h_{\theta\theta,t-1}|$, $|\gamma_{\theta,t}^* - \gamma_{\theta,t}||h_{\theta,t-1}^*| + |\gamma_{\theta,t}||h_{\theta,t-1}^* - h_{\theta,t-1}|$, and $|\delta_t^* - \delta_t||h_{\theta,t-1}^* h_{\theta,t-1}^{*\prime}| + |\delta_t||h_{\theta,t-1}^* h_{\theta,t-1}^{*\prime} - h_{\theta,t-1}h_{\theta,t-1}'|$, respectively. In the last of these upper bounds, $|h_{\theta,t-1}^* h_{\theta,t-1}^{*\prime} - h_{\theta,t-1}h_{\theta,t-1}'| \leq 2|h_{\theta,t-1}^*||h_{\theta,t-1}^* - h_{\theta,t-1}| + |h_{\theta,t-1}^* - h_{\theta,t-1}|^2$. Using these inequalities and Lemma C.1 we obtain the following inequalities for the four distinct terms on the majorant side of (30):

$$\begin{aligned} \left|\alpha_{\theta\theta,t}^* - \alpha_{\theta\theta,t}\right| &\leq C_{t-1}\left|h_{t-1}^* - h_{t-1}\right|, \\ \left|\beta_t^* h_{\theta\theta,t-1}^* - \beta_t h_{\theta\theta,t-1}\right| &\leq C_{t-1}\left|h_{t-1}^* - h_{t-1}\right|\left|h_{\theta\theta,t-1}^*\right| + \kappa\left|h_{\theta\theta,t-1}^* - h_{\theta\theta,t-1}\right|, \\ \left|\gamma_{\theta,t}^* h_{\theta,t-1}^{*\prime} - \gamma_{\theta,t}h_{\theta,t-1}'\right| &\leq C_{t-1}\left|h_{t-1}^* - h_{t-1}\right|\left|h_{\theta,t-1}^*\right| + \left|\gamma_{\theta,t}\right|\left|h_{\theta,t-1}^* - h_{\theta,t-1}\right|, \\ \left|\delta_t^* h_{\theta,t-1}^* h_{\theta,t-1}^{*\prime} - \delta_t h_{\theta,t-1}h_{\theta,t-1}'\right| &\leq C_{t-1}\left|h_{t-1}^* - h_{t-1}\right|\left|h_{\theta,t-1}^* h_{\theta,t-1}^{*\prime}\right| \\ &\quad + \kappa'(2\left|h_{\theta,t-1}^*\right|\left|h_{\theta,t-1}^* - h_{\theta,t-1}\right| + \left|h_{\theta,t-1}^* - h_{\theta,t-1}\right|^2). \end{aligned}$$

Denoting $b_{t-1} = C_{t-1}(1 + 2|h_{\theta,t-1}^*| + |h_{\theta,t-1}^*|^2 + |h_{\theta\theta,t-1}^*|)$ and $c_{t-1} = 2|\gamma_{\theta,t}| + 2\kappa'|h_{\theta,t-1}^*|$ we obtain

$$\left|h_{\theta\theta,t}^* - h_{\theta\theta,t}\right| \leq b_{t-1}\left|h_{t-1}^* - h_{t-1}\right| + c_{t-1}\left|h_{\theta,t-1}^* - h_{\theta,t-1}\right| + \kappa'\left|h_{\theta,t-1}^* - h_{\theta,t-1}\right|^2 + \kappa\left|h_{\theta\theta,t-1}^* - h_{\theta\theta,t-1}\right|.$$

By repeated substitution

$$\left|h^*_{\theta\theta,t}-h_{\theta\theta,t}\right| \leq \sum_{j=0}^{t-1}\kappa^{t-1-j}\left(b_j\left|h^*_j-h_j\right|+c_j\left|h^*_{\theta,j}-h_{\theta,j}\right|+\kappa'\left|h^*_{\theta,j}-h_{\theta,j}\right|^2\right)+\kappa^t\left|h^*_{\theta\theta,0}-h_{\theta\theta,0}\right|,$$

where $h_{\theta\theta,0}=0$. Using Lemma A.1, Hölder's inequality, and the norm inequality

$$\begin{aligned}
&\Delta^{-1}_{r/8,3t+1}\left\|\sup_{\theta\in\Theta_0}\left|h^*_{\theta\theta,t}-h_{\theta\theta,t}\right|\right\|_{r/8} \\
\leq\ & \sum_{j=0}^{t-1}\kappa^{t-1-j}\left\|\sup_{\theta\in\Theta_0}b_j\right\|_{r/5}\left\|\sup_{\theta\in\Theta_0}\left|h^*_j-h_j\right|\right\|_r+\sum_{j=0}^{t-1}\kappa^{t-1-j}\left\|\sup_{\theta\in\Theta_0}c_j\right\|_{r/2}\left\|\sup_{\theta\in\Theta_0}\left|h^*_{\theta,j}-h_{\theta,j}\right|\right\|_{r/4} \\
&+\sum_{j=0}^{t-1}\kappa^{t-1-j}\kappa'\left\|\sup_{\theta\in\Theta_0}\left|h^*_{\theta,j}-h_{\theta,j}\right|^2\right\|_{r/8}+\kappa^t\left\|\sup_{\theta\in\Theta_0}\left|h^*_{\theta\theta,0}\right|\right\|_{r/4}.
\end{aligned}$$

By arguments already used, the terms $\left\|\sup_{\theta\in\Theta_0}b_j\right\|_{r/5}$, $\left\|\sup_{\theta\in\Theta_0}c_j\right\|_{r/2}$, and $\left\|\sup_{\theta\in\Theta_0}|h^*_{\theta\theta,0}|\right\|_{r/4}$ are bounded by a finite constant. Furthermore, by (21) and (29), the terms $\left\|\sup_{\theta\in\Theta_0}|h^*_j-h_j|\right\|_r$ and $\left\|\sup_{\theta\in\Theta_0}|h^*_{\theta,j}-h_{\theta,j}|\right\|_{r/4}$ are bounded from above by $C'\kappa^j$ and $C'\max\{j,j^{4/r}\}\kappa^j$, respectively, for some finite C'. Therefore, for some finite C'',

$$\left\|\sup_{\theta\in\Theta_0}\left|h^*_{\theta\theta,t}-h_{\theta\theta,t}\right|\right\|_{r/8}\leq C''\Delta_{r/8,3t+1}\left(t\kappa^{t-1}+t\max\{t,t^{4/r}\}\kappa^{t-1}+\kappa^t\right),$$

from which the result follows. ■

Appendix D: Proofs for Section 5

Recall from Section 3 that $L_T(\theta)=T^{-1}\sum_{t=1}^T l_t(\theta)$ and $L^*_T(\theta)=T^{-1}\sum_{t=1}^T l^*_t(\theta)$, where $l_t(\theta)=\log(h_t)+u_t^2/h_t$ and $l^*_t(\theta)=\log(h^*_t)+u_t^2/h^*_t$. Let $L_{\theta,T}(\theta)=\partial L_T(\theta)/\partial\theta$ and $l_{\theta,t}(\theta)=\partial l_t(\theta)/\partial\theta$, and denote the analogous first and second partial derivatives of $L^*_T(\theta)$ and $l^*_t(\theta)$ with $L^*_{\theta,T}$, $L^*_{\theta\theta,T}$, $l^*_{\theta,t}$, and $l^*_{\theta\theta,t}$. As an intermediate step in the proof of Theorem 2, we first establish (in Lemmas D.1–D.4 below) the asymptotic normality of the infeasible estimator $\tilde{\theta}_T$ based on minimizing $L^*_T(\theta)$. This is done by using a standard mean value expansion of the score $L^*_{\theta,T}(\theta)$ given by

$$T^{1/2}L^*_{\theta,T}(\tilde{\theta}_T)=T^{1/2}L^*_{\theta,T}(\theta_0)+\dot{L}^*_{\theta\theta,T}T^{1/2}(\tilde{\theta}_T-\theta_0), \tag{31}$$

where $\dot{L}^*_{\theta\theta,T}$ signifies the matrix $L^*_{\theta\theta,T}(\theta)$ with each row evaluated at an intermediate point $\dot{\theta}_{i,T}$ $(i=1,\ldots,m+l)$ lying between $\tilde{\theta}_T$ and θ_0. Subsequently, in Lemmas D.5 and D.6 we show the asymptotic equivalence of the estimators $\hat{\theta}_T$ and $\tilde{\theta}_T$. The result of Theorem 2 is then obtained as an immediate consequence of the conclusions of Lemmas D.4 and D.6.

Lemma D.1. *If the assumptions of Theorem 2 hold, then* $T^{1/2}L^*_{\theta,T}(\theta_0) \xrightarrow{d} N(0, \mathcal{I}(\theta_0))$*, where* $\mathcal{I}(\theta_0) = E\left[l^*_{\theta,t}(\theta_0)l^{*\prime}_{\theta,t}(\theta_0)\right]$ *is finite and can be expressed as*

$$\begin{aligned}\mathcal{I}(\theta_0) &= \begin{bmatrix} 4E\left[\frac{f_{\mu,t}(\mu_0)}{\sigma_t}\frac{f'_{\mu,t}(\mu_0)}{\sigma_t}\right] & 0_{m\times l} \\ 0_{l\times m} & 0_{l\times l}\end{bmatrix} + E\left[\varepsilon_t^4 - 1\right]E\left[\frac{h^*_{\theta,t}(\theta_0)}{\sigma_t^2}\frac{h^{*\prime}_{\theta,t}(\theta_0)}{\sigma_t^2}\right] \\ &\quad +2E\left[\varepsilon_t^3\right]\begin{bmatrix} E\left[\frac{f_{\mu,t}(\mu_0)}{\sigma_t}\frac{h^{*\prime}_{\mu,t}(\theta_0)}{\sigma_t^2} + \frac{h^*_{\mu,t}(\theta_0)}{\sigma_t^2}\frac{f'_{\mu,t}(\mu_0)}{\sigma_t}\right] & E\left[\frac{f_{\mu,t}(\mu_0)}{\sigma_t}\frac{h^{*\prime}_{\lambda,t}(\theta_0)}{\sigma_t^2}\right] \\ E\left[\frac{h^*_{\lambda,t}(\theta_0)}{\sigma_t^2}\frac{f'_{\mu,t}(\mu_0)}{\sigma_t}\right] & 0_{l\times l}\end{bmatrix}. \end{aligned} \tag{32}$$

Proof. Partitioning $l^*_{\theta,t}$ as $l^*_{\theta,t} = (l^*_{\mu,t}, l^*_{\lambda,t})$, direct calculation yields

$$l^*_{\mu,t} = -2\frac{f_{\mu,t}}{h_t^{*1/2}}\frac{u_t}{h_t^{*1/2}} - \frac{h^*_{\mu,t}}{h^*_t}\left(\frac{u_t^2}{h^*_t} - 1\right) \quad \text{and} \quad l^*_{\lambda,t} = -\frac{h^*_{\lambda,t}}{h^*_t}\left(\frac{u_t^2}{h^*_t} - 1\right), \tag{33}$$

and hence

$$l^*_{\mu,t}(\theta_0) = -2\frac{f_{\mu,t}(\mu_0)}{\sigma_t}\varepsilon_t - \frac{h^*_{\mu,t}(\theta_0)}{\sigma_t^2}\left(\varepsilon_t^2 - 1\right) \quad \text{and} \quad l^*_{\lambda,t}(\theta_0) = -\frac{h^*_{\lambda,t}(\theta_0)}{\sigma_t^2}\left(\varepsilon_t^2 - 1\right). \tag{34}$$

By straightforward calculation one now obtains the expression (32). As seen in the proof of Lemma C.1, $f_{\mu,t}$ is L_{2r}–dominated in Θ_0. From this fact and Assumptions C5 and N4 it follows that $E\left[l^*_{\theta,t}(\theta_0)l^{*\prime}_{\theta,t}(\theta_0)\right]$ is finite. Noting that $l^*_{\theta,t}(\theta_0)$ is a stationary ergodic martingale difference sequence and $T^{1/2}L^*_{\theta,T}(\theta_0) = T^{-1/2}\sum_{t=1}^T l^*_{\theta,t}(\theta_0)$, the stated convergence is obtained from Billingsley's (1961) central limit theorem in conjunction with the Cramér-Wold device. ■

Lemma D.2. *If the assumptions of Theorem 2 hold, then* $l^*_{\theta\theta,t}(\theta)$ *is* L_1*–dominated in* Θ_0 *and*

$$\sup_{\theta\in\Theta_0}\left|L^*_{\theta\theta,T}(\theta) - \mathcal{J}(\theta)\right| \to 0 \quad a.s.,$$

where $\mathcal{J}(\theta) = E\left[l^*_{\theta\theta,t}(\theta)\right]$ *is continuous at* θ_0*. Moreover,* $\mathcal{J}(\theta_0)$ *can be expressed as*

$$\mathcal{J}(\theta_0) = \begin{bmatrix} 2E\left[\frac{f_{\mu,t}(\mu_0)}{\sigma_t}\frac{f'_{\mu,t}(\mu_0)}{\sigma_t}\right] & 0_{m\times l} \\ 0_{l\times m} & 0_{l\times l}\end{bmatrix} + E\left[\frac{h^*_{\theta,t}(\theta_0)}{\sigma_t^2}\frac{h^{*\prime}_{\theta,t}(\theta_0)}{\sigma_t^2}\right]. \tag{35}$$

Proof. The first partial derivatives of l^*_t were obtained in (33), whereas the second ones are

$$\begin{aligned} l^*_{\mu\mu,t} &= -\frac{h^*_{\mu\mu,t}}{h^*_t}\left(\frac{u_t^2}{h^*_t} - 1\right) + \frac{h^*_{\mu,t}}{h^*_t}\frac{h^{*\prime}_{\mu,t}}{h^*_t}\left(2\frac{u_t^2}{h^*_t} - 1\right) - 2\frac{f_{\mu\mu,t}}{h_t^{*1/2}}\frac{u_t}{h_t^{*1/2}} \\ &\quad +2\frac{f_{\mu,t}}{h_t^{*1/2}}\frac{f'_{\mu,t}}{h_t^{*1/2}} + 2\left(\frac{f_{\mu,t}}{h_t^{*1/2}}\frac{h^{*\prime}_{\mu,t}}{h^*_t} + \frac{h^*_{\mu,t}}{h^*_t}\frac{f'_{\mu,t}}{h_t^{*1/2}}\right)\frac{u_t}{h_t^{*1/2}}, \\ l^*_{\mu\lambda,t} &= -\frac{h^*_{\mu\lambda,t}}{h^*_t}\left(\frac{u_t^2}{h^*_t} - 1\right) + \frac{h^*_{\mu,t}}{h^*_t}\frac{h^{*\prime}_{\lambda,t}}{h^*_t}\left(2\frac{u_t^2}{h^*_t} - 1\right) + 2\frac{f_{\mu,t}}{h_t^{*1/2}}\frac{h^{*\prime}_{\lambda,t}}{h^*_t}\frac{u_t}{h_t^{*1/2}}, \\ l^*_{\lambda\lambda,t} &= -\frac{h^*_{\lambda\lambda,t}}{h^*_t}\left(\frac{u_t^2}{h^*_t} - 1\right) + \frac{h^*_{\lambda,t}}{h^*_t}\frac{h^{*\prime}_{\lambda,t}}{h^*_t}\left(2\frac{u_t^2}{h^*_t} - 1\right). \end{aligned}$$

It follows from Assumption DGP and Propositions 1, 2, and 3 that $l^*_{\theta\theta,t}$ forms a stationary and ergodic sequence in $\mathbb{C}\left(\Theta_0, \mathbb{R}^{(m+l)\times(m+l)}\right)$ and hence the uniform strong law of large numbers in Theorem 2.7 of Straumann and Mikosch (2006) applies if $E\left[\sup_{\theta\in\Theta_0}|l^*_{\theta\theta,t}(\theta)|\right]$ is finite. Thus, the stated convergence is proved if

$$\left\|\sup_{\theta\in\Theta_0}|u_t|\right\|_4, \quad \left\|\sup_{\theta\in\Theta_0}|f_{\mu,t}|\right\|_4, \quad \left\|\sup_{\theta\in\Theta_0}|f_{\mu\mu,t}|\right\|_4, \quad \left\|\sup_{\theta\in\Theta_0}\frac{1}{h^*_t}\right\|_\infty, \quad \left\|\sup_{\theta\in\Theta_0}\frac{|h^*_{\theta,t}|}{h^*_t}\right\|_4, \text{ and } \left\|\sup_{\theta\in\Theta_0}\frac{|h^*_{\theta\theta,t}|}{h^*_t}\right\|_2$$

are all finite. For the first three norms, this has already been justified in the proofs of Propositions 1 and Lemma C.1, whereas Assumption C5 implies the finiteness of the fourth norm. The last two are finite by Assumption N4. Finally, the continuity of $\mathcal{J}(\theta)$ at θ_0 also follows from the aforementioned theorem of Straumann and Mikosch (2006), and that $\mathcal{J}(\theta_0)$ can be expressed as in (35) is seen by straightforward calculation. ■

Lemma D.3. *If the assumptions of Theorem 2 hold, then the matrices $\mathcal{I}(\theta_0)$ and $\mathcal{J}(\theta_0)$ are positive definite.*

Proof. Consider the matrix $\mathcal{I}(\theta_0)$. For an arbitrary $x = (x_\mu, x_\lambda) \in \mathbb{R}^m \times \mathbb{R}^l$, suppose $x'\mathcal{I}(\theta_0)x = E\left[\left(x' l^*_{\theta,t}(\theta_0)\right)^2\right] = 0$. Then, by (34), we must have

$$x' l^*_{\theta,t}(\theta_0) = 2\varepsilon_t x' \frac{f_{\theta,t}(\theta_0)}{\sigma_t} + \left(\varepsilon_t^2 - 1\right) x' \frac{h^*_{\theta,t}(\theta_0)}{\sigma_t^2} = 0 \quad \text{a.s.}$$

Following exactly the same steps as in Francq and Zakoïan (2004) (their arguments between equations (4.52) and (4.53)) we can use Assumption N5(i) to show that, almost surely, $x'_\mu f_{\mu,t}(\theta_0) = 0$ and $x' h^*_{\theta,t}(\theta_0) = 0$. By Assumption N5(ii), $x_\mu = 0$, and hence $x'_\lambda h^*_{\lambda,t}(\theta_0) = 0$. By equation (11) and the definitions preceding it in Section 4,

$$\begin{aligned} h^*_{\lambda,t}(\theta_0) &= \alpha^*_{\lambda,t}(\theta_0) + \beta^*_t(\theta_0)\, h^*_{\lambda,t-1}(\theta_0) \\ &= \partial g(u_{0,t-1}, \sigma^2_{t-1}; \theta_0)/\partial\lambda + \partial g(u_{0,t-1}, \sigma^2_{t-1}; \theta_0)/\partial h \cdot h^*_{\lambda,t-1}(\theta_0). \end{aligned}$$

By stationarity, also $x'_\lambda h^*_{\lambda,t-1}(\theta_0) = 0$, and hence $x'_\lambda \partial g(u_{0,t-1}, \sigma^2_{t-1}; \theta_0)/\partial\lambda = 0$. By Assumption N5(iii), $x_\lambda = 0$, and hence we have proved that $\mathcal{I}(\theta_0)$ is positive definite.

Regarding the matrix $\mathcal{J}(\theta_0)$, note that $x'\mathcal{J}(\theta_0)x = 0$ now directly implies that

$$2E\left[\left(x'_\mu f_{\mu,t}(\theta_0)\right)^2 \sigma_t^{-2}\right] + E\left[\left(x' h^*_{\theta,t}(\theta_0)\right)^2 \sigma_t^{-4}\right] = 0.$$

This can only happen if $x'_\mu f_{\mu,t}(\theta_0) = 0$ and $x' h^*_{\theta,t}(\theta_0) = 0$ a.s. As above, this implies that $x = 0$. Hence also $\mathcal{J}(\theta_0)$ is positive definite. ■

Lemma D.4. *If the assumptions of Theorem 2 hold, then*

$$T^{1/2}(\tilde{\theta}_T - \theta_0) \xrightarrow{d} N\left(0, \mathcal{J}(\theta_0)^{-1}\mathcal{I}(\theta_0)\mathcal{J}(\theta_0)^{-1}\right).$$

Proof. First note that from the proof of Theorem 1 it can be seen that $\tilde{\theta}_T \to \theta_0$ a.s. (because $\liminf_{T\to\infty} \inf_{\theta\in B(\theta_0,\delta)^c} (L_T^*(\theta) - L_T^*(\theta_0))$ equals the sum of the last two terms on the minorant side of (22)). Recalling the mean value expansion of $L_{\theta,T}^*(\theta)$ in (31), by the strong consistency of $\tilde{\theta}_T$ we also have $\dot{\theta}_{i,T} \to \theta_0$ a.s. as $T\to\infty$ $(i = 1,\ldots,m+l)$. This, together with the uniform convergence result for $L_{\theta\theta,T}^*(\theta)$ in Lemma D.2, now yields $\dot{L}_{\theta\theta,T}^* \to \mathcal{J}(\theta_0)$ a.s. as $T\to\infty$ (see Amemiya (1973), Lemma 4). By Lemma D.3, $\mathcal{J}(\theta_0)$ is invertible, and hence for all T sufficiently large also $\dot{L}_{\theta\theta,T}^*$ is invertible and $\dot{L}_{\theta\theta,T}^{*-1} \to \mathcal{J}(\theta_0)^{-1}$ a.s. as $T\to\infty$ (see Lemma A.1 of Pötscher and Prucha (1991b)). Multiplying the mean value expansion (31) with the Moore-Penrose inverse $\dot{L}_{\theta\theta,T}^{*+}$ of $\dot{L}_{\theta\theta,T}^*$ (this inverse exists for all T) and rearranging we obtain

$$T^{1/2}(\tilde{\theta}_T - \theta_0) = (I - \dot{L}_{\theta\theta,T}^{*+}\dot{L}_{\theta\theta,T}^*)T^{1/2}(\tilde{\theta}_T - \theta_0) + \dot{L}_{\theta\theta,T}^{*+}T^{1/2}L_{\theta,T}^*(\tilde{\theta}_T) - \dot{L}_{\theta\theta,T}^{*+}T^{1/2}L_{\theta,T}^*(\theta_0). \quad (36)$$

The first two terms on the right hand side of (36) converge to zero a.s. (more precisely, for all events ω on a set with probability one, there exists a $T(\omega)$ such that for all $T \geq T(\omega)$ the first two terms are identically equal to zero). For the first term, this follows from the fact that for all T sufficiently large $\dot{L}_{\theta\theta,T}^*$ is invertible. For the second one, this holds because $\tilde{\theta}_T$ being a minimizer of $L_T^*(\theta)$ and θ_0 being an interior point of Θ_0 yield $L_{\theta,T}^*(\tilde{\theta}_T) = 0$ for all T sufficiently large. Furthermore, the eventual a.s. invertibility of $\dot{L}_{\theta\theta,T}^*$ also means that $\dot{L}_{\theta\theta,T}^{*+} - \mathcal{J}(\theta_0)^{-1} \to 0$ a.s. Hence, (36) becomes

$$T^{1/2}(\tilde{\theta}_T - \theta_0) = o_1(1) - (\mathcal{J}(\theta_0)^{-1} + o_2(1))T^{1/2}L_{\theta,T}^*(\theta_0),$$

where $o_1(1)$ and $o_2(1)$ (a vector- and a matrix-valued process, respectively) converge to zero a.s. Combining this with the result of Lemma D.1 completes the proof. ■

Lemma D.5. *If the assumptions of Theorem 2 hold, then for some $\gamma > 1$,*

$$\gamma^t \sup_{\theta\in\Theta_0} \left| l_{\theta,t}^*(\theta) - l_{\theta,t}(\theta)\right| \to 0 \text{ in } L_{1/3}\text{–norm as } t\to\infty.$$

Proof. In this proof we assume $r = 2$, but retain the notation r for ease of comparison to previous results. First consider the difference $h_{\theta,t}^*/h_t^* - h_{\theta,t}/h_t$ and use Assumption C5 to obtain $|h_{\theta,t}^*/h_t^* - h_{\theta,t}/h_t| \leq \underline{g}^{-2}|h_{\theta,t}^*||h_t^* - h_t| + \underline{g}^{-1}|h_{\theta,t}^* - h_{\theta,t}|$. By Lemma A.1, Hölder's inequality, and the norm inequality, we now find that

$$\Delta_{r/4,2}^{-1}\left\|\sup_{\theta\in\Theta_0}\left|\frac{h_{\theta,t}^*}{h_t^*} - \frac{h_{\theta,t}}{h_t}\right|\right\|_{r/4} \leq \underline{g}^{-2}\left\|\sup_{\theta\in\Theta_0}\left|h_{\theta,t}^*\right|\right\|_{r/2}\left\|\sup_{\theta\in\Theta_0}\left|h_t^* - h_t\right|\right\|_r + \underline{g}^{-1}\left\|\sup_{\theta\in\Theta_0}\left|h_{\theta,t}^* - h_{\theta,t}\right|\right\|_{r/4}.$$

Thus, Proposition 2 and inequalities (21) and (29) give

$$\left\|\sup_{\theta\in\Theta_0}\left|h_{\theta,t}^*/h_t^* - h_{\theta,t}/h_t\right|\right\|_{r/4} \leq C\max\{t, t^{4/r}\}\kappa^t \quad (37)$$

for some finite C.

Now consider the difference $l^*_{\theta,t}(\theta) - l_{\theta,t}(\theta)$. Making use of Assumption C5 and the inequality $|x^*y^* - xy| \leq |x^* - x|\,|y^*| + |x^* - x|\,|y^* - y| + |x^*|\,|y^* - y|$ for any conformable vectors we obtain

$$\begin{aligned}
\left|l^*_{\theta,t}(\theta) - l_{\theta,t}(\theta)\right| &= \left|-\frac{h^*_{\theta,t}}{h^*_t}\left(\frac{u_t^2}{h^*_t} - 1\right) + \frac{h_{\theta,t}}{h_t}\left(\frac{u_t^2}{h_t} - 1\right) - 2\frac{f_{\theta,t}}{h^*_t}u_t + 2\frac{f_{\theta,t}}{h_t}u_t\right| \\
&\leq \left|\frac{h^*_{\theta,t}}{h^*_t} - \frac{h_{\theta,t}}{h_t}\right|\left|\frac{u_t^2}{h^*_t} - 1\right| + \left|\frac{h^*_{\theta,t}}{h^*_t} - \frac{h_{\theta,t}}{h_t}\right|\left|\frac{u_t^2}{h^*_t} - \frac{u_t^2}{h_t}\right| \\
&\quad + \left|\frac{h^*_{\theta,t}}{h^*_t}\right|\left|\frac{u_t^2}{h^*_t} - \frac{u_t^2}{h_t}\right| + 2\left|f_{\theta,t}\right|\left|u_t\right|\left|\frac{1}{h^*_t} - \frac{1}{h_t}\right| \\
&\leq \left|\frac{h^*_{\theta,t}}{h^*_t} - \frac{h_{\theta,t}}{h_t}\right|(\underline{g}^{-1}u_t^2 + 1) + \left|\frac{h^*_{\theta,t}}{h^*_t} - \frac{h_{\theta,t}}{h_t}\right|u_t^2\underline{g}^{-2}\left|h^*_t - h_t\right| \\
&\quad + \left|h^*_{\theta,t}\right|u_t^2\underline{g}^{-3}\left|h^*_t - h_t\right| + 2\left|f_{\theta,t}\right|\left|u_t\right|\underline{g}^{-2}\left|h^*_t - h_t\right|.
\end{aligned}$$

By Lemma A.1, Hölder's inequality, and the norm inequality

$$\begin{aligned}
\Delta^{-1}_{r/6,4}\left\|\sup_{\theta\in\Theta_0}\left|l^*_{\theta,t}(\theta) - l_{\theta,t}(\theta)\right|\right\|_{r/6} &\leq \left\|\sup_{\theta\in\Theta_0}\left|\frac{h^*_{\theta,t}}{h^*_t} - \frac{h_{\theta,t}}{h_t}\right|\right\|_{r/4}\left\|\sup_{\theta\in\Theta_0}(\underline{g}^{-1}u_t^2 + 1)\right\|_r \\
&\quad + \underline{g}^{-2}\left\|\sup_{\theta\in\Theta_0}\left|\frac{h^*_{\theta,t}}{h^*_t} - \frac{h_{\theta,t}}{h_t}\right|\right\|_{r/4}\left\|\sup_{\theta\in\Theta_0}u_t^2\right\|_r\left\|\sup_{\theta\in\Theta_0}\left|h^*_t - h_t\right|\right\|_r \\
&\quad + \underline{g}^{-3}\left\|\sup_{\theta\in\Theta_0}\left|h^*_{\theta,t}\right|\right\|_{r/2}\left\|\sup_{\theta\in\Theta_0}u_t^2\right\|_r\left\|\sup_{\theta\in\Theta_0}\left|h^*_t - h_t\right|\right\|_r \\
&\quad + 2\underline{g}^{-2}\left\|\sup_{\theta\in\Theta_0}\left|f_{\theta,t}\right|\right\|_{2r}\left\|\sup_{\theta\in\Theta_0}\left|u_t\right|\right\|_{2r}\left\|\sup_{\theta\in\Theta_0}\left|h^*_t - h_t\right|\right\|_r.
\end{aligned}$$

The result now follows from inequalities (21) and (37) and arguments already used. ■

Lemma D.6. *If the assumptions of Theorem 2 hold, then $T^{1/2}(\hat{\theta}_T - \tilde{\theta}_T) \to 0$ a.s. as $T \to \infty$.*

Proof. Because both $\hat{\theta}_T$ and $\tilde{\theta}_T$ are strongly consistent estimators of θ_0 (see Theorem 1 and the proof of Lemma D.4), we can assume that T is so large that $\hat{\theta}_T, \tilde{\theta}_T \in \Theta_0$ with probability one. From the identity $L^*_{\theta,T}(\tilde{\theta}_T) = L_{\theta,T}(\hat{\theta}_T) = 0$ and the mean value theorem one then obtains

$$T^{1/2}\big(L_{\theta,T}(\hat{\theta}_T) - L^*_{\theta,T}(\hat{\theta}_T)\big) = T^{1/2}\big(L^*_{\theta,T}(\tilde{\theta}_T) - L^*_{\theta,T}(\hat{\theta}_T)\big) = \ddot{L}^*_{\theta\theta,T}T^{1/2}(\tilde{\theta}_T - \hat{\theta}_T), \qquad (38)$$

where $\ddot{L}^*_{\theta\theta,T}$ signifies the matrix $L^*_{\theta\theta,T}(\theta)$ with each row evaluated at an intermediate point $\ddot{\theta}_{i,T}$ $(i = 1, \ldots, m + l)$ lying between $\tilde{\theta}_T$ and $\hat{\theta}_T$. Concerning the term on the left hand side of (38),

$$T^{1/2}\left|L_{\theta,T}(\hat{\theta}_T) - L^*_{\theta,T}(\hat{\theta}_T)\right| \leq T^{-1/2}\sum_{t=1}^{T}\sup_{\theta\in\Theta_0}\left|l^*_{\theta,t}(\theta) - l_{\theta,t}(\theta)\right|,$$

where the majorant side converges to zero a.s. by Lemmas D.5 and A.2. On the other hand, similarly to the proof of Lemma D.4 it can be shown that the matrix $\ddot{L}^*_{\theta\theta,T}$ on the right hand side of (38) is invertible for all T sufficiently large and $\ddot{L}^{*-1}_{\theta\theta,T} \to \mathcal{J}(\theta_0)^{-1}$ a.s. as $T \to \infty$. Hence $T^{1/2}(\hat{\theta}_T - \tilde{\theta}_T) \to 0$ a.s. as $T \to \infty$. ■

Proof of (17). In this proof we assume $r = 2$, but retain the notation r for ease of comparison to previous results. It suffices to show that the four quantities in (16) are strongly consistent estimators of the corresponding four expectations in (15). Due to the strong consistency of $\hat{\theta}_T$, it suffices to prove that

$$\left| T^{-1} \sum_{t=1}^{T} A_t^{*(i)} - E\big[A_t^{*(i)}\big] \right|, \quad i = 1, \ldots, 4, \quad \text{and} \quad \left| T^{-1} \sum_{t=1}^{T} \big(A_t^{*(i)} - A_t^{(i)}\big) \right|, \quad i = 1, \ldots, 4, \tag{39}$$

converge to zero almost surely uniformly over Θ_0 as $T \to \infty$, where

$$A_t^{*(1)} = \frac{f_{\mu,t}}{h_t^{*1/2}} \frac{f'_{\mu,t}}{h_t^{*1/2}}, \quad A_t^{*(2)} = u_t^4 / h_t^{*2}, \quad A_t^{*(3)} = \frac{h^*_{\theta,t}}{h^*_t} \frac{h^{*\prime}_{\theta,t}}{h^*_t}, \quad \text{and} \quad A_t^{*(4)} = \frac{f_{\mu,t}}{h_t^{*1/2}} \frac{h^{*\prime}_{\theta,t}}{h^*_t},$$

and $A_t^{(i)}$, $i = 1, \ldots, 4$, are defined similarly but with h^*_t and $h^*_{\theta,t}$ replaced with h_t and $h_{\theta,t}$, respectively.

Concerning the former four convergences in (39), these can be deduced from Theorem 2.7 of Straumann and Mikosch (2006) if $E\big[\sup_{\theta \in \Theta_0} |A_t^{*(i)}|\big] < \infty$, $i = 1, \ldots, 4$, holds. For $i = 1$, this follows from the fact that $f_{\mu,t}$ is L_{2r}–dominated in Θ_0 (see the proof of Lemma C.1) and Assumption C5. For $i = 2$, the finiteness follows in view of Assumption C5 and the fact that $\big\| \sup_{\theta \in \Theta} |u_t| \big\|_{2r} < \infty$ (see the proof of Proposition 1). For $i = 3$, this holds due to Assumption N4, whereas for $i = 4$, this follows from Assumptions DGP, C5, N3(i), and N4.

Now consider the latter four convergences in (39). For $i = 1$, use Assumption C5 to obtain

$$|A_t^{*(1)} - A_t^{(1)}| = |h_t^{*-1} f_{\mu,t} f'_{\mu,t} - h_t^{-1} f_{\mu,t} f'_{\mu,t}| \le \underline{g}^{-2} |f_{\mu,t} f'_{\mu,t}| |h^*_t - h_t|.$$

Thus, by the Cauchy-Schwartz inequality, the aforementioned L_{2r}–dominance of $f_{\mu,t}$, and (21),

$$\left\| \sup_{\theta \in \Theta_0} |A_t^{*(1)} - A_t^{(1)}| \right\|_{r/2} \le \underline{g}^{-2} \left\| \sup_{\theta \in \Theta_0} |f_{\mu,t} f'_{\mu,t}| \right\|_r \left\| \sup_{\theta \in \Theta_0} |h^*_t - h_t| \right\|_r \le C \kappa^t$$

for some finite C. The required convergence for $i = 1$ now follows from Lemma A.2.

The cases $i = 2$, 3, and 4 can be handled in a similar way. We only note that for $i = 2$,

$$|A_t^{*(2)} - A_t^{(2)}| = u_t^4 |h_t / h_t^{*2} h_t - h^*_t / h_t^{*2} h_t + h_t / h^*_t h_t^2 - h^*_t / h^*_t h_t^2| \le 2 \underline{g}^{-3} u_t^4 |h^*_t - h_t|,$$

and Hölder's inequality, the fact that $\big\| \sup_{\theta \in \Theta} |u_t| \big\|_{2r} < \infty$, and (21), give

$$\left\| \sup_{\theta \in \Theta_0} |A_t^{*(2)} - A_t^{(2)}| \right\|_{r/3} \le 2 \underline{g}^{-3} \left\| \sup_{\theta \in \Theta_0} u_t^4 \right\|_{r/2} \left\| \sup_{\theta \in \Theta_0} |h^*_t - h_t| \right\|_r \le C \kappa^t.$$

For $i = 3$, note that

$$|A_t^{*(3)} - A_t^{(3)}| \leq 2|h_{\theta,t}^*/h_t^*||h_{\theta,t}^*/h_t^* - h_{\theta,t}/h_t| + |h_{\theta,t}^*/h_t^* - h_{\theta,t}/h_t|^2,$$

and thus, by Lemma A.1, the Cauchy-Schwartz inequality, Assumption N4, and inequality (37),

$$\begin{aligned}
&\left\| \sup_{\theta\in\Theta_0} |A_t^{*(3)} - A_t^{(3)}| \right\|_{r/8} \\
\leq\ & \Delta_{r/8,2} \left(2 \left\| \sup_{\theta\in\Theta_0} |h_{\theta,t}^*/h_t^*| \right\|_{r/4} \left\| \sup_{\theta\in\Theta_0} |h_{\theta,t}^*/h_t^* - h_{\theta,t}/h_t| \right\|_{r/4} + \left\| \sup_{\theta\in\Theta_0} |h_{\theta,t}^*/h_t^* - h_{\theta,t}/h_t|^2 \right\|_{r/8} \right) \\
\leq\ & C\big(\max\{t, t^{4/r}\}\kappa^t + \max\{t^2, t^{8/r}\}\kappa^{2t}\big),
\end{aligned}$$

for some finite C. For $i = 4$, using the inequality $|x^*y^* - xy| \leq |x^* - x|\,|y^*| + |x^*|\,|y^* - y| + |x^* - x|\,|y^* - y|$ for any conformable vectors, the mean value theorem for the function $x^{-1/2}$, and Assumption C5,

$$\begin{aligned}
|A_t^{*(4)} - A_t^{(4)}| \leq\ & |f_{\mu,t}||h_t^{*-1/2} - h_t^{-1/2}||h_{\theta,t}^*/h_t^*| + |f_{\mu,t}||h_t^{*-1/2}||h_{\theta,t}^*/h_t^* - h_{\theta,t}/h_t| \\
& + |f_{\mu,t}||h_t^{*-1/2} - h_t^{-1/2}||h_{\theta,t}^*/h_t^* - h_{\theta,t}/h_t| \\
\leq\ & 2^{-1}\underline{g}^{-3/2}|f_{\mu,t}||h_t^* - h_t||h_{\theta,t}^*/h_t^*| + \underline{g}^{-1/2}|f_{\mu,t}||h_{\theta,t}^*/h_t^* - h_{\theta,t}/h_t| \\
& + 2^{-1}\underline{g}^{-3/2}|f_{\mu,t}||h_t^* - h_t||h_{\theta,t}^*/h_t^* - h_{\theta,t}/h_t|.
\end{aligned}$$

First using Lemma A.1 and Hölder's inequality, then Assumption N4, the L_{2r}–dominance of $f_{\mu,t}$, and the inequalities (21) and (37), and finally Lemma A.2, yields the required convergence result. Thus, we have justified (17). ■

Appendix E: Technical details of the examples

Example 1: Linear AR–GARCH

We first show that the conditions in (a) suffice for the validity of Assumption DGP. First consider the process σ_t. Because $u_{0,t} = \sigma_t\varepsilon_t$, the conditional variance process $\sigma_t^2 = g\left(\sigma_{t-1}\varepsilon_{t-1}, \sigma_{t-1}^2; \theta_0\right) = \omega_0 + \alpha_0\sigma_{t-1}^2\varepsilon_{t-1}^2 + \beta_0\sigma_{t-1}^2$ is a function of its own past value and ε_{t-1}^2 only. By Francq and Zakoïan (2004, Proposition 1) and Straumann and Mikosch (2006, Theorem 3.5) the condition $E[\ln(\beta_0 + \alpha_0\varepsilon_t^2)] < 0$ in (a.i) implies the existence of a unique strictly stationary ergodic solution σ_t^2 to this difference equation. By the same reference, this solution is measurable with respect to the σ–algebra generated by $(\varepsilon_{t-1}, \varepsilon_{t-2}, \ldots)$ and $E[\sigma_t^{2r}] < \infty$ for some $r > 0$. Hence the process $(\sigma_t, \varepsilon_t)$ is stationary and ergodic, measurable with respect to the σ–algebra generated by $(\varepsilon_t, \varepsilon_{t-1}, \ldots)$, and $E[\sigma_t^{2r}] < \infty$ and $E[|\varepsilon_t|^{2r}] < \infty$ for some $r > 0$. Therefore, $u_{0,t} = \sigma_t\varepsilon_t$ is stationary and ergodic with $E[|u_{0,t}|^{2r}] < \infty$. Denote $\phi_0(z) = 1 - \sum_{j=1}^{p} \phi_{0,j} z^j$ and let $\phi_0(z)^{-1} = \sum_{j=0}^{\infty} \pi_{0,j} z^j$ be the power series expansion of $\phi_0(z)^{-1}$. As is well known, condition (a.ii) implies that $|\pi_{0,j}| \le C\rho^j$ for some $0 \le \rho < 1$ and $0 < C < \infty$, so that the expansion of $\phi_0(z)^{-1}$ is well defined for $|z| \le 1$. Moreover, from Lemma A.2 we find that the series $y_t = \sum_{j=0}^{\infty} \pi_{0,j} u_{0,t-j}$ converges almost surely. Thus, using Theorem 2.6 of Straumann and Mikosch (2006), (y_t, σ_t^2) is stationary and ergodic. Furthermore, from Lemma A.2 we can also conclude that $E[|y_t|^{2r}] < \infty$. Thus, Assumption DGP holds.

For the assumptions required for consistency, first note that the parameter space is compact by definition so that it is immediate that Assumptions C1, C3, and C5 hold (the last one because ω is bounded away from zero for all $\theta \in \Theta$). The compactness also implies that, for all $\theta \in \Theta$, $\beta \le \bar{\beta} < 1$ for some $\bar{\beta}$, yielding Assumption C2 except for the continuity of the function g, which is obvious. Assumption C4 is similarly obvious. To see that Assumption C6 holds (cf. Francq and Zakoïan (2004), result (ii) in their proof of Theorem 2.1 and result (ii) in their proof of Theorem 3.1), first assume that $f(y_{t-1}, \ldots, y_{t-p}; \mu) = f(y_{t-1}, \ldots, y_{t-p}; \mu_0)$ a.s. for some $\mu \neq \mu_0$, which implies the existence of a linear combination of $y_{t-1}, \ldots, y_{t-p}$ that is a.s. equal to a constant. Hence, to have $\mu \neq \mu_0$, we must have y_{t-i} for some $i = 1, \ldots, p$ being a.s. equal to a deterministic function of y_{t-i-j}, $j \ge 1$. However, by definition $y_{t-i} = f(y_{t-i-1}, \ldots, y_{t-i-p}; \mu_0) + \sigma_{t-i}\varepsilon_{t-i}$ and, conditional on y_{t-i-j}, $j \ge 1$, y_{t-i} is not deterministic because $\sigma_{t-i} \ge \underline{\omega} > 0$ and ε_{t-i} is not degenerate (because $E[\varepsilon_{t-i}] = 0$ and $E[\varepsilon_{t-i}^2] = 1$). Hence $\mu = \mu_0$. Similarly it can be shown that $h_t^*(\mu_0, \lambda) = h_t^*(\mu_0, \lambda_0)$ a.s. implies $\lambda = \lambda_0$ given condition (b.i) and the fact that $\alpha_0 > 0$.

Now consider the validity of the assumptions required for asymptotic normality. Assumption N1 holds by condition (c.i), and Assumptions N2 and N3 are clearly satisfied (N3(iii) with $\kappa' = 1$). For Assumption N4, first note that condition (c.ii) ensures that $E[\sigma_t^4] < \infty$ in the case of a pure GARCH model (see, e.g., Francq and Zakoïan (2004)). Therefore, above in the justification of assumption DGP, the arguments remain valid with $r = 2$. Hence it can be seen that Assumption DGP holds with $r = 2$. The two moment conditions for the derivatives of the process h_t^* can be verified as in Francq and Zakoïan (2004, p. 635), derivation of their equations (4.59) and (4.60). Assumption N5(i) is identical to condition (c.iii). For Assumption N5(ii), note that having $x_\mu' \partial f_t(\mu_0)/\partial\mu = 0$ a.s. with $x_\mu \neq 0$ implies the existence of a linear combination of $y_{t-1}, \ldots, y_{t-p}$ that is a.s. equal to a constant, and a contradiction follows exactly as in verifying Assumption C6. For N5(iii), suppose that $x_\lambda' \partial g(u_{0,t-1}, \sigma_{t-1}^2; \theta_0)/\partial\lambda = x_{\lambda 1} + x_{\lambda 2}\sigma_{t-1}^2\varepsilon_{t-1}^2 + x_{\lambda 3}\sigma_{t-1}^2 = 0$. First, $x_{\lambda 2} = 0$, because otherwise ε_{t-1}^2 would be a (measurable) function of $(\varepsilon_{t-2}, \varepsilon_{t-3}, \ldots)$. Then, we must also have $x_{\lambda 3} = 0$, because otherwise σ_{t-1}^2 would be a.s. equal to a constant, which is impossible due to $\alpha_0 > 0$ and (b.i). Thus, we also get $x_{\lambda 1} = 0$ and $x_\lambda = 0$

so that Assumption N5 holds.

Example 2: AR–AGARCH

Similarly to the case of the linear AR–GARCH model, but now using Theorem 3.5 and Remark 3.6 of Straumann and Mikosch (2006), it can be shown that for the AR–AGARCH model the process $(\sigma_t, \varepsilon_t)$ is stationary and ergodic, measurable with respect to the σ–algebra generated by $(\varepsilon_t, \varepsilon_{t-1}, \ldots)$, and $E[\sigma_t^{2r}] < \infty$ and $E[|\varepsilon_t|^{2r}] < \infty$ for some $r > 0$. Continuing exactly as in the case of the linear AR–GARCH model Assumption DGP can be verified.

For the assumptions required for consistency, C1–C5 and C6(i) can be checked in a manner similar to that of the linear AR–GARCH case whereas C6(ii) can be verified exactly as in Straumann and Mikosch (2006, Lemmas 5.2–5.4). Details are omitted.

Example 3: Nonlinear AR–GARCH

We begin with supplementing conditions (a)–(c) given in Section 6 with conditions required for the nonlinear functions F and G. Subscripts in F and G will denote partial derivatives with respect to the variable(s) in question.

(a) **(v)** The derivatives of $F(\cdot;\varphi_0)$ and $G(\cdot;\gamma_0)$ exist up to any order and are continuous, and $G(\cdot;\gamma_0)$ is strictly increasing (or, alternatively, strictly decreasing).

(b) **(iii)** The functions $F(\cdot;\cdot)$ and $G(\cdot;\cdot)$ are continuous.

- **(iv)** For all φ, $\lim_{y\to-\infty} yF(y;\varphi) = 0$ and $\lim_{y\to\infty} y(1 - F(y;\varphi)) = 0$; if $\varphi \neq \varphi_0$, then for some $\bar{y}$, $F(\bar{y};\varphi) \neq F(\bar{y};\varphi_0)$.
- **(v)** For all γ, $\lim_{u\to-\infty} u^2 G(u;\gamma) = 0$ and $\lim_{u\to\infty} u^2(1 - G(u;\gamma)) = 0$ (or, alternatively, $\lim_{u\to\infty} u^2 G(u;\gamma) = 0$ and $\lim_{u\to-\infty} u^2(1 - G(u;\gamma)) = 0$); if $\gamma \neq \gamma_0$, then for some $\bar{u}$, $G(\bar{u};\gamma) \neq G(\bar{u};\gamma_0)$.

(c) **(iii)** There exist open neighbourhoods $N(\varphi_0)$ and $N(\gamma_0)$ of φ_0 and γ_0 such that $F(\cdot;\cdot)$ and $G(\cdot;\cdot)$ are twice continuously partially differentiable on $\mathbb{R} \times N(\varphi_0)$ and $\mathbb{R} \times N(\gamma_0)$, respectively. Moreover, these partial derivatives are bounded in absolute value uniformly over $\mathbb{R} \times N(\varphi_0)$ and $\mathbb{R} \times N(\gamma_0)$, respectively.

- **(iv)** $\lim_{y\to\pm\infty} yF_\varphi(y;\varphi_0) = 0$; if $(x_1, x_2) \neq (0, 0)$, then for some $\bar{y}$, $(x_1, x_2)' F_\varphi(\bar{y};\varphi_0) \neq 0$.
- **(v)** $\lim_{u\to\pm\infty} u^2 G_\gamma(u;\gamma_0) = 0$; if $(x_1, x_2) \neq (0, 0)$, then for some $\bar{u}$, $(x_1, x_2)' G_\gamma(\bar{u};\gamma_0) \neq 0$.
- **(vi)** $G_u(u;\gamma)\, u^2$, $G_{uu}(u;\gamma)\, u^2$, and $G_{u\gamma}(u;\gamma)\, u^2$ are bounded in absolute value uniformly over $\mathbb{R} \times N(\gamma_0)$.

All of the conditions above are satisfied if F and G are, for example, cumulative distribution functions of either the logistic or the normal distribution. Condition (a.v) is required to apply the results in Meitz and Saikkonen (2008b). Here, as well as in condition (b.v), we separately consider the cases of G being either strictly increasing or strictly decreasing. Condition (b.iii) is needed for the continuity requirement in Assumptions C2 and C4. It is also used to verify the identification conditions in Assumption C6, for which also (b.iv) and (b.v) are needed. Note that (b.iv) implies that $\lim_{y\to-\infty} F(y;\varphi) = 0$ and $\lim_{y\to\infty} F(y;\varphi) = 1$. Condition (c.iii) ensures the differentiability requirements in Assumptions N2 and N3(i)–(ii), and is also used to verify the identification conditions in Assumption N5. Conditions (c.iv) and (c.v) are also needed for Assumption N5 to hold. Finally, (c.vi) is required for the moment conditions for the derivatives in Assumption N4.

We now show that Assumptions DGP, C1–C6, and N1–N5 hold under the conditions made.

Verification of Assumption DGP.

The validity of Assumption DGP follows from the conditions in (a) due to the results in Meitz and Saikkonen (2008b). Specifically, the conditions in (a) imply that Assumptions 1–4, 5(b), and 6 of Meitz and Saikkonen (2008b) hold so that from Theorem 1 of that paper we can conclude that Assumption DGP holds. To see this, note first that, because $E\left[\varepsilon_t^2\right] = 1$, condition (a.i) implies that Assumption 1 of Meitz and Saikkonen (2008b) holds with the value of r therein equal to unity, whereas the conditions imposed on the function F in (a.v) and the assumed range of F imply Assumption 2 of the same paper. That Assumption 3 of Meitz and Saikkonen (2008b) holds follows from the discussion given in Section 4 of that paper and condition (a.ii). Finally, (a.iii), (a.iv), and the conditions assumed about the function G in (a.v) and its range imply that the model satisfies the assumptions required for the model for conditional variance in Proposition 1 of Meitz and Saikkonen (2008b).[9] Of the two alternative cases in that proposition, (a) and (b), the latter is relevant, and it follows that Theorem 1 of Meitz and Saikkonen (2008b) applies with some $r_0 \in (0,1)$. Thus, Assumption DGP holds with $r = r_0$.

Verification of Assumptions for consistency.

For the assumptions required for consistency, first note that Assumption C1 holds due to the definition of the permissible parameter space. The continuity condition in Assumption C2 is an immediate consequence of condition (b.iii). The other conditions in Assumption C2 hold

[9] If G is strictly decreasing, a reparameterization is required in order to apply Proposition 1 of Meitz and Saikkonen (2008b): defining $\alpha^*_{0,1} = \alpha_{0,1} + \alpha_{0,2}$, $\alpha^*_{0,2} = -\alpha_{0,2}$, and $G^*(u;\gamma_{0,1},\gamma_{0,2}) = 1 - G(u;\gamma_{0,1},\gamma_{0,2})$ this proposition applies (this reparameterization is only used when applying the aforementioned proposition of Meitz and Saikkonen (2008b); for parameter estimation, the relevant parameters are still $\alpha_{0,1}$ and $\alpha_{0,2}$).

because the range of the function G is $[0,1]$ and because, for all $\theta \in \Theta$, $\beta \leq \bar{\beta} < 1$ for some $\bar{\beta}$ in view of the assumed compactness of the parameter space. Assumption C3 is satisfied because the range of the function F is $[0,1]$, whereas Assumption C4 is implied by condition (b.iii). Assumption C5 holds because, due to compactness, ω is bounded away from zero for all $\theta \in \Theta$.

In order to verify Assumption C6(i), we first demonstrate that if A_i, $i = 0, \ldots, p$, are any nonempty open subsets of $\mathbb{R}$, the event

$$\{(y_t, \ldots, y_{t-p}) \in A_0 \times \cdots \times A_p\} \tag{40}$$

has a positive probability. To this end, by the aforementioned results of Meitz and Saikkonen (2008b), $\left(y_t, \ldots, y_{t-p}, \sigma_t^2\right)$ is a (geometrically ergodic) Markov chain to which Proposition 4.2.2(iii) and Theorem 10.4.9 of Meyn and Tweedie (1993) apply. By these two results, the event in (40) has positive probability if, from any fixed initial value, the (nonstationary) chain $(y_t^{\dagger}, \ldots, y_{t-p}^{\dagger}, \sigma_t^{\dagger 2})$ eventually reaches the set $A_0 \times \cdots \times A_p \times \mathbb{R}_+$ with positive probability (here we need to distinguish between the chain $\left(y_t, \ldots, y_{t-p}, \sigma_t^2\right)$ initialized from the stationary distribution and the nonstationary one obtained by using a fixed initial value). Because ε_t has a density that is positive everywhere, the nonstationary chain can reach the set $A_p \times \mathbb{R}^p \times \mathbb{R}_+$ in one step with positive probability. Next, making use of the Chapman-Kolmogorov equations (see Meyn and Tweedie (1993, Theorem 3.4.2)), the set $A_{p-1} \times A_p \times \mathbb{R}^{p-1} \times \mathbb{R}_+$ can be reached in the next step with positive probability. Continuing inductively, in $p+1$ steps the set $A_0 \times \cdots \times A_p \times \mathbb{R}_+$ can be reached with positive probability. Because this holds for any initial value, the event in (40) has a positive probability.

Consider now the identification condition in Assumption C6(i). To this end, define $A_j(y; \mu, \mu_0) = \phi_j - \phi_{0,j} + \psi_j F(y; \varphi) - \psi_{0,j} F(y; \varphi_0)$, $j = 0, \ldots, p$, let $\bar{y}_1, \ldots, \bar{y}_p$ denote real numbers, and choose a $\mu \in \mathsf{M}$ such that $f\left(y_{t-1}, \ldots, y_{t-p}; \mu\right) = f\left(y_{t-1}, \ldots, y_{t-p}; \mu_0\right)$ a.s. Then

$$A_0(y_{t-d}; \mu, \mu_0) + \sum_{j=1}^{p} A_j(y_{t-d}; \mu, \mu_0) y_{t-j} = 0 \quad \text{a.s.} \tag{41}$$

We first demonstrate that $\phi_j = \phi_{0,j}$, $j = 0, \ldots, p$. First suppose that $\phi_d \neq \phi_{0,d}$, and consider the set $S(d, y_\bullet) = \{(\bar{y}_1, \ldots, \bar{y}_p) : \bar{y}_d \in (y_\bullet - 1, y_\bullet), \bar{y}_j \in (-1, 1), j \neq d\}$, where $y_\bullet < 0$. Concerning the deterministic sum $A_0(\bar{y}_d; \mu, \mu_0) + \sum_{j=1,\ldots,p, j\neq d} A_j(\bar{y}_d; \mu, \mu_0)\bar{y}_j$, we can find an $M > 0$ (not depending on $y_\bullet$) such that this sum is bounded in absolute value by M on the set $S(d, y_\bullet)$ for any $y_\bullet < 0$ (this holds because the range of F is $[0,1]$). On the other hand, because $\phi_d \neq \phi_{0,d}$, it follows from condition (b.iv) that the term $A_d(\bar{y}_d; \mu, \mu_0)\bar{y}_d$ will attain values arbitrarily large in absolute value on the set $S(d, y_\bullet)$ when $y_\bullet$ is chosen small enough. In particular, for $y_\bullet$ small enough, $|A_d(\bar{y}_d; \mu, \mu_0)\bar{y}_d| > M$. Because the event $\{(y_{t-1}, \ldots, y_{t-p}) \in S(d, y_\bullet)\}$ has positive probability for any $y_\bullet$, we can contradict (41), and hence $\phi_d = \phi_{0,d}$.

Next suppose that $\phi_k \neq \phi_{0,k}$ for some $k = 1, \ldots, p$, $k \neq d$, and consider the set $S(k, y_\bullet) = \{(\bar{y}_1, \ldots, \bar{y}_p) : \bar{y}_k, \bar{y}_d \in (y_\bullet - 1, y_\bullet), \bar{y}_j \in (-1, 1), j \neq k, d\}$, where $y_\bullet < 0$. First note that because $\phi_d = \phi_{0,d}$, $A_d(\bar{y}_d; \mu, \mu_0)\bar{y}_d = (\psi_d F(\bar{y}_d; \varphi) - \psi_{0,d} F(\bar{y}_d; \varphi_0))\bar{y}_d$ will approach 0 as $\bar{y}_d \to -\infty$ due to condition (b.iv). Hence, the deterministic sum $A_0(\bar{y}_d; \mu, \mu_0) + \sum_{j=1,\ldots,p, j\neq k} A_j(\bar{y}_d; \mu, \mu_0)\bar{y}_j$ will be bounded in absolute value by some $M > 0$ on the set $S(k, y_\bullet)$ for all sufficiently small $y_\bullet < 0$ (and M does not depend on $y_\bullet$). Again, because $\phi_k \neq \phi_{0,k}$, the term $A_k(\bar{y}_d; \mu, \mu_0)\bar{y}_k$ will attain values arbitrarily large in absolute value on the set $S(k, y_\bullet)$ when $y_\bullet$ is chosen small enough, and a contradiction is found in the same way as above. Therefore $\phi_j = \phi_{0,j}$ for all $j = 1, \ldots, p$.

Finally, to show that $\phi_0 = \phi_{0,0}$, consider the set $S(y_\bullet) = \{(\bar{y}_1, \ldots, \bar{y}_p) : \bar{y}_j \in (y_\bullet - 1, y_\bullet), j = 1, \ldots, p\}$, where $y_\bullet < 0$. Under the restrictions derived so far and making use of condition (b.iv), the deterministic sum $A_0(\bar{y}_d; \mu, \mu_0) + \sum_{j=1}^{p} A_j(\bar{y}_d; \mu, \mu_0)\bar{y}_j$ will tend to $\phi_0 - \phi_{0,0}$ on the set $S(y_\bullet)$ when $y_\bullet$ is chosen small enough. As above, a contradiction is found, and thus $\phi_0 = \phi_{0,0}$.

In an analogous manner it can be shown that $\psi_j = \psi_{0,j}$, $j = 0, \ldots, p$, and we only give an outline of the required steps. First, supposing that $\psi_d \neq \psi_{0,d}$ and considering the set $S(d, y_\bullet)$ with arbitrarily large positive values of $y_\bullet$, one arrives to a contradiction as above by using condition (b.iv). Then, under the restriction $\psi_d = \psi_{0,d}$, one first notes that $A_d(y; \mu, \mu_0) = \psi_{0,d}\big[\big(1 - F(y; \varphi_0)\big) - \big(1 - F(y; \varphi)\big)\big]$. This fact, and the use of the sets $S(k, y_\bullet)$ for sufficiently large $y_\bullet > 0$, yields $\psi_j = \psi_{0,j}$ for $j = 1, \ldots, p$, $j \neq d$. Finally, making use of the sets $S(y_\bullet)$ with large positive $y_\bullet$, one obtains $\psi_0 = \psi_{0,0}$.

The identity (41) now takes the form

$$\big(F(y_{t-d}; \varphi) - F(y_{t-d}; \varphi_0)\big)\left[\psi_{0,0} + \sum_{j=1}^{p} \psi_{0,j} y_{t-j}\right] = 0 \quad \text{a.s.} \tag{42}$$

If $\varphi \neq \varphi_0$, then by the last part of condition (b.iv) we can find a $\bar{y}$ such that $F(\bar{y}; \varphi) - F(\bar{y}; \varphi_0) \neq 0$. The continuity of $F(\cdot; \cdot)$ assumed in (b.iii) now ensures the existence of some $y_\bullet < \bar{y} < y^\bullet$ such that $F(\bar{y}_d; \varphi) - F(\bar{y}_d; \varphi_0)$ is bounded away from zero for all $\bar{y}_d \in (y_\bullet, y^\bullet)$. On the other hand, by condition (b.i), at least one of the $\psi_{0,j}$, $j = 0, \ldots, p$, is nonzero. First suppose that $\psi_{0,d} \neq 0$, and consider the set $S(d, \delta) = \{(\bar{y}_1, \ldots, \bar{y}_p) : \bar{y}_d \in (y_\bullet, y^\bullet), \bar{y}_j \in (-\delta, \delta), j \neq d\}$, where $\delta > 0$. The deterministic sum $\psi_{0,0} + \sum_{j=1,\ldots,p, j\neq d} \psi_{0,j}\bar{y}_j$ will take values in a small neighborhood of $\psi_{0,0}$ on the set $S(d, \delta)$ when δ is sufficiently small. On the other hand, $\psi_{0,d}\bar{y}_d$ takes the values between $\psi_{0,d} y_\bullet$ and $\psi_{0,d} y^\bullet$ on the set $S(d, \delta)$. Because the event $\{(y_{t-1}, \ldots, y_{t-p}) \in S(d, \delta)\}$ has positive probability for any $\delta > 0$, we find by choosing δ small enough that the term in square brackets in (42) cannot be equal to zero with probability one. Hence, unless $\varphi = \varphi_0$, a contradiction has been found. Now suppose that $\psi_{0,d} = 0$ but $\psi_{0,k} \neq 0$ for some $k = 1, \ldots, p$, $k \neq d$. Consider the set $S(k, \delta) = \{(\bar{y}_1, \ldots, \bar{y}_p) : \bar{y}_k, \bar{y}_d \in (y_\bullet, y^\bullet), \bar{y}_j \in (-\delta, \delta), j \neq k, d\}$, where $\delta > 0$. Using similar arguments as above, a contradiction is again found unless $\varphi = \varphi_0$. Finally, if $\psi_{0,j} = 0$ for all $j = 1, \ldots, p$ but $\psi_{0,0} \neq 0$, a contradiction is obvious unless $\varphi = \varphi_0$. Therefore $\varphi = \varphi_0$, which completes the proof of $\mu = \mu_0$ and hence the verification of the identification condition C6(i).

In order to prove part (ii) of Assumption C6, we first show that for some $\underline{\sigma} > 0$ (which will be defined below) and all $\underline{\sigma} < \sigma_\bullet < \sigma^\bullet$, the probability of the event

$$\{\sigma_t^2 \in (\sigma_\bullet, \sigma^\bullet)\} \tag{43}$$

is positive. As when considering the event in (40), it suffices to show that the nonstationary chain $(y_t^\dagger, \ldots, y_{t-p}^\dagger, \sigma_t^{\dagger 2})$ eventually reaches the set $\mathbb{R}^{p+1} \times (\sigma_\bullet, \sigma^\bullet)$ with positive probability from any initial value. The components $y_t^\dagger, \ldots, y_{t-p}^\dagger$ are not essential here, so we concentrate only on $\sigma_t^{\dagger 2}$. In one step from a fixed initial value σ_0^2, the process $\sigma_t^{\dagger 2}$ reaches

$$\sigma_1^{\dagger 2} = \omega_0 + (\alpha_{0,1} + \alpha_{0,2} G(\sigma_0 \varepsilon_0; \gamma_0)) \sigma_0^2 \varepsilon_0^2 + \beta_0 \sigma_0^2.$$

Because ε_0 has a density that is positive everywhere, $P\left\{\varepsilon_0^2 \leq (\alpha_{0,1} + \alpha_{0,2})^{-1}(1 - \beta_0)/2\right\}$ is positive for all t. For all ε_0 taking such values,

$$\sigma_1^{\dagger 2} \leq \omega_0 + (1 + \beta_0)/2 \cdot \sigma_0^2 \stackrel{def}{=} \omega_0 + \bar{\beta}_0 \sigma_0^2,$$

where $\bar{\beta}_0 < 1$. Moreover, because $\varepsilon_1, \ldots, \varepsilon_{k-1}$ also take such values with positive probability, an application of the Chapman-Kolmogorov equations and an inductive argument yields that $\sigma_k^{\dagger 2} \leq \omega_0(1 + \bar{\beta}_0 + \ldots + \bar{\beta}_0^{k-1}) + \bar{\beta}_0^k \sigma_0^2$ with positive probability. Setting $\underline{\sigma} = \omega_0/(1 - \bar{\beta}_0) + \delta$ for some $\delta > 0$ it is clear that $\sigma_k^{\dagger 2} \leq \underline{\sigma}$ with positive probability in a finite number of steps k. Next, because ε_k has an everywhere positive density, in one step $\sigma_{k+1}^{\dagger 2}$ can take values in any set $(\sigma_\bullet, \sigma^\bullet)$ such that $\underline{\sigma} < \sigma_\bullet < \sigma^\bullet$ with positive probability. Hence, $P\{\sigma_t^2 \in (\sigma_\bullet, \sigma^\bullet)\} > 0$.

Now, to prove part (ii) of Assumption C6, choose a $\lambda \in \Lambda$ such that $h_t^*(\mu_0, \lambda) = \sigma_t^2$ a.s. By stationarity, also $h_{t+1}^*(\mu_0, \lambda) = \sigma_{t+1}^2$ a.s., and hence we obtain

$$(\omega - \omega_0) + (\alpha_1 - \alpha_{0,1}) \varepsilon_t^2 \sigma_t^2 + (\alpha_2 G(\sigma_t \varepsilon_t; \gamma) - \alpha_{0,2} G(\sigma_t \varepsilon_t; \gamma_0)) \varepsilon_t^2 \sigma_t^2 + (\beta - \beta_0) \sigma_t^2 = 0 \quad \text{a.s.}$$

By Assumption C5, $\sigma_t^2 \geq \underline{g} > 0$, implying

$$(\alpha_1 - \alpha_{0,1}) \varepsilon_t^2 = -(\beta - \beta_0) - \sigma_t^{-2} \left[(\omega - \omega_0) + (\alpha_2 G(\sigma_t \varepsilon_t; \gamma) - \alpha_{0,2} G(\sigma_t \varepsilon_t; \gamma_0)) \varepsilon_t^2 \sigma_t^2\right] \quad \text{a.s.} \tag{44}$$

By the same assumption and because ε_t has a density that is positive everywhere, the event $\{\sigma_t^2 \geq \underline{g}, \varepsilon_t \leq \underline{g}^{-1/2} M\}$ has positive probability for all $M < 0$, and on this event $\sigma_t \varepsilon_t \leq M$. By condition (b.v), the term in square brackets in (44) can be made arbitrarily close to $(\omega - \omega_0)$

on the event $\{\sigma_t\varepsilon_t \leq M\}$ by choosing a small enough M.[10] Because σ_t^{-2} is bounded by $\underline{g}^{-1}$, the right hand side of (44) is bounded on $\{\sigma_t\varepsilon_t \leq M\}$ whereas the left hand side may attain values arbitrarily large in absolute value if $\alpha_1 \neq \alpha_{0,1}$ and M is chosen small enough. Thus, because $\sigma_t\varepsilon_t \leq M$ with positive probability for every $M < 0$, we must have $\alpha_1 = \alpha_{0,1}$. Under this restriction, (44) can be rearranged as

$$(\alpha_2 - \alpha_{0,2})\,\varepsilon_t^2 = -\,(\beta - \beta_0) - \sigma_t^{-2}\left[(\omega - \omega_0) + (\alpha_2(G(\sigma_t\varepsilon_t;\gamma) - 1) - \alpha_{0,2}(G(\sigma_t\varepsilon_t;\gamma_0) - 1))\,\varepsilon_t^2\sigma_t^2\right] \quad \text{a.s.}$$

Exactly as above, but now considering the event $\{\sigma_t^2 \geq \underline{g}, \varepsilon_t \geq \underline{g}^{-1/2}M\}$ with M taking large positive values, we can deduce $\alpha_2 = \alpha_{0,2}$ by making use of condition (b.v). With the restrictions derived so far,

$$(\omega - \omega_0) + \alpha_{0,2}\,(G(\sigma_t\varepsilon_t;\gamma) - G(\sigma_t\varepsilon_t;\gamma_0))\,\varepsilon_t^2\sigma_t^2 + (\beta - \beta_0)\,\sigma_t^2 = 0 \quad \text{a.s.,} \tag{45}$$

where $\alpha_{0,2} > 0$ by condition (b.ii). Now consider events $\{\sigma_t^2 \in (\sigma_\bullet, \sigma^\bullet), \varepsilon_t \leq \underline{\sigma}^{-1/2}M\}$ with $\underline{\sigma} < \sigma_\bullet < \sigma^\bullet$ and $M < 0$, which, by (43) and the independence of σ_t^2 and ε_t, have positive probability. Moreover, on these events $\sigma_t\varepsilon_t \leq M$ regardless of the values of $\sigma_\bullet$ and $\sigma^\bullet$. Therefore, by condition (b.v) and choosing a small enough M, the sum of the first two terms in (45) can be made arbitrarily close to $(\omega - \omega_0)$ with positive probability. However, considering events with different values of $\sigma_\bullet$ and $\sigma^\bullet$, (45) is clearly violated unless $\beta = \beta_0$. Furthermore, similar reasoning using (45) and the restriction $\beta = \beta_0$ also yields $\omega - \omega_0$. Hence $[G(\sigma_t\varepsilon_t;\gamma) - G(\sigma_t\varepsilon_t;\gamma_0)]\varepsilon_t^2\sigma_t^2 = 0$ a.s. If $\gamma \neq \gamma_0$, then by the last condition in (b.v) and the continuity of $G(\cdot\,;\cdot)$ assumed in (b.iii), we can find some $u_\bullet < u^\bullet$ such that on the event $\{\sigma_t\varepsilon_t \in (u_\bullet, u^\bullet)\}$ the term in square brackets is bounded away from zero. As this event clearly has positive probability, we can conclude that $\gamma = \gamma_0$. Therefore $\lambda = \lambda_0$ and Assumption C6(ii) holds.

Verification of Assumptions for asymptotic normality.

Now consider the validity of the assumptions required for asymptotic normality. Assumption N1 holds by condition (c.i), and Assumption N2 by condition (c.iii). Assumptions N3(i) and N3(ii) can be verified by condition (c.iii), whereas Assumption N3(iii) is clearly satisfied with $\kappa' = 1$. That Assumption DGP holds with $r = 2$ follows from conditions (a) and (c.ii). Specifically, part (a) of Proposition 1 of Meitz and Saikkonen (2008b) now applies with $r = 2$, and thus the validity of Assumption DGP with $r = 2$ follows from Theorem 1 of the same paper (cf. the verification of Assumption DGP above).

[10] This concerns the case of a strictly increasing G. If G is strictly decreasing, consider the event $\{\sigma_t^2 \geq \underline{g}, \varepsilon_t \geq \underline{g}^{-1/2}M\}$ with $M > 0$, on which event $\sigma_t\varepsilon_t \geq M$. Now, considering sufficiently large values of M, the same conclusion is obtained. An analogous change to the arguments is needed also in two other instances in the rest of the verification of C6(ii), but we omit the details.

Verifying the moment conditions for the first and second derivatives of h_t^* in Assumption N4 requires considerably more work. In what follows, we assume that $\theta \in \Theta_0$. Moreover, without loss of generality we may assume Θ_0 is small enough to ensure that $\theta \in \Theta_0$ implies $0 < \underline{\omega} \leq \omega \leq \overline{\omega} < \infty$, $0 < \underline{\alpha}_1 \leq \alpha_1 \leq \overline{\alpha}_1 < \infty$, $0 < \underline{\alpha}_2 \leq \alpha_2 \leq \overline{\alpha}_2 < \infty$, $0 < \underline{\beta} \leq \beta \leq \overline{\beta} < 1$, $\varphi \in N(\varphi_0)$, and $\gamma \in N(\gamma_0)$. Now, for the first norm in Assumption N4 concerning the vector $h_{\theta,t}^*/h_t^*$, recall that in the present case $h_t^* = \omega + (\alpha_1 + \alpha_2 G(u_{t-1};\gamma))\, u_{t-1}^2 + \beta h_{t-1}^*$ (where the argument θ has been suppressed from h_t^* and u_t) and, in the notation of Section 4, $h_{\theta,t}^* = g_{\theta,t}^* - g_{u,t}^* f_{\theta,t-1} + g_{h,t}^* h_{\theta,t-1}^*$ (see equations (7) and (11)). Partitioning $h_{\theta,t}^*$ as $h_{\theta,t}^* = (h_{\mu,t}^*, h_{\lambda,t}^*)$ we obtain $h_{\mu,t}^* = -g_{u,t}^* f_{\mu,t-1} + \beta h_{\mu,t-1}^*$ and $h_{\lambda,t}^* = g_{\lambda,t}^* + \beta h_{\lambda,t-1}^*$ as immediate consequences of the definitions. Because $\beta \leq \bar{\beta} < 1$ by assumption, $h_{\mu,t}^*$ and $h_{\lambda,t}^*$ have the representations

$$h_{\mu,t}^* = -\sum_{j=0}^{\infty} \beta^j g_{u,t-j}^* f_{\mu,t-j-1} \quad \text{and} \quad h_{\lambda,t}^* = \sum_{j=0}^{\infty} \beta^j g_{\lambda,t-j}^*, \tag{46}$$

respectively, where the convergence of the infinite sums follows from Lemmas A.2 and C.1. By straightforward derivation,

$$g_{u,t}^* = 2\left(\alpha_1 + \alpha_2 G(u_{t-1};\gamma)\right) u_{t-1} + \alpha_2 G_u\left(u_{t-1};\gamma\right) u_{t-1}^2,$$

whereas the components of the vector $g_{\lambda,t}^*$ are seen to be

$$1,\ u_{t-1}^2,\ G(u_{t-1};\gamma)u_{t-1}^2,\ h_{t-1}^*,\ \text{and}\ \alpha_2 G_\gamma(u_{t-1};\gamma)u_{t-1}^2. \tag{47}$$

Because the range of G is $[0,1]$, and $G_u\left(u;\gamma\right) u^2$ and $G_\gamma(u;\gamma)$ are bounded in absolute value uniformly over $\mathbb{R}\times N(\gamma_0)$ by conditions (c.iii) and (c.vi), the finiteness of $\left\|\sup_{\theta\in\Theta_0}|h_{\mu,t}^*|/h_t^*\right\|_4$ and $\left\|\sup_{\theta\in\Theta_0}|h_{\lambda,t}^*|/h_t^*\right\|_4$, and hence of the first norm in Assumption N4, follows if we show that

$$\left\|\sup_{\theta\in\Theta_0} \sum_{j=0}^{\infty} \beta^j a_{t-1-j}^{(i)}/h_t^*\right\|_4 < \infty, \quad i = 1,\ldots,4, \tag{48}$$

where $a_t^{(1)} = u_t^2$, $a_t^{(2)} = h_t^*$, $a_t^{(3)} = |u_t||f_{\mu,t}|$, and $a_t^{(4)} = |f_{\mu,t}|$.

To show this, first express h_t^* as

$$h_t^* = \sum_{k=0}^{\infty} \beta^k \left(\omega + (\alpha_1 + \alpha_2 G(u_{t-1-k};\gamma))\, u_{t-1-k}^2\right), \tag{49}$$

where the convergence of the infinite sum follows fom Lemma A.2 and the result $\left\|\sup_{\theta\in\Theta}|u_t|\right\|_{2r} < \infty$ obtained in the proof of Proposition 1. Because $\omega \geq \underline{\omega} > 0$, $\alpha_1 \geq \underline{\alpha}_1 > 0$, $\alpha_2 \geq \underline{\alpha}_2 > 0$, and $\beta \geq \underline{\beta} > 0$ is assumed, we have

$$h_t^* \geq \sum_{k=0}^{\infty} \beta^k \left(\omega + \alpha_1 u_{t-1-k}^2\right) \geq \underline{\omega} + \beta^j \underline{\alpha}_1 u_{t-1-j}^2 \tag{50}$$

for any $j \geq 0$.

Now, considering (48) with $i = 1$ and making use of (50) and the fact that $x/(1+x) \leq x^s$ for all $x \geq 0$ and any $s \in (0,1)$ (cf. Francq and Zakoïan (2004), above their equation (4.25)), we obtain that, for any $j \geq 0$ and $s \in (0,1)$,

$$\begin{aligned} \beta^j u_{t-1-j}^2/h_t^* &\leq \underline{\alpha}_1^{-1} \frac{\beta^j \underline{\alpha}_1 u_{t-1-j}^2/\underline{\omega}}{1+\beta^j \underline{\alpha}_1 u_{t-1-j}^2/\underline{\omega}} \\ &\leq \underline{\alpha}_1^{s-1} \underline{\omega}^{-s} \beta^{sj} \left|u_{t-1-j}\right|^{2s} \\ &\leq \underline{\alpha}_1^{s-1} \underline{\omega}^{-s} \bar{\beta}^{sj} \sup_{\theta\in\Theta_0} \left|u_{t-1-j}\right|^{2s}. \end{aligned} \tag{51}$$

As was noted above, $\left\|\sup_{\theta\in\Theta}|u_t|\right\|_{2r} < \infty$, or $\left\|\sup_{\theta\in\Theta}|u_t|\right\|_4 < \infty$ when $r = 2$ is assumed. Thus, choosing $s \leq 1/2$ and making use of the norm inequality we obtain $\left\|\sup_{\theta\in\Theta_0}|u_t|^{2s}\right\|_4 \leq \left\|\sup_{\theta\in\Theta_0}|u_t|\right\|_4^{2s}$. Using this fact, (51), and Minkowski's inequality we find that

$$\left\|\sup_{\theta\in\Theta_0} \sum_{j=0}^{\infty} \beta^j u_{t-1-j}^2/h_t^*\right\|_4 \leq \underline{\alpha}_1^{s-1} \underline{\omega}^{-s} \sum_{j=0}^{\infty} \bar{\beta}^{sj} \left\|\sup_{\theta\in\Theta_0} |u_{t-1-j}|\right\|_4^{2s},$$

where the majorant side is finite, and hence we have established (48) with $i = 1$.

Now consider (48) with $i = 2$ and conclude from (49) and (50) that

$$\frac{h_{t-1-j}^*}{h_t^*} \leq \sum_{k=0}^{\infty} \beta^k \frac{\omega + (\alpha_1 + \alpha_2 G(u_{t-2-j-k};\gamma))\, u_{t-2-j-k}^2}{\underline{\omega} + \beta^{j+k+1} \underline{\alpha}_1 u_{t-2-j-k}^2}$$

for any $j \geq 0$. Because $\omega \leq \overline{\omega}$ and $\alpha_1 + \alpha_2 G(u_{t-2-j-k};\gamma) \leq C$ for some finite C, we have

$$\frac{h_{t-1-j}^*}{h_t^*} \leq \frac{\overline{\omega}}{\underline{\omega}} \sum_{k=0}^{\infty} \bar{\beta}^k + C \sum_{k=0}^{\infty} \beta^k \frac{u_{t-2-j-k}^2}{\underline{\omega} + \beta^{j+k+1} \underline{\alpha}_1 u_{t-2-j-k}^2}.$$

Hence, by arguments similar to those used to derive (51) we have, for any $j \geq 0$ and $s \in (0,1)$,

$$\begin{aligned} \beta^j \frac{h_{t-1-j}^*}{h_t^*} &\leq \beta^j \frac{\overline{\omega}}{\underline{\omega}} \sum_{k=0}^{\infty} \bar{\beta}^k + \frac{C}{\underline{\alpha}_1 \beta} \sum_{k=0}^{\infty} \frac{\beta^{j+k+1} \underline{\alpha}_1 u_{t-2-j-k}^2/\underline{\omega}}{1+\beta^{j+k+1} \underline{\alpha}_1 u_{t-2-j-k}^2/\underline{\omega}} \\ &\leq \beta^j \frac{\overline{\omega}}{\underline{\omega}} \left(1-\bar{\beta}\right)^{-1} + \frac{C \underline{\alpha}_1^{s-1}}{\underline{\omega}^s \underline{\beta}} \sum_{k=0}^{\infty} \beta^{(j+k+1)s} \left|u_{t-2-j-k}\right|^{2s} \\ &\leq \bar{\beta}^j \frac{\overline{\omega}}{\underline{\omega}} \left(1-\bar{\beta}\right)^{-1} + \frac{C \underline{\alpha}_1^{s-1} \bar{\beta}^s}{\underline{\omega} \underline{\beta}} \bar{\beta}^{js} \sum_{k=0}^{\infty} \bar{\beta}^{ks} \sup_{\theta\in\Theta_0} \left|u_{t-2-j-k}\right|^{2s}. \end{aligned}$$

Choosing $s \leq 1/2$ and using Minkowski's inequality and the norm inequality in the same way as in the case $i = 1$ we find that the norm in (48) is finite when $i = 2$.

Next consider (48) with $i=3$. In view of (50) and the inequality $x/(1+x^2)\le 1$ (cf. Francq and Zakoïan (2004), above their equation (4.49)) we have, for any $j\ge 0$,

$$\begin{aligned}\beta^j\left|u_{t-1-j}\right|\left|f_{\mu,t-1-j}\right|/h_t^* &\le \frac{(\beta^j u_{t-1-j}^2)^{1/2}}{\underline{\omega}+\beta^j\underline{\alpha}_1 u_{t-1-j}^2}\beta^{j/2}\left|f_{\mu,t-j-1}\right| \\ &\le (\underline{\alpha}_1\underline{\omega})^{-1/2}\frac{(\beta^j\underline{\alpha}_1 u_{t-1-j}^2/\underline{\omega})^{1/2}}{1+\beta^j\underline{\alpha}_1 u_{t-1-j}^2/\underline{\omega}}\bar{\beta}^{j/2}\left|f_{\mu,t-j-1}\right| \\ &\le (\underline{\alpha}_1\underline{\omega})^{-1/2}\bar{\beta}^{j/2}\left|f_{\mu,t-j-1}\right|.\end{aligned}$$

As in the case $i=1$, Minkowski's inequality shows that (48) holds with $i=3$ if $\left\|\sup_{\theta\in\Theta_0}|f_{\mu,t}|\right\|_4<\infty$. To verify this, calculate the partial derivatives of $f\left(y_{t-1},\ldots,y_{t-p};\mu\right)$ as

$$1,\ y_{t-1},\ldots,y_{t-p},\ \ (1,y_{t-1},\ldots,y_{t-p})\,F(y_{t-d};\varphi),\ \text{and}\ (\psi_0+\sum_{j=1}^{p}\psi_j y_{t-j})F_{\varphi}(y_{t-d};\varphi). \tag{52}$$

Because the range of F is $[0,1]$ and the partial derivatives of F are bounded uniformly over $\mathbb{R}\times N(\varphi_0)$ by condition (c.iii), we have $|f_{\mu,t}|\le C(1+\sum_{j=1}^{p}|y_{t-j}|)$ for some finite C. Thus, the desired result follows because $E[y_t^4]<\infty$ in view of Assumption DGP and the fact that $r=2$.

For (48) with $i=4$ it suffices to note that $\beta^j|f_{\mu,t-1-j}|/h_t^*\le \underline{g}^{-1}\bar{\beta}^j|f_{\mu,t-1-j}|$ by Assumption C5, and hence the result follows as in the case $i=3$. Therefore we have verified (48) and thus the finiteness of the first norm in Assumption N4.

Now consider the latter norm in Assumption N4 which involves the matrix $h_{\theta\theta,t}^*/h_t^*$. Recall from Section 4 that

$$h_{\theta\theta,t}^*=\alpha_{\theta\theta,t}^*+\beta_t^* h_{\theta\theta,t-1}^*+\gamma_{\theta,t}^* h_{\theta,t-1}^{*\prime}+h_{\theta,t-1}^*\gamma_{\theta,t}^{*\prime}+\delta_t^* h_{\theta,t-1}^* h_{\theta,t-1}^{*\prime},$$

where $\alpha_{\theta\theta,t}^*$, β_t^*, $\gamma_{\theta,t}^*$, and δ_t^* are as in (7)–(8) but with h_t throughout replaced with h_t^*. As already noticed, $\beta_t^*=g_{h,t}^*=\beta$, which implies that $g_{hh,t}^*=0$ and $g_{uh,t}^*=0$. Moreover, only one element of $g_{\theta h,t}^*$ is nonzero, namely the one related to the component β of θ for which the resulting partial derivative is unity. Thus, $\delta_t^*=0$, $\gamma_{\theta,t}^*=\gamma_\theta^*$ is independent of t, and we get

$$h_{\theta\theta,t}^*=\alpha_{\theta\theta,t}^*+\gamma_\theta^* h_{\theta,t-1}^{*\prime}+h_{\theta,t-1}^*\gamma_\theta^{*\prime}+\beta h_{\theta\theta,t-1}^*$$

giving the representation

$$h_{\theta\theta,t}^*=\sum_{j=0}^{\infty}\beta^j\alpha_{\theta\theta,t-j}^*+\sum_{j=0}^{\infty}\beta^j\gamma_\theta^* h_{\theta,t-1-j}^{*\prime}+\sum_{j=0}^{\infty}\beta^j h_{\theta,t-1-j}^*\gamma_\theta^{*\prime}$$

(the infinite sums converge due to Lemmas A.2 and C.1 and Proposition 2). This, and the definition of $\alpha_{\theta\theta,t}^*$, show that for $\left\|\sup_{\theta\in\Theta_0}|h_{\theta\theta,t}^*|/h_t^*\right\|_2<\infty$ it suffices to establish that

$$\left\|\sup_{\theta\in\Theta_0}\sum_{j=0}^{\infty}\beta^j a_{t-1-j}^{(i)}/h_t^*\right\|_2<\infty,\quad i=5,\ldots,9, \tag{53}$$

where $a_t^{(5)} = |g^*_{\theta\theta,t+1}|$, $a_t^{(6)} = |g^*_{uu,t+1}||f_{\theta,t}|^2$, $a_t^{(7)} = |g^*_{u\theta,t+1}||f_{\theta,t}|$, $a_t^{(8)} = |g^*_{u,t+1}||f_{\theta\theta,t}|$, and $a_t^{(9)} = |h^*_{\theta,t}|$.

Because the details of verifying (53) are similar to those already used to deduce (48), we only sketch the required steps. For (53) with $i = 5$, note that because $g^*_{\mu,t+1} = 0$, also $g^*_{\mu\mu,t+1} = 0$ and $g^*_{\lambda\mu,t+1} = 0 = g^{*\prime}_{\mu\lambda,t+1}$. Moreover, by direct calculation, it can be seen that the only nonnull elements of $g^*_{\lambda\lambda,t+1}$ are $G_\gamma(u_t;\gamma)u_t^2$ and $\alpha_2 G_{\gamma\gamma}(u_t;\gamma)u_t^2$. Therefore, $|g^*_{\theta\theta,t+1}|$ is dominated by Cu_t^2 ($C < \infty$). Arguments similar to those used to show (48) with $i = 1$ can now be applied to verify (53) with $i = 5$ (we omit the details). Next, for (53) with $i = 6$, straightforward differentiation gives $g^*_{uu,t+1} = 2(\alpha_1 + \alpha_2 G(u_t;\gamma)) + 4\alpha_2 G_u(u_t;\gamma)u_t + \alpha_2 G_{uu}(u_t;\gamma)u_t^2$. By condition (c.vi), $\sup_{\theta\in\Theta_0}|g^*_{uu,t+1}|$ is bounded, and therefore arguments already used to show (48) with $i = 4$ can be used to obtain the desired result. For (53) with $i = 7$, consider $g^*_{u\theta,t+1}$ and note that $g^*_{u\mu,t+1} = 0$ whereas the nonnull elements of the matrix $g^*_{u\lambda,t+1}$ are $2u_t$, $2G(u_t;\gamma)u_t + G_u(u_t;\gamma)u_t^2$, and $2\alpha_2 G_\gamma(u_t;\gamma)u_t + \alpha_2 G_{u\gamma}(u_t;\gamma)u_t^2$. By conditions (c.iii) and (c.vi), $|g^*_{u\theta,t+1}|$ is dominated by $C(1 + |u_t|)$ ($C < \infty$), and arguments already used to verify (48) with $i = 3$ can be applied to deduce (53) with $i = 7$.

Now consider (53) with $i = 8$. By (52) and direct calculation, the nonnull elements of $f_{\theta\theta,t-1}$ are $(1, y_{t-1}, \ldots, y_{t-p})\, F_\varphi(y_{t-d};\varphi)$ and $(\psi_0 + \sum_{j=1}^p \psi_j y_{t-j}) F_{\varphi\varphi}(y_{t-d};\varphi)$. Thus, similarly to $\sup_{\theta\in\Theta_0}|f_{\theta,t-1}| = \sup_{\theta\in\Theta_0}|f_{\mu,t-1}|$ also $\sup_{\theta\in\Theta_0}|f_{\theta\theta,t-1}|$ is dominated by a term of the form $C(1 + \sum_{j=1}^p |y_{t-j}|)$ with a finite C. Arguments used for (48) with $i = 3$ and 4 can now be used to deduce (53). Finally, for (53) with $i = 9$, recall that we have shown that $\left\|\sup_{\theta\in\Theta_0}|h^*_{\theta,t}|/h^*_t\right\|_4$ is finite, and thus Minkowski's inequality gives $\left\|\sup_{\theta\in\Theta_0}\sum_{j=0}^\infty \beta^j |h^*_{\theta,t-1-j}|/h^*_t\right\|_2 < \infty$. To conclude, we have shown that (53) holds with $i = 5, \ldots, 9$, and therefore that $\left\|\sup_{\theta\in\Theta_0}|h^*_{\theta\theta,t}|/h^*_t\right\|_2 < \infty$. This completes the verification of Assumption N4.

As for Assumption N5, part (i) clearly holds due to condition (a.i). Consider now verifying the condition for the conditional mean in Assumption N5(ii). Recall that the partial derivatives of $f(y_{t-1}, \ldots, y_{t-p};\mu)$ were given in (52), and choose an $x = (x_1, \ldots, x_{2p+4}) \in \mathbb{R}^{2p+4}$ such that $x' \frac{\partial f_t(\mu_0)}{\partial \mu} = 0$ a.s. By (52) and rearranging terms,

$$\begin{aligned}&\left[x_1 + x_{p+2}F(y_{t-d};\varphi_0) + \psi_{0,0}(x_{2p+3}, x_{2p+4})' F_\varphi(y_{t-d};\varphi_0)\right]\\&+\sum_{j=1}^{p}\left[x_{1+j} + x_{p+2+j}F(y_{t-d};\varphi_0) + \psi_{0,j}(x_{2p+3}, x_{2p+4})' F_\varphi(y_{t-d};\varphi_0)\right] y_{t-j} = 0 \quad \text{a.s.}\end{aligned}$$

Using conditions (b.iv) and (c.iv) and arguments similar to those used to verify Assumption C6(i), we can deduce that $x_1 = \cdots = x_{2p+2} = 0$ (without going into details, first, making use of the sets $S(d, y_\bullet)$ with $y_\bullet < 0$ (see the verification of C6(i)), we obtain $x_{1+d} = 0$; next, using the sets $S(k, y_\bullet)$ with $y_\bullet < 0$ and $k \neq d$, we obtain $x_{1+k} = 0$, $k = 1, \ldots, p$, $k \neq d$; similar considerations but now with $y_\bullet > 0$ first give $x_{p+2+d} = 0$ and then $x_{p+2+k} = 0$, $k = 1, \ldots,$

p, $k \neq d$; finally, considering first sufficiently small values of $y_\bullet$ we obtain $x_1 = 0$, and then sufficiently large values of $y_\bullet$ we obtain $x_{p+2} = 0$). Hence

$$\left((x_{2p+3}, x_{2p+4})' F_\varphi(y_{t-d}; \varphi_0)\right) \left[\psi_{0,0} + \sum_{j=1}^{p} \psi_{0,j} y_{t-j}\right] = 0 \quad \text{a.s.}$$

If either $x_{2p+3} \neq 0$ or $x_{2p+4} \neq 0$, then by the last part of condition (c.iv) we can find a $\bar{y}$ such that $(x_{2p+3}, x_{2p+4})' F_\varphi(\bar{y}; \varphi_0) \neq 0$. The continuity of $F_\varphi(\cdot; \cdot)$ assumed in (c.iii) now ensures the existence of some $y_\bullet < \bar{y} < y^\bullet$ such that $(x_{2p+3}, x_{2p+4})' F_\varphi(\bar{y}_d; \varphi_0)$ is bounded away from zero for all $\bar{y}_d \in (y_\bullet, y^\bullet)$. By condition (b.i), at least one of the $\psi_{0,j}$, $j = 0, \ldots, p$, is nonzero, and the arguments used when verifying condition C6(i) can be used to arrive at contradiction (see equation (42) and the discussion following it). Hence, we must have $x_{2p+3} = x_{2p+4} = 0$, and thus $x = 0$. Therefore, Assumption N5(ii) holds.

Now consider Assumption N5(iii), and suppose that for some $x_\lambda = (x_1, \ldots, x_6) \in \mathbb{R}^6$, $x'_\lambda \partial g(u_{0,t}, \sigma_t^2; \theta_0)/\partial\lambda = 0$ a.s. or, using the expressions of the partial derivatives in (47),

$$x_1 + x_2\sigma_t^2\varepsilon_t^2 + x_3 G(\sigma_t\varepsilon_t; \gamma_0)\sigma_t^2\varepsilon_t^2 + x_4\sigma_t^2 + \alpha_{0,2}(x_5, x_6)' G_\gamma(\sigma_t\varepsilon_t; \gamma_0)\sigma_t^2\varepsilon_t^2 = 0 \text{ a.s.} \tag{54}$$

Now, similarly to the verification of Assumption C6(ii), consider the events $\{\sigma_t^2 \in (\sigma_\bullet, \sigma^\bullet), \varepsilon_t \leq \underline{\sigma}^{-1/2} M\}$ with $\underline{\sigma} < \sigma_\bullet < \sigma^\bullet$ and $M < 0$, which by (43) and the independence of σ_t and ε_t have positive probability, and, moreover, on these events $\sigma_t\varepsilon_t \leq M$ regardless of the values of $\sigma_\bullet$ and $\sigma^\bullet$. For fixed $\sigma_\bullet$ and $\sigma^\bullet$ and for arbitrarily small values of M, all the other terms in (54) are bounded (due to conditions (b.v) and (c.v)) except the second one, which takes values arbitrarily large in absolute value unless $x_2 = 0$.[11] Next, under the restriction $x_2 = 0$, writing $x_3 G(\sigma_t\varepsilon_t; \gamma_0)\sigma_t^2\varepsilon_t^2 = x_3\sigma_t^2\varepsilon_t^2 + x_3(G(\sigma_t\varepsilon_t; \gamma_0) - 1)\sigma_t^2\varepsilon_t^2$ and considering the events $\{\sigma_t^2 \in (\sigma_\bullet, \sigma^\bullet), \varepsilon_t \geq \underline{\sigma}^{-1/2} M\}$ with M positive, we can similarly conclude that $x_3 = 0$. With the restrictions derived so far,

$$x_1 + x_4\sigma_t^2 + \alpha_{0,2}(x_5, x_6)' G_\gamma(\sigma_t\varepsilon_t; \gamma_0)\sigma_t^2\varepsilon_t^2 = 0 \quad \text{a.s.} \tag{55}$$

Now, consider again the events $\{\sigma_t^2 \in (\sigma_\bullet, \sigma^\bullet), \varepsilon_t \geq \underline{\sigma}^{-1/2} M\}$ with M positive. Letting M be arbitrarily large, but this time considering these events with different values for $\sigma_\bullet$ and $\sigma^\bullet$, (55) is clearly violated unless $x_4 = 0$. With a similar reasoning, also $x_1 = 0$. Hence $(x_5, x_6)' G_\gamma(\sigma_t\varepsilon_t; \gamma_0)\sigma_t^2\varepsilon_t^2 = 0$ a.s., from which $x_5 = x_6 = 0$ follows by using the last condition in (c.v) and arguments similar to those used at the end of the verification of Assumption C6(ii). Thus, Assumption N5(iii) holds, and the verification of Assumption N5 is complete.

[11] This concerns the case of a strictly increasing G. In the case of a strictly decreasing G, a slight change in the argument is required here and once more below; cf. footnote 10. We omit the details.

References

Amemiya, T. (1973): "Regression analysis when the dependent variable is truncated normal," *Econometrica*, 41, 997–1016.

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

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

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

Billingsley, P. (1995): *Probability and Measure*, 3rd ed., Wiley, New York.

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

Bollerslev, T., R. F. Engle, and D. B. Nelson (1994): "ARCH models," in *Handbook of Econometrics*, ed. by R. F. Engle and D. L. McFadden, vol. IV, chap. 49. Elsevier Science Publishers.

Bougerol, P. (1993): "Kalman filtering with random coefficients and contractions," *SIAM Journal on Control and Optimization*, 31, 942–959.

Boussama, F. (2000): "Normalité asymptotique de l'estimateur du pseudo-maximum de vraisemblance d'un modèle GARCH ," *Comptes rendus de l'Académie des Sciences de Paris, Série I*, 331, 81–84.

Chan, K. S. (1993): "Consistency and limiting distribution of the least squares estimator of a threshold autoregressive model," *Annals of Statistics*, 21, 520–533.

Chan, K. S., and H. Tong (1986): "On estimating thresholds in autoregressive models," *Journal of Time Series Analysis*, 7, 179–190.

Chen, R., and R. S. Tsay (1993): "Functional-coefficient autoregressive models," *Journal of the American Statistical Association*, 88, 298–308.

Cline, D. B. H. (2007): "Stability of nonlinear stochastic recursions with application to nonlinear AR–GARCH models," *Advances in Applied Probability*, 39, 462–491.

Davidson, J. (1994): *Stochastic Limit Theory.* Oxford University Press, Oxford.

Ding, Z., C. W. J. Granger, and R. F. Engle (1993): "A long memory property of stock market returns and a new model," *Journal of Empirical Finance*, 1, 83–106.

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

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

Francq, C., and J.-M. Zakoïan (2006): "Mixing properties of a general class of GARCH(1,1) models without moment assumptions," *Econometric Theory*, 22, 815–834.

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

Glosten, L. W., R. Jaganathan, and D. E. Runkle (1993): "On the relation between the expected value and the volatility of the nominal excess return on stocks," *Journal of Finance*, 48, 1779–1801.

González-Rivera, G. (1998): "Smooth-transition GARCH models," *Studies in Nonlinear Dynamics and Econometrics*, 3, 61–78.

Haggan, V., and T. Ozaki (1981): "Modelling nonlinear random vibrations using an amplitude-dependent autoregressive time series model," *Biometrika*, 68, 189–196.

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

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

Kristensen, D. and A. Rahbek (2005a): "Asymptotics of the QMLE for a class of ARCH(q) models," *Econometric Theory*, 21, 946–961.

Kristensen, D. and A. Rahbek (2005b): "Asymptotics of the QMLE for general ARCH(q) models," Preprint No. 5, Department of Mathematical Sciences, University of Copenhagen.

Lai, T. L. (1994): "Asymptotic properties of nonlinear least squares estimates in stochastic regression models," *Annals of Statistics*, 22, 1917–1930.

Lange, T., A. Rahbek, and S. T. Jensen (2006): "Estimation and asymptotic inference in the first order AR–ARCH model," Preprint No. 4, Department of Mathematical Sciences, University of Copenhagen.

Lanne, M. and P. Saikkonen (2005): "Nonlinear GARCH models for highly persistent volatility," *Econometrics Journal*, 8, 251–276.

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

Liebscher, E. (2005): "Towards a unified approach for proving geometric ergodicity and mixing properties of nonlinear autoregressive processes," *Journal of Time Series Analysis*, 26, 669–689.

Ling, S. (2007a): "Self-weighted and local quasi-maximum likelihood estimators for ARMA–GARCH/IGARCH models," *Journal of Econometrics*, 140, 849–873.

Ling, S. (2007b): "A double AR(p) model: structure and estimation," *Statistica Sinica*, 17, 161–175.

Ling, S., and W. K. Li (1997): "On fractionally integrated autoregressive moving-average time series models with conditional heteroscedasticity," *Journal of the American Statistical Association*, 92, 1184–1194.

Ling, S., and 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., and M. McAleer (2003): "Asymptotic theory for a vector ARMA–GARCH model," *Econometric Theory*, 19, 280–310.

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

Lundbergh, S., and T. Teräsvirta (2002): "Evaluating GARCH models," *Journal of Econometrics*, 110, 417–435.

Meitz, M., and P. Saikkonen (2008a): "Ergodicity, mixing, and existence of moments of a class of Markov models with applications to GARCH and ACD models," *Econometric Theory*, (forthcoming).

Meitz, M., and P. Saikkonen (2008b): "Stability of nonlinear AR–GARCH models," *Journal of Time Series Analysis*, 29, 453–475.

Meyn, S. P., and R. L. Tweedie (1993): *Markov Chains and Stochastic Stability*. Springer-Verlag, London.

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

Pantula, S. G. (1988): "Estimation of autoregressive models with ARCH errors," *Sankhya B*, 50, 119–138.

Pfanzagl, J. (1969): "On the measurability and consistency of minimum contrast estimates," *Metrika*, 14, 249–272.

Pötscher, B. M., and I. R. Prucha (1991a): "Basic structure of the asymptotic theory in dynamic nonlinear econometric models, Part I: Consistency and approximation concepts," *Econometric Reviews*, 10, 125–216.

Pötscher, B. M., and I. R. Prucha (1991b): "Basic structure of the asymptotic theory in dynamic nonlinear econometric models, Part II: Asymptotic normality," *Econometric Reviews*, 10, 253–325.

Straumann, D., and T. Mikosch (2006): "Quasi-maximum-likelihood estimation in conditionally heteroscedastic time series: a stochastic recurrence equations approach," *Annals of Statistics*, 34, 2449–2495.

Teräsvirta, T. (1994): "Specification, estimation, and evaluation of smooth transition autoregressive models," *Journal of the American Statistical Association*, 89, 208–218.

Tjøstheim, D. (1986): "Estimation in nonlinear time series models," *Stochastic Processes and their Applications*, 21, 251–273.

van Dijk, D., T. Teräsvirta, and P. H. Franses (2002): "Smooth transition autoregressive models — A survey of recent developments," *Econometric Reviews*, 21, 1–47.

Weiss, A. A. (1986): "Asymptotic theory for ARCH models: estimation and testing," *Econometric Theory*, 2, 107–131.

White, H. (1980): "Nonlinear regression on cross-section data," *Econometrica*, 48, 721–746.

Zakoïan, J.-M. (1994): "Threshold heteroskedastic models," *Journal of Economic Dynamics and Control*, 18, 931–955.