Econometric Theory, **27**, 2011, 1236–1278.
doi:10.1017/S0266466611000041

PARAMETER ESTIMATION IN NONLINEAR AR–GARCH MODELS

MIKA MEITZ
Koç University

PENTTI SAIKKONEN
University of Helsinki

This paper develops an asymptotic estimation theory for nonlinear autoregressive models with conditionally heteroskedastic errors. We consider a general nonlinear 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. We do not require the rescaled errors to be independent, but instead only to form a stationary and ergodic martingale difference sequence. 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.

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

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). We thank a co-editor and two anonymous referees for helpful comments and suggestions. The first version of this paper was completed in May 2008 while the first author was a post-doctoral research fellow at University of Oxford's Department of Economics. Parts of this research were also 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. Material from the paper has been presented at the Second Brussels-Waseda Seminar on Time Series and Financial Statistics, Brussels, June 2008; ESRC Econometric Study Group Annual Conference, Bristol, July 2008; Workshop on Nonparametric Function Estimation with Applications in Finance, Oulu, June 2009; Econometrics, Time Series Analysis and Systems Theory – A Conference in Honor of Manfred Deistler, Vienna, June 2009; 64th European Meeting of the Econometric Society, Barcelona, August 2009; International Symposium on Econometric Theory and Applications, Singapore, April 2010; and in seminars at Bilkent University, Graz University of Technology, Koç University, University of Aarhus, and University of Vienna. We thank the participants in these occasions for their comments. Address correspondence to: Mika Meitz, Department of Economics, Koç University, Rumelifeneri Yolu, 34450 Sariyer, Istanbul, Turkey; e-mail: mmeitz@ku.edu.tr; 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.

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), Berkes, Horváth, and Kokoszka (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 (2007). Of these papers, Weiss (1986), Pantula (1988), and Lange et al. (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 et al. (2006) and Ling (2007) consider weighted QML estimators instead of the usual one. As these previous papers indicate, the inclusion of an autoregressive conditional mean entails nontrivial complications for the development of asymptotic estimation theory.

The aforementioned papers are all confined to the linear case. Estimation in nonlinear pure ARCH, but not GARCH, models is considered by Kristensen and Rahbek (2005, 2009). To the best of our knowledge, Straumann and Mikosch (2006) and Pan, Wang, and Tong (2008) are the only ones to consider asymptotic estimation theory in nonlinear GARCH models. The former authors study QML estimation in a rather general nonlinear pure GARCH model, whereas the latter focus on so-called power transformed and threshold GARCH models. To exemplify their general theory, Straumann and Mikosch (2006) consider the so-called AGARCH and EGARCH models. 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 the work of the aforementioned authors indicates, allowing for nonlinearities in GARCH models considerably complicates the development of asymptotic estimation theory.

The aforementioned papers also differ in regard to what is assumed of the rescaled error term (i.e., the process obtained by centering the observed variable by the conditional mean and then dividing by the conditional standard deviation). In nearly all of these papers, the rescaled errors are assumed to be independent and identically distributed (i.i.d.). This is the case, for instance, in Berkes et al. (2003), Francq and Zakoïan (2004), and Straumann and Mikosch (2006), which are the papers closest to ours in their method of proof. Serial dependence in the rescaled errors is allowed by Lee and Hansen (1994), who assume them to form a stationary and ergodic process, and, quite recently, also by Escanciano (2009) and Linton, Pan, and Wang (2010). All three of these papers are confined to the linear pure GARCH case.

In this paper we consider QML estimation in autoregressive models with GARCH errors and allow both the conditional mean and conditional variance to be nonlinear. Specifically, the conditional mean can be a general nonlinear autoregression of order p (AR(p)), whereas the conditional variance is specified as a general nonlinear first-order GARCH model (GARCH(1,1)). Moreover, the rescaled errors are not required to be independent, but instead to form a stationary and ergodic martingale difference sequence. 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. We have decided to leave the extension to higher-order GARCH models for future research because the technical difficulties are considerable already in the first-order case. An instance of such difficulties are conditions under which certain stochastic difference equations possess stationary ergodic solutions (one is led to such considerations when examining the conditional variance process in detail, see Proposition 1 in Section 2 and Propositions 2 and 3 in Section 4). By confining ourselves to the leading case of GARCH(1,1) models, we are able to present the required rather intricate theory in a relatively transparent way and give the required conditions in explicit and easily verifiable forms. The extension to the higher-order case is feasible but leads to rather unintuitive formulations and conditions that appear very difficult to verify. Another instance of the arising difficulties is that our results are not directly applicable to the EGARCH model and the so-called log–GARCH model and 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 references using a similar approach are 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, and Kristensen and Rahbek (2009), where nonlinear ARCH models are considered. 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 also extend the asymptotic estimation theory of homoskedastic nonlinear autoregressions. In addition to Chan and Tong (1986), already mentioned, one can refer to Tjøstheim (1986), who derives asymptotic properties of least squares and QML estimators in general nonlinear autoregressions (to some extent, conditional heteroskedasticity is also allowed for but 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, 1991b). 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 with independent and identically distributed rescaled errors, 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 with martingale difference errors 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 the Appendixes.

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 $x = (x_1, \ldots, x_n)$ 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 $\|X\|_p = \left(\mathrm{E}\left[|X|^p\right]\right)^{1/p}$, 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

Suppose our interest is to model a univariate stationary time series and, especially, its conditional mean and conditional variance. We wish to consider a fairly general model and, therefore, our introductory discussion is partly informal and involves assumptions that will be weakened and made precise later.

Many of the models used so far in the literature assume a data generation process y_t that can be described by the general equation

$$y_t = f(y_{t-1}, \ldots, y_{t-p}; \mu_0) + \sigma_t \varepsilon_t, \qquad t = 1, 2, \ldots, \tag{1}$$

where $f(y_{t-1}, \ldots, y_{t-p}; \mu_0)$ and σ_t^2 represent the conditional mean and conditional variance, respectively; ε_t is an unobservable error term; and the (positive) volatility σ_t depends on the variables $\{y_s, s < t\}$. This discussion obviously assumes that the data generation process has finite variance and that suitable assumptions are imposed on the error term. For instance, it has been common to make the following assumption.

The random variables ε_t are i.i.d. with zero mean and unit variance, and such that ε_t is independent of the variables $\{y_s, s < t\}$.

This "i.i.d. assumption" is not necessary for our general results but appears convenient for expository purposes. As will be seen later, the i.i.d. assumption may also be of interest in our subsequent developments, for it can be used to weaken the moment conditions needed to permit dependence in the error term ε_t. The assumptions required for the error term (Assumption E) will be introduced later in this section.

Equation (1) characterizes the conditional mean as a general nonlinear function of p lagged values of y_t and the $m \times 1$ parameter vector μ_0. 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

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

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)–(3) 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. Unlike in the preceding discussion, existence of finite variance or even finite mean will not be assumed although, for convenience, we continue to use the terms conditional mean and conditional variance. Specifically, we make the following assumption.

Assumption DGP. The process (y_t, σ_t^2) defined by equations (1)–(3) is stationary and ergodic, with $\mathrm{E}[|y_t|^{2r}] < \infty$ and $\mathrm{E}[\sigma_t^{2r}] < \infty$ for some $r > 0$.

This is a high-level assumption that can be checked by using results available in the literature. A discussion of this issue is provided shortly after completing the model specification and discussing conditions required for the error term ε_t. Letting $\mathcal{F}_t = \sigma(y_t, y_{t-1}, \ldots)$ denote the σ–algebra generated by present and past observations, we impose the following assumption.

Assumption E. The random variables ε_t satisfy $\mathrm{E}[\varepsilon_t^2] < \infty$ for all t. Moreover, $\mathrm{E}[\varepsilon_t \mid \mathcal{F}_{t-1}] = 0$ a.s. and $\mathrm{E}[\varepsilon_t^2 \mid \mathcal{F}_{t-1}] = 1$ a.s.

As stated, this assumption alone is not very informative. It becomes more transparent when combined with Assumption DGP and Assumptions C1–C3, to be imposed in the next section. Using these assumptions one can justify that both $f(y_{t-1}, \ldots, y_{t-p}; \mu_0)$ and σ_t^2 are stationary, ergodic, and $\mathcal{F}_{t-1}$–measurable (see Proposition 1), which in conjunction with equation (1) and Assumption E implies that ε_t is a stationary and ergodic martingale difference sequence and that the conditional mean and variance of y_t (when they exist) are equal to $f(y_{t-1}, \ldots, y_{t-p}; \mu_0)$ and σ_t^2, respectively. Thus, Assumption E (together with the aforementioned other assumptions) enables us to somewhat weaken the i.i.d. assumption. Previously, conditions similar to Assumption E have been employed by Lee and Hansen (1994) and Escanciano (2009) to develop estimation theory for linear GARCH models (see also Linton et al., 2010).

We now discuss sufficient conditions for Assumption DGP. Such conditions were recently obtained by Meitz and Saikkonen (2008c) by assuming the conditional mean function is of the form

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

where $a(z; \mu_0) = (a_1(z; \mu_0), \ldots, a_p(z; \mu_0))$ and $b(z; \mu_0)$ are nonlinear bounded functions ($z \in \mathbb{R}^p$). Using theory developed for Markov chains, Meitz and Saikkonen (2008c) give conditions for (geometric) ergodicity and existence of moments in general AR–GARCH models of this type. For their results to hold, they have to assume (in addition to a number of technical assumptions) that the error term ε_t satisfies the i.i.d. assumption and has a positive and lower semicontinuous (Lebesgue) density on $\mathbb{R}$. The latter requirement is more than needed in some recent work on the estimation of GARCH and (linear) ARMA–GARCH models (see Berkes et al., 2003; Francq and Zakoïan, 2004; and Straumann and Mikosch, 2006). Meitz and Saikkonen (2008c) also need rather stringent smoothness conditions on the nonlinear functions in (2) and (4). Such conditions are not needed by

Cline (2007), who also uses Markov chain theory to establish (geometric) ergodicity in nonlinear AR–GARCH models. Similarly to Meitz and Saikkonen (2008c) he also needs the i.i.d. assumption on the error term. Cline considers a very general model, but his assumptions are not easy to check. Indeed, Cline only verifies all of 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 (2008c), Assumption DGP can be justified for several widely used models. The conditional mean can be 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. Besides the standard linear GARCH model, the conditional variance can be a smooth transition GARCH model proposed by 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 (2008c), although this may be difficult for general nonlinear models. 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; Zakoïan, 1994). In this particular example Assumption DGP can be verified without the i.i.d. assumption, but in general doing so in models involving nonlinearity appears to be very difficult.

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 are needed (see Assumption N4, in Section 5, and the discussion following it).

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 et al. (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 (Sect. 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, 2005, 2009).

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, 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

$$h_0(\theta) = \varsigma_0 \quad \text{and} \quad h_t(\theta) = g(u_{t-1}, h_{t-1}(\theta); \theta), \qquad t = 1, 2, \ldots, \tag{5}$$

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 $h_t(\theta)$, $t = 1, 2, \ldots$, 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.[1] 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 shorthand 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 (2008c) 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 process $h_t(\theta)$ for parameter values different from the true value θ_0, but the results of Meitz and Saikkonen (2008c) 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, 2005, 2009).

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 $C < \infty$, and all $u \in \mathbb{R}$, $x \in \mathbb{R}_+$, and $\theta \in \Theta$, $g(u,x;\theta) \leq \varrho x + C(1+u^2)$.
(ii) For some $0 < \kappa < 1$, and all $u \in \mathbb{R}$, $x_1, x_2 \in \mathbb{R}_+$, and $\theta \in \Theta$, $|g(u,x_1;\theta) - g(u,x_2;\theta)| \leq \kappa |x_1 - x_2|$.

Assumption C3. The function $f : \mathbb{R}^p \times \mathsf{M} \to \mathbb{R}$ is such that $f(\cdot;\mu)$ is Borel measurable for every μ and $f(z;\cdot)$ is continuous for every $z \in \mathbb{R}^p$. Furthermore, for some $C < \infty$ and all $z \in \mathbb{R}^p$ and $\mu \in \mathsf{M}$, $|f(z_1,\ldots,z_p;\mu)| \leq C\left(1+\sum_{j=1}^p |z_j|\right)$.

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 even when the i.i.d. assumption is used for ε_t (see Assumption 4 in Meitz and Saikkonen, 2008c). This particularly holds for the Lipschitz condition in Assumption C2(ii), which is restrictive in that it rules out the so-called EGARCH and log-GARCH models. One might consider relaxing this condition along the lines in Straumann and Mikosch (2006), but this does not seem straightforward. For instance, allowing κ in Assumption C2(ii) to depend on u leads to additional technical difficulties because in our proofs we have to replace u by u_t and, unlike in the pure GARCH case, our u_t also depends on the parameter μ. This dependence makes the verification of the resulting condition more difficult than in the pure GARCH case, and it may also necessitate imposing additional restrictions on the conditional mean. Assumption C3 appears fairly mild. The measurability and continuity requirements are common in nonlinear estimation problems, and the dominance condition holds, for example, for several functional-coefficient autoregressive models, including the exponential autoregressive model and various smooth transition autoregressive models.

Using Assumptions C1–C3 we can prove the following result, which will be used to prove the consistency and asymptotic normality of the QML estimator of the parameter vector θ_0.

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*

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

This solution is continuous in θ, $\mathcal{F}_{t-1}$*–measurable, and unique when (6) is extended to all* $t \in \mathbb{Z}$. *Furthermore, the solution* $h_t^*(\theta)$ *has the properties* $h_t^*(\theta_0) = \sigma_t^2$ *a.s. and* $\mathrm{E}\left[\sup_{\theta\in\Theta} h_t^{*r}(\theta)\right] < \infty$, *and, if* $h_t(\theta)$, $\theta \in \Theta$, *are any other solutions to 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 (a.s.) 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 (2008c). 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.

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)–(3) (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 $h_t(\theta)$ $(t = 1, \ldots, T)$. It follows from Proposition 1 that $h_t(\theta)$ approximates the stationary solution to equation (6), which for $\theta = \theta_0$ coincides (a.s.) with the true conditional variance σ_t^2. We also define

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

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.

The continuity of the functions f and g imposed in Assumptions C2 and C3 ensures that the Gaussian log-likelihood function $L_T(\theta)$ is continuous, which in conjunction with the assumed compactness of the parameter space Θ 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, Lem. 3.4). In view of the continuity of $h_t^*(\theta)$ established in Proposition 1, the same is true for a minimizer of $L_T^*(\theta)$.

In addition to the assumptions already made, we have to supplement Assumption C2 concerning the conditional variance by the following technical condition.

Assumption C4. 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 requires the function g to be bounded away from zero in the same way as, for example, Assumption C.3 of Straumann and Mikosch (2006). Such a requirement is somewhat unnatural and, like Assumption C2(ii), rules out the log-GARCH model. However, it appears very useful in the proofs. For instance, it facilitates proving convergence results for $l_t^*(\theta)$ and other quantities and is also used to show the finiteness of certain moments. In pure ARCH models this condition can be relaxed (cf. Condition C.3 in Kristensen and Rahbek, 2009), but doing so here would require strengthening of other assumptions.

Consistency of the QML estimator $\hat{\theta}_T$ also requires the following identification condition.

Assumption C5.

(i) $f(y_{t-1},\ldots,y_{t-p};\mu) = f(y_{t-1},\ldots,y_{t-p};\mu_0)$ a.s. only if $\mu = \mu_0$.
(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 C5 is equivalent to $\mathrm{E}[L_T^*(\theta)]$ 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, Def. 3.1).[2] Although more explicit than an identifiable uniqueness condition, the conditions in Assumption C5 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 Assumption C5(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 Assumption C5(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 challenging to verify Assumption C5(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 Assumption C5(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 Assumption C5(i) clearly holds if the event $\{(y_{t-1},\ldots,y_{t-p}) \in A\}$ 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 Assumption C5(ii), it agrees with the identification condition used by Straumann and Mikosch (2006) in their nonlinear GARCH model. However, in their examples they do not consider nonlinearities as complicated as those we consider. Therefore, unlike in one of our examples, they do not need to rely on Markov chain theory to verify their counterpart of Assumption C5(ii). So far, Straumann and Mikosch (2006) seems to be the only published paper dealing with identification in nonlinear GARCH models. However, identification in nonlinear ARCH models has recently been considered by Kristensen and Rahbek (2009). These authors also use Markov chain techniques to verify identification conditions similar to Assumption C5(ii), but their approach is quite different from ours. In particular, Kristensen and Rahbek (2009) also make use of the differentiability of $h_t^*(\theta)$, which we do not assume; see, e.g., proof of their Corollary 2.

As a final remark we note that in the verification of Assumption C5 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 nondegenerate, and a similar condition also appears in Straumann and Mikosch (2006, Thms. 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, E, and C1–C5 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 et al. (2003), Francq and Zakoïan (2004), Straumann and Mikosch (2006), and Escanciano (2009). 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 Section 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). In this case, the resulting asymptotic distribution is no longer normal. We leave this for future research.) Combined with the consistency of the QML estimator, Assumption N1 also implies that in the subsequent analysis we 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 of the objective functions $L_T(\theta)$and $L_T^*(\theta)$ and their components, including the processes h_t and h_t^*.

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

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

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 f and g. Denote $f_\mu = \partial f(z;\mu)/\partial\mu$, $f_{\mu\mu} = \partial^2 f(z;\mu)/\partial\mu\partial\mu'$, and the first and second partial derivatives of g with $g_{v_1} = \partial g(u,h;\theta)/\partial v_1$ and $g_{v_1 v_2} = \partial^2 g(u,h;\theta)/\partial v_1 \partial v_2'$, where v_1 and v_2 can be any of u, h, or θ.

Assumption N3.

(i) For some $C < \infty$ and all $z \in \mathbb{R}^p$ and $\mu \in \mathsf{M}_0$, the quantities $|f_\mu|$ and $|f_{\mu\mu}|$ (evaluated at $(z;\mu)$) are bounded by $C(1+\sum_{j=1}^p |z_j|)$.

(ii) For some $C < \infty$ and all $u \in \mathbb{R}$, $x \in \mathbb{R}_+$, and $\theta \in \Theta_0$, 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 $u \in \mathbb{R}$, $x_1, x_2 \in \mathbb{R}_+$, and $\theta \in \Theta_0$,

$$|g_v(u,x_1;\theta) - g_v(u,x_2;\theta)| \le \kappa'|x_1 - x_2|, \qquad v = u,h,\theta,$$

$$|g_{v_1 v_2}(u,x_1;\theta) - g_{v_1 v_2}(u,x_2;\theta)| \le \kappa'|x_1 - x_2|, \qquad v_1, v_2 = u,h,\theta.$$

Assumption N3(i) places further restrictions on the behavior of the function f that specifies the conditional mean. Like the dominance condition already imposed on the function f in Assumption C3, this condition may be stringent from a mathematical point of view but holds for various commonly used functional-coefficient autoregressive models of the type (4). The second and third parts of Assumption N3 are related to Assumptions 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). Assumption 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 $g_{\upsilon_1,t} = [g_{\upsilon_1}]_{u=u_{t-1}(\theta),h=h_{t-1}(\theta)} = \partial g\left(u_{t-1}(\theta), h_{t-1}(\theta);\theta\right)/\partial \upsilon_1$ 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 $h_t = g(u_{t-1}, h_{t-1};\theta)$, $t = 1,2,\ldots$, can now be derived by straightforward but tedious differentiation. We have

$$h_{\theta,t} = g_{\theta,t} - g_{u,t} f_{\theta,t-1} + g_{h,t} h_{\theta,t-1}, \qquad t = 1,2,\ldots,$$

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

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

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

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

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

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

$$\begin{aligned} h_{\theta\theta,t} &= \alpha_{\theta\theta,t} + \beta_t h_{\theta\theta,t-1} + \gamma_{\theta,t} h_{\theta,t-1}' + h_{\theta,t-1}\gamma_{\theta,t}' + \delta_t h_{\theta,t-1} h_{\theta,t-1}', \\ &\qquad t = 1,2,\ldots. \end{aligned} \tag{10}$$

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–C5, and N1–N3 hold.*

(a) For all $\theta \in \Theta_0$ there exists a stationary ergodic solution $h_{\theta,t}^{*}(\theta)$ to the equation

$$h_{\theta,t}(\theta) = \alpha_{\theta,t}^{*} + \beta_t^{*} h_{\theta,t-1}(\theta), \qquad t = 1,2,\ldots. \tag{11}$$

This solution is $\mathcal{F}_{t-1}$–measurable, unique when (11) is extended to all $t \in \mathbb{Z}$, and $\mathrm{E}\left[\sup_{\theta\in\Theta_0}|h^*_{\theta,t}(\theta)|^{r/2}\right] < \infty$. *Furthermore, the stationary ergodic solution $h^*_t(\theta)$ obtained from Proposition 1 is a.s. continuously partially differentiable on Θ_0 for every $t \in \mathbb{Z}$ and $\partial h^*_t(\theta)/\partial\theta = h^*_{\theta,t}(\theta)$ a.s.*

(b) *If $h_t(\theta)$ and $h_{\theta,t}(\theta)$, $\theta \in \Theta_0$, are any solutions to the difference equations (6) and (9), 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 (a.s.) continuously differentiable and that its derivative coincides (a.s.) 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–C5, 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*

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

This solution is $\mathcal{F}_{t-1}$–measurable, unique when (12) is extended to all $t \in \mathbb{Z}$, and $\mathrm{E}\left[\sup_{\theta\in\Theta_0}|h^*_{\theta\theta,t}(\theta)|^{r/4}\right] < \infty$. *Furthermore, the stationary ergodic solution $h^*_t(\theta)$ obtained from Proposition 1 is a.s. twice continuously partially differentiable on Θ_0 for every $t \in \mathbb{Z}$ and $\partial^2 h^*_t(\theta)/\partial\theta\partial\theta' = h^*_{\theta\theta,t}(\theta)$ a.s.*

(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), 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. Note that in the moment and convergence results obtained for $h^*_{\theta,t}$ and $h^*_{\theta\theta,t}$ in Propositions 2 and 3, 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, Props. 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. 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. Our next assumption imposes further restrictions on the moments of the observed process and the derivatives of the process $h_t^*(\theta)$.

Assumption N4.

(i) Assumption DGP holds with $r = 2$ and the random variables ε_t satisfy $\mathrm{E}[\varepsilon_t^8] < \infty$.

(ii) $\left\| \sup_{\theta \in \Theta_0} \frac{|h_{\theta,t}^*(\theta)|}{h_t^*(\theta)} \right\|_4 < \infty$ and $\left\| \sup_{\theta \in \Theta_0} \frac{|h_{\theta\theta,t}^*(\theta)|}{h_t^*(\theta)} \right\|_2 < \infty$.

The conditions in Assumption N4(i) imply that finiteness of fourth moments is assumed for the observed process y_t, which is much more than needed to prove consistency. However, proving asymptotic normality of the QML estimator without this assumption has proved difficult even in the linear ARMA–GARCH case (see the discussions in Francq and Zakoïan, 2004; Ling, 2007). (In the pure GARCH case the situation is different, for then no additional moment conditions on the observed process are required; see Francq and Zakoïan, 2004; Straumann and Mikosch, 2006.) If one is willing to make the i.i.d. assumption the moment condition in Assumption N4(i) can be weakened to $\mathrm{E}[\varepsilon_t^4] < \infty$. Finiteness of eighth moments is needed to ensure that the limiting distribution of the QML estimator has a finite covariance matrix when the errors are dependent and only satisfy Assumption E (see the proof of Lemma D.1 in Appendix D). (An alternative to assuming finite eighth moments is to require that $\mathrm{E}[\varepsilon_t^4] < \infty$ and $\mathrm{E}[\varepsilon_t^4 \mid \mathcal{F}_{t-1}] \le K < \infty$ a.s.; cf. Assumption A.2(i) of Lee and Hansen, 1994.) In this respect, the situation is easier in the case of linear pure GARCH models where similar dependence in the errors can be allowed by assuming only $\mathrm{E}[|\varepsilon_t|^{4+\delta}] < \infty$ for some $\delta > 0$ (see Escanciano, 2009). The moment conditions imposed on the derivatives of h_t^* in Assumption N4(ii) are satisfied when the i.i.d. assumption holds and 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; Ling, 2007). 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

$$\mathcal{I}(\theta_0) \stackrel{def}{=} \mathrm{E}\left[\frac{\partial L_T^*(\theta_0)}{\partial \theta}\frac{\partial L_T^*(\theta_0)}{\partial \theta'}\right] \quad \text{and} \quad \mathcal{J}(\theta_0) \stackrel{def}{=} \mathrm{E}\left[\frac{\partial^2 L_T^*(\theta_0)}{\partial \theta \partial \theta'}\right].$$

Explicit expressions for these matrices are given in Theorem 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. This is guaranteed by the following three conditions.

Assumption N5.

(i) For all t, the conditional distribution of ε_t given $\mathcal{F}_{t-1}$ 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, Assum. 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, it seems difficult to replace them with more transparent counterparts. However, in the case of a standard linear AR–GARCH(1,1) model, these two conditions are automatically satisfied provided that Assumption N5(i) holds and homoskedasticity is ruled out (see Appendix E, Example 1). For a model containing both a conditional mean and a conditional variance, Assumption 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 $\hat{\theta}_T$. If one makes the i.i.d. assumption, the unconditional counterpart of Assumption N5(i) suffices; this condition was also used by Francq and Zakoïan (2004) in the context of their linear ARMA–GARCH model. In the context of a nonlinear GARCH model, a condition at least as strong as Assumption N5(i) often may be needed to ensure that condition N5(iii) holds. We will return to this in the concrete examples of the next section.

Verifying Assumptions 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 C5, and we only mention that we have been forced to use the i.i.d. assumption and Markov chain techniques in order to be able to verify them. A previous example of this kind of approach in the context of a homoskedastic smooth transition autoregressive model is provided by Chan and Tong (1986, App. II) (see also Tjøstheim, 1986, Sect. 4.1; Kristensen and Rahbek, 2009).

Now we can state the main result of this section.

THEOREM 2. *Suppose Assumptions DGP, E, C1–C5, 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 positive definite and can be expressed as

$$\mathcal{I}(\theta_0) = \begin{bmatrix} 4\mathrm{E}\left[\frac{f_{\mu,t}(\mu_0)}{\sigma_t}\frac{f'_{\mu,t}(\mu_0)}{\sigma_t}\right] & 0_{m\times l} \\ 0_{l\times m} & 0_{l\times l} \end{bmatrix} + \mathrm{E}\left[(\varepsilon_t^4 - 1)\frac{h^*_{\theta,t}(\theta_0)}{\sigma_t^2}\frac{h^{*\prime}_{\theta,t}(\theta_0)}{\sigma_t^2}\right] \tag{13}$$
$$+2\begin{bmatrix} \mathrm{E}\left[\varepsilon_t^3\left(\frac{f_{\mu,t}(\mu_0)}{\sigma_t}\frac{h^{*\prime}_{\mu,t}(\theta_0)}{\sigma_t^2} + \frac{h^*_{\mu,t}(\theta_0)}{\sigma_t^2}\frac{f'_{\mu,t}(\mu_0)}{\sigma_t}\right)\right] & \mathrm{E}\left[\varepsilon_t^3\frac{f_{\mu,t}(\mu_0)}{\sigma_t}\frac{h^{*\prime}_{\lambda,t}(\theta_0)}{\sigma_t^2}\right] \\ \mathrm{E}\left[\varepsilon_t^3\frac{h^*_{\lambda,t}(\theta_0)}{\sigma_t^2}\frac{f'_{\mu,t}(\mu_0)}{\sigma_t}\right] & 0_{l\times l} \end{bmatrix},$$

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

As in the consistency proof, we shall follow Berkes et al. (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). The covariance matrix of the limiting distribution obtained in the theorem simplifies if the i.i.d. assumption holds (and even more if the error term ε_t also has a symmetric distribution). Note also that then one can change the assumptions in Theorem 2 so that Assumption E is deleted and in Assumption N4(i) the condition $\mathrm{E}[\varepsilon_t^8] < \infty$ is replaced by $\mathrm{E}[\varepsilon_t^4] < \infty$ (for details on the i.i.d. case, see Meitz and Saikkonen, 2008b).

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 (13) and (14) reveal that it suffices to find consistent estimators for

$$\mathrm{E}\left[\frac{f_{\mu,t}(\mu_0)}{\sigma_t}\frac{f'_{\mu,t}(\mu_0)}{\sigma_t}\right], \qquad \mathrm{E}\left[\frac{h^*_{\theta,t}(\theta_0)}{\sigma_t^2}\frac{h^{*\prime}_{\theta,t}(\theta_0)}{\sigma_t^2}\right],$$
$$\mathrm{E}\left[\varepsilon_t^4\frac{h^*_{\theta,t}(\theta_0)}{\sigma_t^2}\frac{h^{*\prime}_{\theta,t}(\theta_0)}{\sigma_t^2}\right], \quad \text{and} \quad \mathrm{E}\left[\varepsilon_t^3\frac{f_{\mu,t}(\mu_0)}{\sigma_t}\frac{h^{*\prime}_{\theta,t}(\theta_0)}{\sigma_t^2}\right]. \quad (15)$$

The obvious choices are

$$T^{-1}\sum_{t=1}^{T}\frac{\hat{f}_{\mu,t}}{\hat{h}_t^{1/2}}\frac{\hat{f}'_{\mu,t}}{\hat{h}_t^{1/2}}, \qquad T^{-1}\sum_{t=1}^{T}\frac{\hat{h}_{\theta,t}}{\hat{h}_t}\frac{\hat{h}'_{\theta,t}}{\hat{h}_t},$$
$$T^{-1}\sum_{t=1}^{T}\frac{\hat{u}_t^4}{\hat{h}_t^2}\frac{\hat{h}_{\theta,t}}{\hat{h}_t}\frac{\hat{h}'_{\theta,t}}{\hat{h}_t}, \quad \text{and} \quad T^{-1}\sum_{t=1}^{T}\frac{\hat{u}_t^3}{\hat{h}_t^{3/2}}\frac{\hat{f}_{\mu,t}}{\hat{h}_t^{1/2}}\frac{\hat{h}'_{\theta,t}}{\hat{h}_t}, \quad (16)$$

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 $\hat{\mathcal{J}}_T$. For these estimators to be consistent, additional assumptions are needed. It can be shown that, under the conditions of Theorem 2 and the additional requirement that Assumption DGP holds with $r = 4$,

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

(details are available on request). 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 $\hat{\mathcal{J}}_T^{-1}\hat{\mathcal{I}}_T\hat{\mathcal{J}}_T^{-1}$. As discussed after Theorem 2, if the i.i.d. assumption holds, the expression of $\mathcal{I}(\theta_0)$ simplifies, which can accordingly be taken into account in its estimation. Consistency of the resulting estimators of $\mathcal{I}(\theta_0)$ and $\mathcal{J}(\theta_0)$ then also holds under the mentioned weakened assumptions of Theorem 2 (for details, see Meitz and Saikkonen, 2008b).

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, E, C1–C5, and N1–N5. That the stated conditions imply these assumptions is shown in 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

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

where $u_{0,t} = y_t - (\phi_{0,0} + \sum_{j=1}^{p} \phi_{0,j} y_{t-j}) = \sigma_t \varepsilon_t$. The parameter vectors μ and λ are $\mu = (\phi_0,\dots,\phi_p)$ and $\lambda = (\omega,\alpha,\beta)$, and the permissible parameter spaces M and Λ 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. Note that our definition of the parameter space includes the restriction that $\beta < 1$ over Θ.

Let $\mathcal{F}_t^{\varepsilon} = \sigma(\varepsilon_t,\varepsilon_{t-1},\dots)$ denote the σ–algebra generated by present and past errors, and consider the following set of conditions.

(a) (i) The random variables ε_t are stationary and ergodic.
(ii) $\mathrm{E}\left[\ln(\beta_0 + \alpha_0 \varepsilon_t^2)\right] < 0$.
(iii) $\mathrm{E}\left[|\varepsilon_t|^{2r}\right] < \infty$ and $\mathrm{E}\left[(\beta_0 + \alpha_0 \varepsilon_t^2)^r \mid \mathcal{F}_{t-1}^{\varepsilon}\right] \le C < 1$ a.s. for some $r > 0$.
(iv) $1 - \sum_{j=1}^{p} \phi_{0,j} z^j \neq 0$, $|z| \le 1$.
(v) $\mathrm{E}[\varepsilon_t^2] < \infty$, $\mathrm{E}\left[\varepsilon_t \mid \mathcal{F}_{t-1}^{\varepsilon}\right] = 0$ a.s., and $\mathrm{E}\left[\varepsilon_t^2 \mid \mathcal{F}_{t-1}^{\varepsilon}\right] = 1$ a.s.

(b) (i) ε_t^2 has a nondegenerate distribution.
(ii) $\alpha_0 > 0$.

(c) (i) The true parameter value θ_0 is an interior point of Θ.
(ii) $\mathrm{E}[\varepsilon_t^4] < \infty$ and $\mathrm{E}\left[(\beta_0 + \alpha_0 \varepsilon_t^2)^2 \mid \mathcal{F}_{t-1}^{\varepsilon}\right] \le C < 1$ a.s.
(iii) $\mathrm{E}[\varepsilon_t^8] < \infty$.
(iv) For all t, the conditional distribution of ε_t given $\mathcal{F}_{t-1}^{\varepsilon}$ is not concentrated at two points.

The five conditions in part (a) imply the validity of Assumptions DGP and E for the linear AR(p)–GARCH(1,1) model as defined above (for details of this and the following statements, see Appendix E). Of these conditions, (a.i) and (a.ii) ensure the existence of a (strictly) stationary and ergodic solution for the conditional variance process, whereas (a.iii) guarantees that this solution has moments of some (small) order. Condition (a.iv) is needed for these properties to carry over to the observed process y_t and Assumption DGP to hold. Assumption E holds when condition (a.v) is added. If the conditions in part (b) are also assumed, Assumptions C1–C5 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 QML estimator to hold. The second condition ensures that the conditional variance process, and hence also y_t, has finite fourth moments. The moment condition in (c.iii) coincides with the one in Assumption N4(i). Finally, (c.iv) is needed for the identification condition N5 to hold.

If one makes the i.i.d. assumption, conditions (a.i), (a.iii), (a.v), and (c.iii) can be dropped, and (c.ii) and (c.iv) can be replaced by their unconditional counterparts ("$\mathrm{E}[\varepsilon_t^4] < \infty$ and $\mathrm{E}\left[\left(\beta_0 + \alpha_0 \varepsilon_t^2\right)^2\right] < 1$" and "the distribution of ε_t is not concentrated at two points"), see Meitz and Saikkonen (2008b) for details. In this case, the resulting conditions (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.[3] Therefore, although our framework allows for rather general forms of nonlinearity and dependence in the errors, 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 et al. (1993). For this model we are able to show strong consistency, but not asymptotic 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

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

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 et al. (1993) and Zakoïan (1994) are included in the AGARCH model.

Consider the following set of conditions.

(a) (i) The random variables ε_t are stationary and ergodic.
(ii) $\mathrm{E}\left[\ln\left(\beta_0+\alpha_0(|\varepsilon_t|-\gamma_0\varepsilon_t)^2\right)\right] < 0$.
(iii) $\mathrm{E}\left[|\varepsilon_t|^{2r}\right] < \infty$ and $\mathrm{E}\left[\left(\beta_0+\alpha_0(|\varepsilon_t|-\gamma_0\varepsilon_t)^2\right)^r \mid \mathcal{F}_{t-1}^{\varepsilon}\right] \leq C < 1$ a.s. for some $r > 0$.
(iv) $1-\sum_{j=1}^{p}\phi_{0,j}z^j \neq 0$, $|z| \leq 1$.
(v) $\mathrm{E}[\varepsilon_t^2] < \infty$, $\mathrm{E}\left[\varepsilon_t \mid \mathcal{F}_{t-1}^{\varepsilon}\right] = 0$ a.s., and $\mathrm{E}\left[\varepsilon_t^2 \mid \mathcal{F}_{t-1}^{\varepsilon}\right] = 1$ a.s.

(b) (i) For all t, the conditional distribution of ε_t given $\mathcal{F}_{t-1}^{\varepsilon}$ is not concentrated at two points.
(ii) $\alpha_0 > 0$.

Conditions (a.i)–(a.v) ensure the validity of Assumptions DGP and E for the AR–AGARCH model and are analogous to the ones used in Example 1. Altogether the conditions in (a) and (b) ensure that Assumptions C1–C5 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 Assumption C5(ii). The conditions again simplify if one makes the i.i.d. assumption. Then (a.i), (a.iii), and (a.v) can be dropped, and (b.i) can be replaced by its unconditional counterpart (see Meitz and Saikkonen, 2008b, for details).

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 (and under the i.i.d. assumption) 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 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 (2008c). Using similar arguments, other models of interest could also be considered.

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

$$\begin{aligned} f\left(y_{t-1},\ldots,y_{t-p};\mu_0\right) &= \phi_{0,0}+\psi_{0,0}F(y_{t-d};\varphi_0) \\ &\quad + \sum_{j=1}^{p}\left(\phi_{0,j}+\psi_{0,j}F(y_{t-d};\varphi_0)\right)y_{t-j} \quad \text{and} \\ \sigma_t^2 &= g(u_{0,t-1},\sigma_{t-1}^2;\theta_0) \\ &= \omega_0+\left(\alpha_{0,1}+\alpha_{0,2}G(u_{0,t-1};\gamma_0)\right)u_{0,t-1}^2+\beta_0\sigma_{t-1}^2, \qquad \textbf{(19)} \end{aligned}$$

where $u_{0,t} = y_t - f(y_{t-1},\ldots,y_{t-p};\mu_0) = \sigma_t\varepsilon_t$, $\varphi_0 = (\varphi_{0,1},\varphi_{0,2})$, 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 $\lambda = (\omega,\alpha_1,\alpha_2,\beta,\gamma_1,\gamma_2)$, respectively, 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. 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,

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

$$G(u;\gamma_1,\gamma_2) = \left[1+\exp\left(-\gamma_2(u-\gamma_1)\right)\right]^{-1},$$

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 straightforward to verify that these conditions are satisfied with the choice of logistic functions (or, e.g., the Gaussian cumulative distribution functions). In the following we assume that F and G satisfy the additional conditions given in Appendix E.

To present the conditions for this model we require 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

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

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

where I_{p-1} denotes the identity matrix and 0_{p-1} a vector of zeros. 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 $\mathcal{A}^k = \{A_1A_2\cdots A_k : A_i \in \mathcal{A},\ i = 1,\ldots,k\}$ 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 (2008c).

Now consider the following set of conditions.

(a) (i) The random variables ε_t are i.i.d. with zero mean and unit variance, and such that ε_t is independent of the variables $\{y_s,\ s < t\}$.
(ii) The ε_t have a (Lebesgue) density that is positive and lower semicontinuous on $\mathbb{R}$.
(iii) Either $\sum_{j=1}^{p}\max\left\{|\phi_{0,j}|, |\phi_{0,j}+\psi_{0,j}|\right\} < 1$ or $\rho(\{A_{01}, A_{02}\}) < 1$.
(iv) $\mathrm{E}\left[\log\left(\beta_0 + (\alpha_{0,1}+\alpha_{0,2})\varepsilon_t^2\right)\right] < 0$.
(v) $\alpha_{0,1} > 0$ 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) $\mathrm{E}[\varepsilon_t^4] < \infty$ and $\mathrm{E}\left[\left(\beta_0 + (\alpha_{0,1}+\alpha_{0,2})\varepsilon_t^2\right)^2\right] < 1$.

Conditions (a.i)–(a.v) ensure the validity of Assumptions DGP and E in the case of the considered nonlinear AR–GARCH model. Conditions (a.i) and (a.ii) restrict 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 (2008c). In particular, we make the i.i.d. assumption (a.i). Conditions (a.i) and (a.ii) also facilitate the verification of the identification conditions in Assumptions C5 and N5. As our discussion following Assumption C5 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 $\{(y_{t-1},\ldots,y_{t-p}) \in A\}$ have a positive probability with suitably defined (Borel) measurable sets $A \subset \mathbb{R}^p$. Conditions (a.i) and (a.ii) will be critical in establishing this. A condition similar to (a.ii) is also used by Kristensen and Rahbek (2009, Cond. C.Z). The two alternative conditions in (a.iii) are both sufficient restrictions on the conditional mean needed to show the validity of Assumption DGP. They are used in Meitz and Saikkonen (2008c, Sect. 4) and, as discussed by Liebscher (2005, p. 682), the latter condition is strictly weaker than the former one. Condition (a.iv) is an unconditional analog of the moment conditions (a.ii) 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.v) excludes the ARCH case, but is required for the results in Meitz and Saikkonen (2008c) to hold. In many applications the estimate of β would typically be rather large (and close to unity), and hence condition (a.v) is not very restrictive in practice.[4]

If conditions (b.i) and (b.ii) are also assumed, Assumptions C1–C5 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. The former condition was already used in Example 1, whereas the latter is an unconditional analogue of condition (c.ii) used therein.

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 AR(p) models with conditionally heteroskedastic errors specified as a general nonlinear GARCH(1,1) model. The assumptions we needed for this theory are comparable to those previously used in linear ARMA–GARCH models and in nonlinear pure GARCH models.

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 (2007) 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.

NOTES

1. 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.

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

3. There appears to be only one small difference. In their Condition A8, 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.iv) 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.

4. The ARCH case could be treated separately, as is also mentioned in Meitz and Saikkonen (2008c, p. 465). We do not pursue this further and only note that in this case many of the required derivations simplify considerably.

REFERENCES

Berkes, I., L. Horváth, & 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.

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

Bollerslev, T., R.F. Engle, & D.B. Nelson (1994) ARCH models. In R.F. Engle & D.L. McFadden (eds.), *Handbook of Econometrics*, vol. IV, chap. 49. Elsevier.

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

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. & H. Tong (1986) On estimating thresholds in autoregressive models. *Journal of Time Series Analysis* 7, 179–190.

Chen, R. & 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.

Ding, Z., C.W.J. Granger, & 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.

Escanciano, J.C. (2009) Quasi-maximum likelihood estimation of semi-strong GARCH models. *Econometric Theory* 25, 561–570.

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

Francq, C. & 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, & 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. & T. Ozaki (1981) Modelling nonlinear random vibrations using an amplitude-dependent autoregressive time series model. *Biometrika* 68, 189–196.

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

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

Kristensen, D. & A. Rahbek (2009) Asymptotics of the QMLE for non-linear ARCH models. *Journal of Time Series Econometrics* 1(1), article 2.

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, & S.T. Jensen (2006) Estimation and Asymptotic Inference in the First Order AR–ARCH Model. Preprint No. 4, University of Copenhagen.

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

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

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

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

Linton, O., J. Pan, & H. Wang (2010) Estimation for a nonstationary semi-strong GARCH(1,1) model with heavy-tailed errors. *Econometric Theory* 26, 1–28.

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. & T. Teräsvirta (2002) Evaluating GARCH models. *Journal of Econometrics* 110, 417–435.

Meitz, M. & P. Saikkonen (2008a) Ergodicity, mixing, and existence of moments of a class of Markov models with applications to GARCH and ACD models. *Econometric Theory* 24, 1291–1320.

Meitz, M. & P. Saikkonen (2008b) Parameter Estimation in Nonlinear AR–GARCH Models. Department of Economics Discussion Paper Series no. 396, University of Oxford.

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

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

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

Pan, J., H. Wang, & H. Tong (2008) Estimation and tests for power-transformed and threshold GARCH models. *Journal of Econometrics* 142, 352–378.

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

APPENDIX A: Auxiliary Results

We shall first give two simple lemmas that are useful in several subsequent derivations. We omit their proofs, which are available from the authors on request.

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 $\Delta_{r,k} = \max\left\{1, k^{1/r-1}\right\}$.

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.

The following lemma presents a result that 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 $\vartheta \in K$, $|G(v, z; \vartheta)| \leq \bar{\varrho}\,|z| + \psi(|v|)$, 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 $(v, \vartheta) \in M_v \times K$, $\left|G(v, z_1; \vartheta) - G(v, z_2; \vartheta)\right| \leq \bar{\kappa}\,|z_1 - z_2|$ for some $0 < \bar{\kappa} < 1$ and all $z_1, z_2 \in M_z$.

By $\mathbb{C}(K, M_z)$ 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 $z_t^*(\vartheta)$ to the equation

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

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

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

Compared to Bougerol (1993, Thm. 3.1) and Straumann and Mikosch (2006, Thm. 2.8), Lemma A.3 is more specific although sufficient for the purpose of this paper. Unlike the above-mentioned theorems, 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 $\sup_{\vartheta\in K}\left|T^{-1}\sum_{t=1}^{T} z_t^*(\vartheta)-\mathrm{E}[z_t^*(\vartheta)]\right| \to 0$ a.s. when $r \geq 1$. Lemma A.3 also states that the solution z_t^* is unique, with which we mean that any two stationary solutions to (A.1) coincide a.s. Hence, z_t^* is uniquely defined up to an event with probability zero.

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 $G_t : \mathbb{C}(K, M_z) \to \mathbb{C}\big(K, M_z\big)$ 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 (A.1) is in $\mathbb{C}(K, M_z)$ and is a solution to the difference equation $x_t = G_t\left(x_{t-1}\right)$, $t \geq 1$. 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),

$$\sup_{\vartheta\in K}\left|G(v_{t-1}(\vartheta), x_1(\vartheta);\vartheta)-G(v_{t-1}(\vartheta), x_2(\vartheta);\vartheta)\right| \leq \bar{\kappa}\sup_{\vartheta\in K}|x_1(\vartheta)-x_2(\vartheta)|,$$

implying $|G_t(x_1)-G_t(x_2)|_K \leq \bar{\kappa}|x_1-x_2|_K$. Thus, $\rho(G_t)$ is a stationary and ergodic process bounded from above by $\bar{\kappa} < 1$. Using this fact and the assumptions imposed it is straightforward to check that assumptions (C1) and (C2) of Theorem 3.1 of Bougerol (1993) hold. Thus, the existence of a stationary ergodic solution $z_t^* \in \mathbb{C}(K, M_z)$ to (A.1) 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 (a.s.) 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 $\left(v_{t-1}(\vartheta), v_{t-2}(\vartheta), \ldots\right)$ (cf. Straumann and Mikosch, 2006, Prop. 2.6).

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

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

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

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

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

$$\left|z_{t,n}(x)\right|_K \leq \bar{\varrho}^{n+1} |x|_K + \sum_{j=0}^{n} \bar{\varrho}^j \sup_{\vartheta \in K} \psi\left(\left|v_{t-j-1}(\vartheta)\right|\right)$$

$$\leq |x|_K + \sum_{j=0}^{\infty} \bar{\varrho}^j \sup_{\vartheta \in K} \psi\left(\left|v_{t-j-1}(\vartheta)\right|\right).$$

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, $\mathrm{E}\left[|w_t|^r\right] < \infty$ where Lemma A.1 is also made use of. Hence, we can conclude that the collection of random variables $\left\{\left|z_{t,n}(x)\right|_K^r,\ n = 1, 2, \ldots\right\}$ is uniformly integrable (see Billingsley, 1995, p. 338). Thus, because $z_t^* = \lim_{n\to\infty} z_{t,n}(x)$ a.s. (in $\mathbb{C}(K, M_z)$) we also have $\lim_{n\to\infty} \left|z_{t,n}(x)\right|_K^r = \left|z_t^*\right|_K^r$ (a.s.) and the above-mentioned uniform integrability allows us to conclude that $\mathrm{E}[\left|z_t^*\right|_K^r]$ ($= \mathrm{E}[\sup_{\vartheta\in K} \left|z_t^*(\vartheta)\right|^r]$) is the finite limit of $\mathrm{E}[\left|z_{t,n}(x)\right|_K^r]$ (see Davidson, 1994, Thm. 12.8).

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

$$\sup_{\vartheta \in K} \left|G(v_{t-1}(\vartheta), z_{t-1}^*(\vartheta); \vartheta) - G(v_{t-1}(\vartheta), z_{t-1}(\vartheta); \vartheta)\right|^r$$
$$\leq \bar{\kappa}^r \sup_{\vartheta \in K} \left|z_{t-1}^*(\vartheta) - z_{t-1}(\vartheta)\right|^r,$$

or, in other words, $\left|z_t^* - z_t\right|_K^r \leq \bar{\kappa}^r \left|z_{t-1}^* - z_{t-1}\right|_K^r$. Continuing iteratively,

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

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 $M_v = \mathbb{R}$, $M_z = \mathbb{R}_+$, $K = \Theta$, $G = g$, $v_t = u_t = y_t - f\left(y_{t-1}, \ldots, y_{t-p}; \mu\right)$, and $z_t(\theta) = h_t(\theta) = g\big(u_{t-1}(\theta), h_{t-1}(\theta); \theta\big)$, it follows from Assumption C2 that Conditions G(i) and (ii) are satisfied with the function $\psi(x) = C(1 + x^2)$. Furthermore, the last condition in Assumption C3, Assumption DGP, and Lemma A.1 give $\left\|\sup_{\theta\in\Theta} |f_t|\right\|_{2r} < \infty$ and $\left\|\sup_{\theta\in\Theta} |u_t|\right\|_{2r} < \infty$, implying the moment condition $\mathrm{E}[\sup_{\vartheta\in K} \psi(|v_t(\vartheta)|)^r] < \infty$. The stated result, except for the equality $h_t^*(\theta_0) = \sigma_t^2$ (a.s.), 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 $h_t^*(\theta_0) = \sigma_t^2$ a.s. (cf. Straumann and Mikosch, 2006, Props 3.7 and 3.12), 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 $|h_{n,n}(\theta_0) - \sigma_n^2| \leq \kappa^n |h_{0,0}(\theta_0) - \sigma_0^2|$, where $|h_{0,0}(\theta_0) - \sigma_0^2| = |g(u_{0,-1}, x(\theta_0); \theta_0) - \sigma_0^2| \leq \varrho x(\theta_0) + C(1 + u_{0,-1}^2) + \sigma_0^2$ 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 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

$$\left\|\sup_{\theta\in\Theta}|h_t^* - h_t|\right\|_r \leq C\kappa^t, \tag{B.1}$$

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

Proof of Theorem 1. For strong consistency of $\hat{\theta}_T$ 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 $\mathrm{E}[l_t^*(\theta)]$ is well defined, taking values in $\mathbb{R} \cup \{+\infty\}$ but $\mathrm{E}[l_t^*(\theta_0)] < \infty$. Next note that

$$\begin{aligned}
&\liminf_{T\to\infty} \inf_{\theta\in B(\theta_0,\delta)^c} \left(L_T(\theta) - L_T(\theta_0)\right) \\
&\quad \geq -\limsup_{T\to\infty} \sup_{\theta\in\Theta} \left|\left(L_T^*(\theta) - L_T^*(\theta_0)\right) - \left(L_T(\theta) - L_T(\theta_0)\right)\right| \\
&\quad\quad + \liminf_{T\to\infty}\left(\mathrm{E}[l_t^*(\theta_0)] - L_T^*(\theta_0)\right) \\
&\quad\quad + \liminf_{T\to\infty} \inf_{\theta\in B(\theta_0,\delta)^c} \left(L_T^*(\theta) - \mathrm{E}[l_t^*(\theta_0)]\right).
\end{aligned} \tag{B.2}$$

The desired result is obtained by showing that the first two terms on the minorant side of (B.2) equal zero a.s., whereas the third term is strictly positive. We begin by showing that

$$\sup_{\theta\in\Theta} \left|L_T^*(\theta) - L_T(\theta)\right| \to 0 \quad \text{a.s. as } T \to \infty, \tag{B.3}$$

from which it follows that the first term on the minorant side of (B.2) equals zero a.s. Note that

$$\begin{aligned}\left|l_t^*(\theta)-l_t(\theta)\right| &= \left|\log\left(h_t^*\right)-\log(h_t)+u_t^2(1/h_t^*-1/h_t)\right| \\ &\leq \underline{g}^{-1}\left|h_t^*-h_t\right|+\underline{g}^{-2}u_t^2\left|h_t^*-h_t\right|,\end{aligned}$$

where the inequality makes use of the mean value theorem and Assumption C4. Using Lemma A.1 and the Cauchy-Schwartz inequality we obtain

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

for some finite C_1. As seen in the proof of Proposition 1, the term in the parenthesis is finite, whereas inequality (B.1) gives the upper bound $C\kappa^t$ for the term $\left\|\sup_{\theta\in\Theta}|h_t^*-h_t|\right\|_r$. Hence, there exists a $\gamma>1$ such that $\gamma^t\sup_{\theta\in\Theta}\left|l_t^*(\theta)-l_t(\theta)\right|$ converges to zero in $L_{r/2}$–norm, and thus $\sum_{t=1}^{\infty}\sup_{\theta\in\Theta}\left|l_t^*(\theta)-l_t(\theta)\right|<\infty$ a.s. by Lemma A.2. Hence, the result in (B.3) follows.

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

$$u_t^2=\sigma_t^2\varepsilon_t^2-2(f_t(\mu)-f_t(\mu_0))\sigma_t\varepsilon_t+(f_t(\mu)-f_t(\mu_0))^2. \tag{B.4}$$

For $\theta=\theta_0$, $u_t^2(\theta_0)=\sigma_t^2\varepsilon_t^2$, and therefore $\mathrm{E}\left[l_t^*(\theta_0)\right]<\infty$ because $\mathrm{E}\left[\varepsilon_t^2\right]<\infty$ by Assumption E and $h_t^*(\theta_0)=\sigma_t^2$ a.s. by Proposition 1. However, for $\theta\neq\theta_0$, we may have $\mathrm{E}\left[u_t^2/h_t^*\right]=\infty$. (We note that if $\mathrm{E}\left[\sup_{\theta\in\Theta}l_t^*(\theta)\right]<\infty$, a uniform law of large numbers applies and the proof simplifies; cf. Straumann and Mikosch (2006), part 1 of the proof of Thm. 4.1.) That the second term on the minorant side of (B.2) equals zero a.s. can now be concluded from the ergodic theorem (because $l_t^*(\theta_0)$ is a stationary ergodic sequence with $\mathrm{E}\left[l_t^*(\theta_0)\right]<\infty$).

Finally, consider the third term on the minorant side of (B.2). As in Pfanzagl's (1969) proof of Lemma 3.11, it can be shown that $\mathrm{E}\left[l_t^*(\theta)\right]$ is a lower semicontinuous function on Θ and

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

(we omit the details, which can be obtained from the authors on request). Thus, the third term on the minorant side of (B.2) is positive if $\mathrm{E}\left[l_t^*(\theta)\right]-\mathrm{E}\left[l_t^*(\theta_0)\right]\geq 0$ with equality if and only if $\theta=\theta_0$. Because $\mathrm{E}\left[l_t^*(\theta_0)\right]<\infty$ this obviously holds if $\mathrm{E}\left[l_t^*(\theta)\right]=\infty$. Therefore in the following we assume that $\mathrm{E}\left[l_t^*(\theta)\right]<\infty$. As h_t^* and $(f_t(\mu)-f_t(\mu_0))$ are functions of $\left(y_{t-1},y_{t-2},\ldots\right)$ only and $h_t^*(\theta_0)=\sigma_t^2$ a.s., (B.4) together with Assumption E yield $\mathrm{E}\left[u_t^2/h_t^*\right]=\mathrm{E}\left[\sigma_t^2/h_t^*\right]+\mathrm{E}[(f_t(\mu)-f_t(\mu_0))^2/h_t^*]$ and thus

$$\mathrm{E}\left[l_t^*(\theta)\right]-\mathrm{E}\left[l_t^*(\theta_0)\right]=\mathrm{E}\left[\log(h_t^*/\sigma_t^2)\right]+\mathrm{E}\left[\sigma_t^2/h_t^*\right]+\mathrm{E}[(f_t(\mu)-f_t(\mu_0))^2/h_t^*]-1. \tag{B.5}$$

By the inequality $x - \log(x) \geq 1$ ($x \in \mathbb{R}_+$) and the identification conditions in Assumption C5, the expression in (B.5) is nonnegative and equals zero if and only if $\theta = \theta_0$. ■

APPENDIX C: Proofs for Section 4

We first present a simple lemma that is used in the proofs of Propositions 2 and 3. We omit the proof, which is available from the authors on request.

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 of Proposition 2. In addition to Lemma C.1 the proof makes use of Lemma A.3 and arguments similar to those in Straumann and Mikosch (2006, pp. 2483–2484). We omit details, which are available on request, but mention that the proof of part (b) boils down to showing

$$\left\| \sup_{\theta\in\Theta_0} |h^*_{\theta,t} - h_{\theta,t}| \right\|_{r/4} \leq C \max\{t, t^{4/r}\}\kappa^{t-1} \qquad (0 < C < \infty). \tag{C.1}$$

This result will also be used later. ■

Proof of Proposition 3. Similar to proof of Proposition 2. Details are available on request. ■

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 $l^*_{\theta\theta,t}$. (Pedantically, these derivatives are only defined on an event with probability one, but as this has no significant consequence on our results, we do not make this explicit.) 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

$$T^{1/2}L^*_{\theta,T}(\tilde{\theta}_T) = T^{1/2}L^*_{\theta,T}(\theta_0) + \dot{L}^*_{\theta\theta,T}T^{1/2}(\tilde{\theta}_T - \theta_0) \quad \text{a.s.,} \tag{D.1}$$

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* $T^{1/2}L^*_{\theta,T}(\theta_0) \xrightarrow{d} N\big(0, \mathcal{I}(\theta_0)\big)$, *where* $\mathcal{I}(\theta_0) = \mathrm{E}\big[l^*_{\theta,t}(\theta_0) l^{*\prime}_{\theta,t}(\theta_0)\big]$ *is finite and can be expressed as in (13).*

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

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

and hence

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

$$l^*_{\lambda,t}(\theta_0) = -\frac{h^*_{\lambda,t}(\theta_0)}{\sigma_t^2}\left(\varepsilon_t^2 - 1\right) \quad \text{a.s.} \tag{D.3}$$

By straightforward calculation one now obtains the expression (13). Finiteness of $\mathrm{E}\big[l^*_{\theta,t}(\theta_0) l^{*\prime}_{\theta,t}(\theta_0)\big]$ is an immediate consequence of $f_{\mu,t}$ being L_{2r}–dominated in Θ_0, Assumption N4, and Hölder's inequality. Noting that $l^*_{\theta,t}(\theta_0)$ is a stationary ergodic martingale difference sequence (with respect to $\mathcal{F}_t$) 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.

As noted in the discussion following Assumption N4, the moment condition $\mathrm{E}[\varepsilon_t^8] < \infty$ can be weakened to $\mathrm{E}[\varepsilon_t^4] < \infty$ if the i.i.d. assumption is made. Further details relating to this can be obtained from the authors on request. ■

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 $\mathcal{J}(\theta) = \mathrm{E}\big[l^*_{\theta\theta,t}(\theta)\big]$ *is continuous at* θ_0. *Moreover,* $\mathcal{J}(\theta_0)$ *can be expressed as in (14).*

Proof. The first partial derivatives of l^*_t were obtained in (D.2). The second ones are (a.s.)

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

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

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

are all finite. For the first norm, this has already been shown in the proof of Proposition 1. For the second and third norms, the justification is similar but now based on Assumption N3(i). Assumption C4 implies the finiteness of the fourth norm. The last two are finite by Assumption N4(ii). 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 (14) is seen by straightforward calculation. ■

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

Proof. The proof is similar to that in Francq and Zakoïan (2004), their derivation between equations (4.52) and (4.53). Details are available on request. ■

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. The proof makes use of standard arguments similar to those in Pötscher and Prucha (1991b). Details are available on request. ■

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

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

Proof. In this proof we assume $r = 2$, but retain the notation r for ease of comparison to previous results. By Assumption C4, $|h^*_{\theta,t}/h^*_t - h_{\theta,t}/h_t| \le \underline{g}^{-2}|h^*_{\theta,t}||h^*_t - h_t| + \underline{g}^{-1}|h^*_{\theta,t} - h_{\theta,t}|$, and thus by Lemma A.1, Hölder's inequality, and the norm inequality,

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

Thus, Proposition 2 and inequalities (B.1) and (C.1) give, for some finite C,

$$\left\|\sup_{\theta\in\Theta_0}\left|h^*_{\theta,t}/h^*_t - h_{\theta,t}/h_t\right|\right\|_{r/4} \le C\max\{t, t^{4/r}\}\kappa^t. \tag{D.4}$$

Now consider the difference $l^*_{\theta,t}(\theta) - l_{\theta,t}(\theta)$. Using Assumption C4, Lemma A.1, and standard inequalities, it can be shown that (details available on request)

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

The result now follows from inequalities (B.1) and (D.4) 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 a.s. identity $L^*_{\theta,T}(\tilde{\theta}_T) = L_{\theta,T}(\hat{\theta}_T) = 0$ and the mean value theorem, one then obtains

$$\begin{aligned}
T^{1/2}(L_{\theta,T}(\hat{\theta}_T) - L^*_{\theta,T}(\hat{\theta}_T)) &= T^{1/2}(L^*_{\theta,T}(\tilde{\theta}_T) - L^*_{\theta,T}(\hat{\theta}_T)) \\
&= \ddot{L}^*_{\theta\theta,T} T^{1/2}(\tilde{\theta}_T - \hat{\theta}_T) \quad \text{a.s.,}
\end{aligned} \tag{D.5}$$

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 (D.5),

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

and the majorant side converges to zero a.s. by Lemmas D.5 and A.2. Using the uniform convergence result for $L^*_{\theta\theta,T}(\theta)$ (see Lemma D.2) and the invertibility of $\mathcal{J}(\theta_0)$ (see Lemma D.3) it can be shown that, for all T sufficiently large, $\ddot{L}^*_{\theta\theta,T}$ is invertible (a.s.)

and $\ddot{L}^{*-1}_{\theta\theta,T} \to \mathcal{J}(\theta_0)^{-1}$ a.s. as $T \to \infty$ (cf. Lemma A.1 of Pötscher and Prucha, 1991b). Hence, the result follows. ■

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 Assumptions DGP and E. First consider the process σ_t^2. Under conditions (a.i) and (a.ii), the process σ_t^2 defined by $\sigma_t^2 = \omega_0\{1 + \sum_{j=1}^{\infty}\prod_{i=1}^{j}(\alpha_0\varepsilon_{t-i}^2 + \beta_0)\}$ is finite for all t a.s., strictly stationary, ergodic, and $\mathcal{F}_{t-1}^{\varepsilon}$–measurable. This can be seen as in Nelson (1990, proof of Thm. 2), replacing the strong law of large numbers for i.i.d. random variables used therein with one for ergodic stationary variables (cf. Lee and Hansen, 1994, proof of Lem. 2(1), and Linton et al., 2010, proof of Thm. 1). Moreover, a process σ_t^2 defined like this clearly is a solution to $\sigma_t^2 = g(\sigma_{t-1}\varepsilon_{t-1}, \sigma_{t-1}^2; \theta_0) = \omega_0 + \alpha_0\sigma_{t-1}^2\varepsilon_{t-1}^2 + \beta_0\sigma_{t-1}^2$. Making use of condition (a.iii), Lemma A.2, and the law of iterated expectations it can also be shown that $\mathrm{E}[\sigma_t^{2r}] < \infty$. Hence the process $(\sigma_t, \varepsilon_t)$ is stationary and ergodic, $\mathcal{F}_t^{\varepsilon}$–measurable, and $\mathrm{E}[\sigma_t^{2r}] < \infty$ and $\mathrm{E}[|\varepsilon_t|^{2r}] < \infty$ for some $r > 0$. Therefore, $u_{0,t} = \sigma_t\varepsilon_t$ is stationary and ergodic with $\mathrm{E}[|u_{0,t}|^{2r}] < \infty$. 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.iv) 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 a.s. Thus, using Proposition 2.6 of Straumann and Mikosch (2006), (y_t, σ_t^2) is stationary, ergodic, and $\mathcal{F}_t^{\varepsilon}$–measurable. Furthermore, from Lemma A.2 we can also conclude that $\mathrm{E}[|y_t|^{2r}] < \infty$. Thus, Assumption DGP holds. Finally, as (y_t, σ_t^2) was shown to be $\mathcal{F}_t^{\varepsilon}$–measurable, we have $\mathcal{F}_t \subseteq \mathcal{F}_t^{\varepsilon}$. This together with condition (a.v) ensures that Assumption E 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 C4 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 g, which is obvious. To see that Assumption C5 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), 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 $\mu \ne \mu_0$, 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 $\mu \ne \mu_0$, 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 the conditional distribution of ε_{t-i} is not degenerate (because $\mathrm{E}[\varepsilon_{t-i} \mid \mathcal{F}_{t-i-1}] = 0$ a.s. and $\mathrm{E}[\varepsilon_{t-i}^2 \mid \mathcal{F}_{t-i-1}] = 1$ a.s.). 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 conditions (b.i) and (b.ii).

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, condition (c.ii) ensures that in the preceding justification of Assumption DGP the arguments remain valid with $r = 2$. Hence it can be seen that Assumption DGP holds with $r = 2$. The latter part of Assumption N4(i) holds by condition (c.iii). Assumption N4(ii) can be verified as in Francq and Zakoïan's (2004, p. 635) derivation of their equations (4.59) and (4.60). Assumption N5(i) follows from condition (c.iv) because $\mathcal{F}_t \subseteq \mathcal{F}_t^{\varepsilon}$. For Assumption N5(ii), note that $x_\mu' \partial f_t(\mu_0)/\partial\mu = 0$ 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 C5. For Assumption 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 $(\varepsilon_{t-2},\varepsilon_{t-3},\ldots)$, which is ruled out by condition (c.iv). 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
Assumptions DGP, E, C1–C4, and C5(i) can be checked in a manner similar to that of the linear AR–GARCH case, the only modification being that the term $\beta_0 + \alpha_0\varepsilon_t^2$ is replaced with $\beta_0 + \alpha_0(|\varepsilon_{t-1}| - \gamma_0\varepsilon_{t-1})^2$ throughout. Assumption C5(ii) can be verified using arguments analogous to those in Straumann and Mikosch (2006, Lems. 5.2–5.4), replacing unconditional distributions with conditional ones in relevant places. Details are omitted.

Example 3: Nonlinear AR–GARCH
We first supplement 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.

(a) (vi) 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 neighborhoods $N(\varphi_0)$ 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.vi) is required to apply the results in Meitz and Saikkonen (2008c). Here, as well as in condition (b.v), we separately consider the cases of G being either strictly increasing or strictly decreasing.

In the subsequent derivations, we confine ourselves to the former case; details of the latter can be found in Meitz and Saikkonen (2008b). Condition (b.iii) is needed for the continuity requirement in Assumptions C2 and C3. It is also used to verify the identification conditions in Assumption C5, for which (b.iv) and (b.v) are needed also. 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, whereas (c.vi) ensures that Assumption N4(ii) holds.

Verification of Assumptions DGP and E. Assumption DGP follows from the conditions in (a) due to the results in Meitz and Saikkonen (2008c). Specifically, the conditions in (a) imply that Assumptions 1–4, 5(b), and 6 of Meitz and Saikkonen (2008c) hold so that from Theorem 1 of that paper we can conclude that Assumption DGP holds (details available on request). Assumption E follows from the i.i.d. assumption (a.i).

Verification of Assumptions for Consistency. Of the assumptions required for consistency, 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 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 of (b.iii) and the fact that F has range $[0, 1]$, and Assumption C4 holds because, due to compactness, ω is bounded away from zero for all $\theta \in \Theta$.

In order to verify Assumption C5(i), we first demonstrate that if A_i, $i = 0, \ldots, p$, are any nonempty open subsets of $\mathbb{R}$, the event

$$\left\{\left(y_t, \ldots, y_{t-p}\right) \in A_0 \times \cdots \times A_p\right\} \tag{E.1}$$

has a positive probability. By the aforementioned results of Meitz and Saikkonen (2008c), $\left(y_t, \ldots, y_{t-p}, \sigma_t^2\right)$ 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 (E.1) has positive probability if, from any fixed initial value, the (nonstationary) chain $\left(y_t^{\dagger}, \ldots, y_{t-p}^{\dagger}, \sigma_t^{\dagger 2}\right)$ 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 an everywhere positive density, the nonstationary chain can reach the set $A_p \times \mathbb{R}^p \times \mathbb{R}_+$ in one step with positive probability. Making use of the Chapman-Kolmogorov equations (Meyn and Tweedie, 1993, Thm. 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 (E.1) has a positive probability.

Consider now the identification condition in Assumption C5(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 $\mu \in \mathrm{M}$ 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

$$A_0(y_{t-d}; \mu, \mu_0) + \sum_{j=1}^{p} A_j(y_{t-d}; \mu, \mu_0) y_{t-j} = 0 \quad \text{a.s.} \tag{E.2}$$

To show 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 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 F has range [0, 1]). However, as $\phi_d \neq \phi_{0,d}$, it follows from (b.iv) that $A_d(\bar{y}_d; \mu, \mu_0)\bar{y}_d$ attains values arbitrarily large in absolute value on the set $S(d, y_\bullet)$ when $y_\bullet$ is small enough. Specifically, for $y_\bullet$ small enough, $|A_d(\bar{y}_d; \mu, \mu_0)\bar{y}_d| > M$. As the event $\{(y_{t-1}, \ldots, y_{t-p}) \in S(d, y_\bullet)\}$ has positive probability for any $y_\bullet$, we can contradict (E.2), 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 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 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 $\phi_0 = \phi_{0,0}$.

Using similar arguments it can be shown that $\psi_j = \psi_{0,j}$, $j = 0, \ldots, p$ (we omit the details). Therefore, the identity (E.2) takes the form

$$\left(F(y_{t-d}; \varphi) - F(y_{t-d}; \varphi_0)\right)\left[\psi_{0,0} + \sum_{j=1}^{p} \psi_{0,j} y_{t-j}\right] = 0 \quad \text{a.s.} \tag{E.3}$$

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 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 (E.3) 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 of Assumption C5(i).

In order to prove part (ii) of Assumption C5, 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

$$\{\sigma_t^2 \in (\sigma_\bullet, \sigma^\bullet)\} \tag{E.4}$$

is positive. As when considering the event in (E.1), it suffices to show that the nonstationary chain $(y_t^\dagger, \dots, y_{t-p}^\dagger, \sigma_t^{\dagger 2})$ eventually reaches the set $\mathbb{R}^{p+1} \times (\sigma_\bullet, \sigma^\bullet)$ with positive probability from any initial value. The components $y_t^\dagger, \dots, y_{t-p}^\dagger$ are not essential here, so we concentrate only on $\sigma_t^{\dagger 2}$. From a fixed initial value σ_0^2, the process $\sigma_t^{\dagger 2}$ reaches $\sigma_1^{\dagger 2} = \omega_0 + (\alpha_{0,1} + \alpha_{0,2} G(\sigma_0 \varepsilon_0; \gamma_0))\sigma_0^2 \varepsilon_0^2 + \beta_0 \sigma_0^2$ in one step. Because ε_0 has a density that is positive everywhere, $P\{\varepsilon_0^2 \le (\alpha_{0,1} + \alpha_{0,2})^{-1}(1 - \beta_0)/2\}$ is positive for all t. For all ε_0 taking such values, and defining $\bar{\beta}_0 = (1 + \beta_0)/2$ (< 1), we have $\sigma_1^{\dagger 2} \le \omega_0 + \bar{\beta}_0 \sigma_0^2$. Because $\varepsilon_1, \dots, \varepsilon_{k-1}$ also take such values with positive probability, the Chapman-Kolmogorov equations and an inductive argument yield that $\sigma_k^{\dagger 2} \le \omega_0(1 + \bar{\beta}_0 + \dots + \bar{\beta}_0^{k-1}) + \bar{\beta}_0^k \sigma_0^2$ with positive probability. Setting $\underline{\sigma} = \omega_0/(1 - \bar{\beta}_0) + \delta$ for some $\delta > 0$ it is clear that $\sigma_k^{\dagger 2} \le \underline{\sigma}$ 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 C5, 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 by Assumption C4, $\sigma_t^2 \ge \underline{g} > 0$. Hence we obtain

$$\begin{aligned}(\alpha_1 - \alpha_{0,1})\,\varepsilon_t^2 = -(\beta - \beta_0) - \sigma_t^{-2}\Big[(\omega - \omega_0) + \big(\alpha_2 G(\sigma_t \varepsilon_t; \gamma) \\ - \alpha_{0,2} G(\sigma_t \varepsilon_t; \gamma_0)\big)\varepsilon_t^2 \sigma_t^2\Big] \quad \text{a.s.}\end{aligned} \tag{E.5}$$

Because ε_t has an everywhere positive density, the event $\{\sigma_t^2 \ge \underline{g}, \varepsilon_t \le \underline{g}^{-1/2} M\}$ has positive probability for all $M < 0$, and on this event $\sigma_t \varepsilon_t \le M$. By condition (b.v), the term in square brackets in (E.5) can be made arbitrarily close to $(\omega - \omega_0)$ on the event $\{\sigma_t \varepsilon_t \le M\}$ by choosing a small enough M. Because σ_t^{-2} is bounded by $\underline{g}^{-1}$, the right-hand side of (E.5) is bounded on $\{\sigma_t \varepsilon_t \le M\}$, whereas the left-hand side may attain values arbitrarily large in absolute value if $\alpha_1 \ne \alpha_{0,1}$ and M is chosen small enough. Thus, because $\sigma_t \varepsilon_t \le M$ with positive probability for every $M < 0$, we must have $\alpha_1 = \alpha_{0,1}$. Under this restriction, (E.5) can be rearranged as

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

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

$$(\omega - \omega_0) + \alpha_{0,2}\,(G(\sigma_t \varepsilon_t; \gamma) - G(\sigma_t \varepsilon_t; \gamma_0))\,\varepsilon_t^2 \sigma_t^2 + (\beta - \beta_0)\sigma_t^2 = 0 \quad \text{a.s.,} \tag{E.6}$$

where $\alpha_{0,2} > 0$ by (b.ii). Now consider events $\{\sigma_t^2 \in (\sigma_\bullet, \sigma^\bullet), \varepsilon_t \le \underline{\sigma}^{-1/2} M\}$ with $\underline{\sigma} < \sigma_\bullet < \sigma^\bullet$ and $M < 0$, which, by (E.4) and the independence of σ_t^2 and ε_t, have positive probability. On these events $\sigma_t \varepsilon_t \leq M$ regardless of the values of $\sigma_\bullet$ and $\sigma^\bullet$. Thus, by condition (b.v) and choosing a small enough M, the sum of the first two terms in (E.6) can be made arbitrarily close to $(\omega - \omega_0)$ with positive probability. However, considering events with different values of $\sigma_\bullet$ and $\sigma^\bullet$, (E.6) is clearly violated unless $\beta = \beta_0$. Similar reasoning using (E.6) 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 (b.iii) and the last condition in (b.v), 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 must have $\gamma = \gamma_0$. Thus $\lambda = \lambda_0$, and Assumption C5(ii) holds.

Verification of Assumptions 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 $r = 2$ and $\mathrm{E}[\varepsilon_t^4] < \infty$ follows from conditions (a) and (c.ii). (As discussed after Assumption N4, due to the i.i.d. assumption, condition $\mathrm{E}[\varepsilon_t^8] < \infty$ can be weakened to $\mathrm{E}[\varepsilon_t^4] < \infty$.) Specifically, part (a) of Proposition 1 of Meitz and Saikkonen (2008c) 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). Verification of Assumption N4(ii) is long and tedious, and we omit the details (they are available on request).

As for Assumption N5, part (i) clearly holds due to conditions (a.i) and (a.ii). To verify Assumption N5(ii), calculate the partial derivatives of $f\left(y_{t-1}, \ldots, y_{t-p}; \mu\right)$ as

$$1,\ y_{t-1}, \ldots, y_{t-p},\ \left(1, y_{t-1}, \ldots, y_{t-p}\right) F(y_{t-d}; \varphi), \quad \text{and} \ \left(\psi_0 + \sum_{j=1}^{p} \psi_j y_{t-j}\right) F_\varphi(y_{t-d}; \varphi). \tag{E.7}$$

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 (E.7) and rearranging,

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

Using conditions (b.iv) and (c.iv) and arguments similar to those used to verify Assumption C5(i), it can be deduced that $x_1 = \cdots = x_{2p+2} = 0$ (we omit the details). 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 (b.i), at least one of the $\psi_{0,j}$, $j =$

0, ..., p, is nonzero, and the arguments used when verifying Assumption C5(i) can be used to arrive at contradiction (see equation (E.3) and the discussion following it). Hence, $x_{2p+3} = x_{2p+4} = 0$ and $x = 0$. 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. By straightforward derivation, the vector $g_{\lambda,t}^*$ has components 1, u_{t-1}^2, $G(u_{t-1};\gamma)u_{t-1}^2$, h_{t-1}^*, and $\alpha_2 G_\gamma(u_{t-1};\gamma)u_{t-1}^2$, so that

$$x_1 + x_2\sigma_t^2\varepsilon_t^2 + x_3 G(\sigma_t\varepsilon_t;\gamma_0)\sigma_t^2\varepsilon_t^2 + x_4\sigma_t^2 + \alpha_{0,2}(x_5, x_6)' G_\gamma(\sigma_t\varepsilon_t;\gamma_0)\sigma_t^2\varepsilon_t^2 = 0 \quad \text{a.s.} \tag{E.8}$$

Now, similarly to the verification of Assumption C5(ii), consider the events $\{\sigma_t^2 \in (\sigma_\bullet, \sigma^\bullet), \varepsilon_t \le \underline{\sigma}^{-1/2}M\}$ with $\underline{\sigma} < \sigma_\bullet < \sigma^\bullet$ and $M < 0$, which by (E.4) and the independence of σ_t and ε_t have positive probability, and, moreover, on these events $\sigma_t\varepsilon_t \le M$ 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 (E.8) are bounded (due to (b.v) and (c.v)) except the second one, which takes values arbitrarily large in absolute value unless $x_2 = 0$. 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 $\{\sigma_t^2 \in (\sigma_\bullet, \sigma^\bullet), \varepsilon_t \ge \underline{\sigma}^{-1/2}M\}$ with M positive, we can similarly conclude that $x_3 = 0$. With the restrictions derived so far,

$$x_1 + x_4\sigma_t^2 + \alpha_{0,2}(x_5, x_6)' G_\gamma(\sigma_t\varepsilon_t;\gamma_0)\sigma_t^2\varepsilon_t^2 = 0 \quad \text{a.s.} \tag{E.9}$$

Consider again the events $\{\sigma_t^2 \in (\sigma_\bullet, \sigma^\bullet), \varepsilon_t \ge \underline{\sigma}^{-1/2}M\}$. Letting $M > 0$ be arbitrarily large, but this time considering these events with different values for $\sigma_\bullet$ and $\sigma^\bullet$, (E.9) is clearly violated unless $x_4 = 0$. With 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 C5(ii). Thus, Assumption N5(iii) holds.