Available online at www.sciencedirect.com
ScienceDirect
EconomiA 16 (2015) 1–21
ECONOMIA
www.elsevier.com/locate/econ

Structure and asymptotic theory for nonlinear models with GARCH errors☆

Estrutura e Teoria Assintótica para Modelos Não-lineares com Erros GARCH economia e finanças

Felix Chan[a], Michael McAleer[b,c], Marcelo C. Medeiros[d,*]

[a] School of Economics and Finance, Curtin University of Technology, Australia
[b] Department of Quantitative Finance, National Tsing Hua University, Taiwan
[c] Econometric Institute, Erasmus University Rotterdam, The Netherlands
[d] Department of Economics, Pontifical Catholic University of Rio de Janeiro, Brazil

Received 3 June 2014; received in revised form 7 January 2015; accepted 22 January 2015
Available online 7 February 2015

Abstract

Nonlinear time series models, especially those with regime-switching and/or conditionally heteroskedastic errors, have become increasingly popular in the economics and finance literature. However, much of the research has concentrated on the empirical applications of various models, with little theoretical or statistical analysis associated with the structure of the processes or the associated asymptotic theory. In this paper, we derive sufficient conditions for strict stationarity and ergodicity of three different specifications of the first-order smooth transition autoregressions with heteroskedastic errors. This is essential, among other reasons, to establish the conditions under which the traditional LM linearity tests based on Taylor expansions are valid. We also provide sufficient conditions for consistency and asymptotic normality of the Quasi-Maximum Likelihood Estimator for a general nonlinear conditional mean model with first-order GARCH errors.

JEL classification: C22

Keywords: Nonlinear time series; Regime-switching; Smooth transition; STAR; GARCH; Asymptotic theory

☆ This paper was circulated previously as “Structure and Asymptotic Theory for STAR-GARCH(1,1) Models”. The authors wish to thank Thierry Jeantheau, Offer Lieberman, Shiqing Ling, Howell Tong and Alvaro Veiga for insightful discussions. The first author acknowledges the financial support of the Australian Research Council, the second author is most grateful for the financial support of the Australian Research Council, National Science Council, Taiwan, and the Japan Society for the Promotion of Science, and the third author wishes to thank CNPq/Brazil for partial financial support.
* Corresponding author. Tel.: +55 21 3527 1078; fax: +55 21 3527 1084.
E-mail address: mcm@econ.puc-rio.br (M.C. Medeiros).
Peer review under responsibility of National Association of Postgraduate Centers in Economics, ANPEC.

http://dx.doi.org/10.1016/j.econ.2015.01.001
1517-7580

Resumo

Modelos não-lineares com múltiplos regimes e heterpcedasticidade condicional são muito populares em economia e finanças. Neste artigo derivamos condições de estacionaridade para modelos com transição suave e erros heterocedásticos. Além disso derivamos condições suficientes para consistência e normalidade assintótica do estimador de quase-máxima verossimilhança.

Palavras-chave: modelos não-lineares; modelos com transição suave; GARCH; teoria assintótica

1. Introduction

Recent years have witnessed a vast development of nonlinear techniques for modelling the conditional mean and conditional variance of economic and financial time series. In the vast array of new technical developments for conditional mean models, the Smooth Transition AutoRegressive (STAR) specification, proposed by Chan and Tong (1986) and developed by Luukkonen et al. (1988) and Teräsvirta (1994), has found a number of successful applications (see Tweedie (1988) for a recent review).

The term "smooth transition" in its present meaning first appeared in Bacon and Watts (1971). They presented their smooth transition specification as a model of two intersecting lines with an abrupt change from one linear regression to another at an unknown change-point. Goldfeld and Quandt (1972, pp. 263–264) generalized the so-called two-regime switching regression model using the same idea. In the time series literature, the STAR model is a natural generalization of the Self-Exciting Threshold Autoregressive (SETAR) models pioneered by Tong (1978) and Tong and Lim (1980) (see also Tong (1990)).

In terms of the conditional variance, Engle's (1982) Autoregressive Conditional Heteroskedasticity (ARCH) model and Bollerslev's (1986) Generalized ARCH (GARCH) model are the most popular specifications for capturing symmetric time-varying volatility in financial and economic time series data. McAleer (2005) provide an overview of different univariate and multivariate conditional volatility models.

Despite their popularity, the structural and statistical properties of these models were not fully established until recently. Chan and Tong (1986) derived sufficient conditions for strict stationarity and geometric ergodicity of a two-regime STAR model, where the transition function is given by the cumulative Gaussian distribution. Although several papers have been published in the literature with general conditions for strict stationarity and ergodicity of nonlinear time series models, especially threshold-type models, few attempts have been made to comprehend the dynamics of more general smooth transition processes (see Chen and Tsay (1991) for an early reference on the ergodicity of threshold models). In general, only very restrictive sufficient conditions are provided. For general nonlinear homoskedastic autoregressions, see Bhattacharya and Lee (1995), An and Huang (1996), An and Chen (1997), and Lee (1998), among many others. Nonlinear models with ARCH errors (not GARCH) have been considered, for example, by Masry and Tjostheim (1995), Cline and Pu (1998, 1999, 2004), Lu (1998), Lu and Jang (2001), Chen and Chen (2001), Hwang and Woo (2001), Liebscher (2005), and Saikkonen (2007). Stability of nonlinear autoregressions with GARCH-type errors has been analyzed by Liu et al. (1997), Ling (1999), and Cline (2007). Of these articles, those of Liu et al. (1997) and Ling (1999) are restricted to threshold AR-GARCH models, whereas Cline (2007) analyzes a very general nonlinear autoregressive models with GARCH errors. Cline (2007) obtained sharp results for geometric ergodicity but a difficulty with the application of these results is that the assumptions employed are quite general and, hence are difficult to verify. A threshold AR-GARCH model is the only example that is explicitly treated in the paper. Furthermore, conditional heteroskedasticity is driven by the observed series instead of the autoregressive errors as in the usual GARCH specification. Ferrante et al. (2003) considered threshold bilinear Markov processes. Only recently, Meitz and Saikkonen (2008) study the stability of general nonlinear autoregressions or order p with first-order GARCH errors. However, they explicitly analyzed only a STAR model with two limiting regimes.

Consistency and asymptotic normality of the nonlinear least squares estimator are given under the assumption that the errors are homoskedastic and independent. In a recent paper, Mira and Escribano (2000) derived new sufficient conditions for consistency and asymptotic normality of the nonlinear least squares estimator. However, estimation of the conditional variance was not considered in these papers.

Significant efforts have been made to fully understand the properties of univariate and multivariate GARCH models. Nelson (1990) derived the necessary and sufficient log-moment condition for stationarity and ergodicity of the GARCH(1,1) model. This condition was extended to higher-order models by Bougerol and Picard (1992). Weak stationarity and the existence of fourth moments of a family of power GARCH models have been investigated in He and Teräsvirta (1999a,b), while Ling and McAleer (2002a,b) derived the necessary and sufficient conditions for the existence of all moments for these models.

Concerning the estimation of the parameters of GARCH models, Lee and Hansen (1994) and Lumsdaine (1996) proved that the local Quasi-Maximum Likelihood Estimator (QMLE) was consistent and asymptotically normal under strong conditions. Jeantheau (1998) established the consistency results of estimators for multivariate GARCH models. His proofs of consistency did not assume a particular functional form for the conditional mean, but assumed a log-moment condition and some regularity conditions for purposes of identification. More recently, Ling and McAleer (2003) proposed the vector ARMA-GARCH model and proved the consistency of the global QMLE under only the second-order moment condition. They also proved the asymptotic normality of the global (local) QMLE under the sixth-order (fourth-order) moment condition. Comte and Lieberman (2003) studied the asymptotic properties of the QMLE for the BEKK model of Engle and Kroner (1995). Berkes et al. (2003) proved the consistency and asymptotic normality if the QMLE of the parameters of the GARCH(p,q) model under second- and fourth-order moment conditions, respectively. Boussama (2000), McAleer et al. (2007), and Francq and Zakoïan (2004) also considered the properties of the QMLE under different specifications of the symmetric and asymmetric GARCH(p,q) model.

However, most of the theoretical results on GARCH models have assumed a constant or linear conditional mean (see McAleer (2005) for further details). It has not yet been established whether these results would also hold if the conditional mean were nonlinear. Chan and McAleer (2002) combined the general STAR model with GARCH(p,q) errors, but their results were derived under the assumption that the conditional mean parameters were known.

This paper extends existing results in the literature in several respects. The sufficient conditions for strict stationarity and geometric ergodicity of a general class of first-order STAR models with GARCH(1,1) errors are established. STAR models with more than two regimes are also considered. Second, consistency and asymptotic normality of the QMLE of a general nonlinear conditional mean model with first-order GARCH errors are derived under weak conditions. Finally, a simulation experiment highlight the small sample properties of the QMLE.

The plan of the paper is as follows. Section 2 provides a description of the models considered in the paper. Stationarity, ergodicity and the existence of moments are discussed in Section 3. The asymptotic properties of the QMLE are considered in Section 4. In Section 5 we present simulation results concerning the finite sample properties of the QMLE. Finally, Section 6 gives some concluding remarks. All technical proofs are given in the Appendix.

2. Model specification

In this section we consider three different classes of STAR-GARCH models. The first specification is an additive logistic STAR model with multiple regimes in the conditional mean and GARCH errors. This model nests the SETAR-GARCH process of Li and Lam (1995). A similar specification with Gaussian errors was proposed in Suarez-Fariñas et al. (2004) and Medeiros and Veiga (2000, 2005). The second specification is a restricted form of the multiple-regime logistic STAR model with GARCH errors. This particular functional form with homoskedastic errors was discussed in Tweedie (1988). Finally, the third specification is the Exponential STAR-GARCH (ESTAR-GARCH) model, of which the Exponential STAR (ESTAR) model of Teräsvirta (1994) is a special case.

Definition 1. The $\mathbb{R}$-valued process $\{y_t, t \in \mathbb{Z}\}$ follows an autoregressive model with time-varying coefficients and GARCH(1,1) errors if

$$y_t = f_0(s_t) + \sum_{i=1}^{p} f_i(s_t) y_{t-i} + \varepsilon_t, \tag{1}$$

$$\varepsilon_t = \eta_t \sqrt{h_t}, \tag{2}$$

$$h_t = \omega + \alpha\varepsilon_{t-1}^2 + \beta h_{t-1}, \tag{3}$$

where $\{\eta_t\}$ is a sequence of independently and identically distributed zero mean and unit variance random variables, $\eta_t \sim \mathrm{IID}(0, 1)$, and $f_j(s_t) \equiv f_j(s_t; \boldsymbol{\lambda}_j)$, $j = 0, 1, \ldots, p$, are nonlinear functions of the variables s_t and are indexed by the vector of parameters $\boldsymbol{\lambda}_j \in \mathbb{R}^K$.

It is clear that the model defined by Eqs. (1)–(3) is similar to the functional coefficient autoregressive model proposed by Chen and Tsay (1993). Depending on the choice of the functions $f_j(s_t; \boldsymbol{\lambda})$, $j = 0, 1, \ldots, p$, different specifications of the STAR model can be derived. The following cases are considered:

1. The Multiple Regime Logistic STAR(p)-GARCH(1,1) (or MLSTAR(p)-GARCH(1,1)) model:
   Set $s_t = y_{t-d}$, $d \in \mathbb{N}$, and

$$f_j(s_t; \boldsymbol{\lambda}) = \phi_{0j} + \sum_{i=1}^{m} \phi_{ij} G(y_{t-d}; \gamma_i, c_i), \quad j = 0, \ldots, p, \tag{4}$$

   where

$$G(y_{t-d}; \gamma_i, c_i) = \frac{1}{1 + e^{-\gamma_i (y_{t-d} - c_i)}}. \tag{5}$$

2. The Generalized STAR(p)-GARCH(1,1) (or GSTAR(p)-GARCH(1,1)) model:
   Set $s_t = y_{t-d}$, $d \in \mathbb{N}$, and

$$f_j(s_t; \boldsymbol{\lambda}) = \phi_{0j} + \phi_{1j} G(y_{t-d}; \gamma, \mathbf{c}), \tag{6}$$

   where

$$G(y_{t-d}; \gamma, c) = \frac{1}{1 + e^{-\gamma \left[\prod_{i=1}^{m} (y_{t-d} - c_i)\right]}}, \tag{7}$$

   with $\mathbf{c} = (c_1, \ldots, c_m)'$.
3. The Exponential STAR(p)-GARCH(1,1) (or ESTAR(p)-GARCH(1,1)) model:
   Set $s_t = y_{t-d}$, $d \in \mathbb{N}$, and

$$f_j(s_t; \boldsymbol{\lambda}) = \phi_{0j} + \phi_{1j} G(y_{t-d}; \gamma, c), \tag{8}$$

   where

$$G(y_{t-d}; \gamma, c) = 1 - e^{-\gamma (y_{t-d} - c)^2}. \tag{9}$$

Example 1. Consider a three regime MLSTAR(1)-GARCH(1,1) model, where the transition variable is y_{t-1}, $\phi_{00} = -0.001$, $\phi_{10} = 0.001$, $\phi_{20} = 0.001$, $\phi_{01} = -0.001$, $\phi_{11} = 0.001$, $\phi_{21} = 0.001$, $\gamma_1 = 1000$, $\gamma_2 = 1000$, $c_1 = -0.01$, $c_2 = 0.01$, $\omega = 10^{-5}$, $\alpha = 0.05$, and $\beta = 0.85$. Fig. 1 shows the scatter plot $f_0(y_{t-1})$ and $f_1(y_{t-1})$ versus y_{t-1}. One characteristic of such a specification is that the linear parameters in each limiting regimes are allowed to be different.

Example 2. Consider a three regime GSTAR(1)-GARCH(1,1) model, where the transition variable is y_{t-1}, $\phi_{00} = -0.001$, $\phi_{10} = 0.002$, $\phi_{01} = 0.025$, $\phi_{11} = 0.025$, $\gamma = 100, 000$, $c_1 = -0.01$, $c_2 = 0.01$, $\omega = 10^{-5}$, $\alpha = 0.05$, and $\beta = 0.85$. Fig. 2 shows the scatter plot $f_0(y_{t-1})$ and $f_1(y_{t-1})$ versus y_{t-1}. As distinct from the MLSTAR model, the linear parameters in each limiting extreme regime are restricted to be equal.

Example 3. Consider a three regime ESTAR(1)-GARCH(1,1) model, where the transition variable is y_{t-1}, $\phi_{00} = -0.001$, $\phi_{10} = 0.002$, $\phi_{01} = 0.025$, $\phi_{11} = 0.025$, $\gamma = 100, 000$, $c = 0$, $\omega = 10^{-5}$, $\alpha = 0.05$, and $\beta = 0.85$. Fig. 3 shows the scatter plot $f_0(y_{t-1})$ and $f_1(y_{t-1})$ versus y_{t-1}. As in the previous example, the linear parameters in each limiting extreme regime are restricted to be equal.

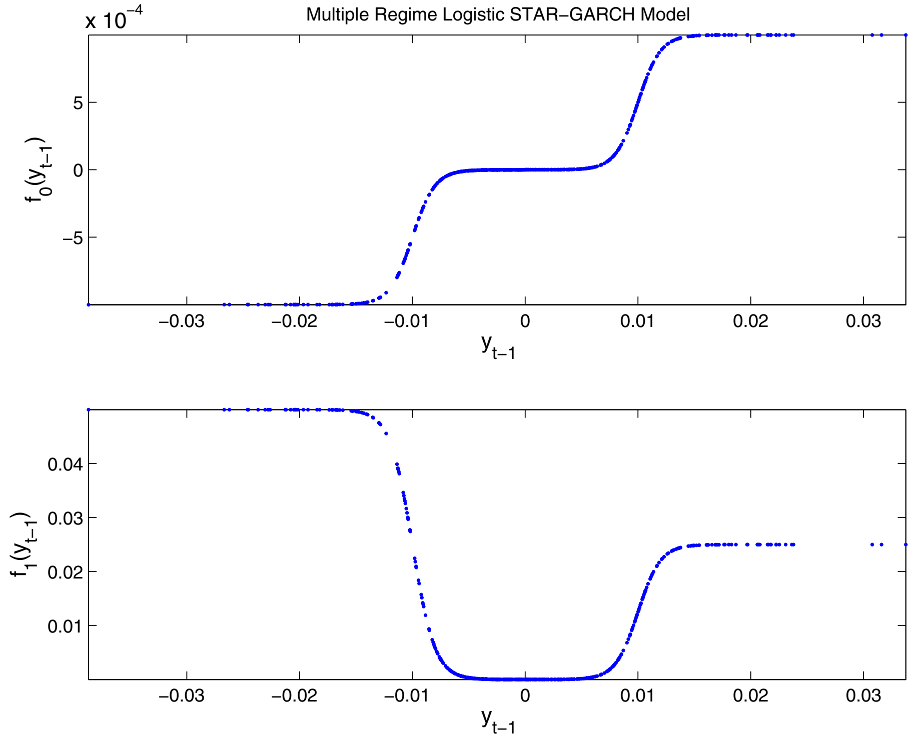

Fig. 1. Upper panel: $f_0(y_{t-1})$ versus y_{t-1} for one realization of the model described in Example 1. Lower panel: $f_1(y_{t-1})$ versus y_{t-1} for one realization of the model described in Example 1.

3. Probabilistic properties

In this section, only first-order models will be considered, while in Section 4 general nonlinear models will be analyzed. Consider the following set of assumptions.

Assumption 1 *(Error Term).* The sequence $\{\eta_t\}$ of IID(0, 1) random variables is drawn from a continuous (with respect to Lebesgue measure on the real line), unimodal, positive everywhere density, and bounded in a neighborhood of 0.

Assumption 2 *(Model Structure).* $p=1$ and $s_t = y_{t-1}$ in Eq. (1).

Assumption 3 *(Identifiability and Positiveness of the Variance).* The parameters of the model defined by (1)–(3) satisfy the following restrictions: (R.1a) $\gamma_i > 0$, $i = 1, \ldots, m$, and $c_1 < c_2 < \cdots < c_m$ in (4); (R1.b) $\gamma > 0$ and $c_1 \leq c_2 \leq \cdots \leq c_m$ in (6); (R.1c) $\gamma > 0$ in (8); and (R.2) $\omega > 0$, $\alpha > 0$, and $\beta > 0$.

Assumption 1 is standard. Note that we do not assume symmetry of the distribution, which is particularly useful when modelling financial time series. Assumption 2 forces the model to be of first-order. This will be crucial to the results in this section, but will be relaxed in Section 4. The restrictions (R.1a)–(R.1c) in Assumption 3 are important to guarantee that the model is globally identifiable. Restriction (R.2) is a sufficient condition for $h_t > 0$ with probability one.

Note that $\mathbf{z}_t = (y_t, h_t, \eta_t)'$ is a Markov chain with homogenous transition probability, expressed as

$$\mathbf{z}_t = \mathbf{F}(\mathbf{z}_{t-1}) + \mathbf{e}_t, \tag{10}$$

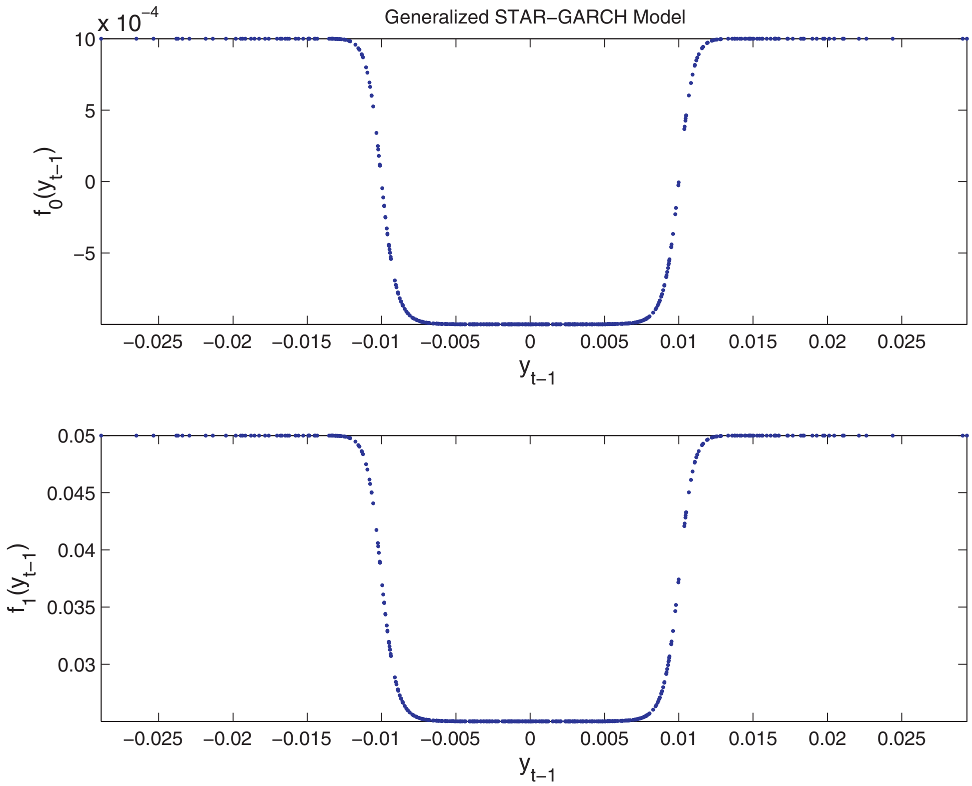

Fig. 2. Upper panel: $f_0(y_{t-1})$ versus y_{t-1} for one realization of the model described in Example 2. Lower panel: $f_1(y_{t-1})$ versus y_{t-1} for one realization of the model described in Example 2.

where

$$\mathbf{F}(\mathbf{z}_{t-1}) = \begin{bmatrix} f_0(y_{t-1}) + f_1(y_{t-1})y_{t-1} \\ \omega + \left(\beta + \alpha\eta_{t-1}^2\right) h_{t-1} \\ 0 \end{bmatrix}$$

and $\mathbf{e}_t = (\varepsilon_t, 0, \eta_t)'$.

The following theorems state the necessary conditions for strict stationarity and geometric ergodicity of the STAR-GARCH models considered in this paper.

Theorem 1 (Stationarity – MRLSTAR(1)-GARCH(1,1) model). *Define $\overline{\phi} = \sum_{i=0}^{m} \phi_{i1}$. Under Assumptions 1 and 2, and if (R.1a) in Assumption 3 holds, the process $\{y_t, t \in \mathbb{Z}\}$ defined by Eqs. (1)–(3) and (4) is strictly stationary and geometrically ergodic if $\alpha + \beta < 1$, $|\phi_{01}| < 1$ and $|\overline{\phi}| < 1$. Furthermore, the process $\{\mathbf{z}_t, t \in \mathbb{Z}\}$ admits a unique causal expansion.*

Theorem 2 (Stationarity – GSTAR(1)-GARCH(1,1) model). *Set $\overline{\phi} = \phi_{01} + \phi_{11}$. Under Assumption 1, and if (R.1b) in Assumption 3 holds, the process $\{y_t, t \in \mathbb{Z}\}$ defined by Eqs. (1)–(3) and (6) is strictly stationary and geometrically ergodic if $\alpha + \beta < 1$, $|\phi_{01}| < 1$ and $|\overline{\phi}| < 1$. Furthermore, the process $\{\mathbf{z}_t, t \in \mathbb{Z}\}$ admits a unique causal expansion.*

Theorem 3 (Stationarity – ESTAR(1)-GARCH(1,1) model). *Set $\overline{\phi} = \phi_{01} + \phi_{11}$. Under Assumption 1, and if (R.1c) in Assumption 3 holds, the process $\{y_t, t \in \mathbb{Z}\}$ defined by equations (1)–(3) and (8) is strictly stationary and geometrically ergodic if $\alpha + \beta < 1$ and $|\overline{\phi}| < 1$. Furthermore, the process $\{\mathbf{z}_t, t \in \mathbb{Z}\}$ admits a unique causal expansion.*

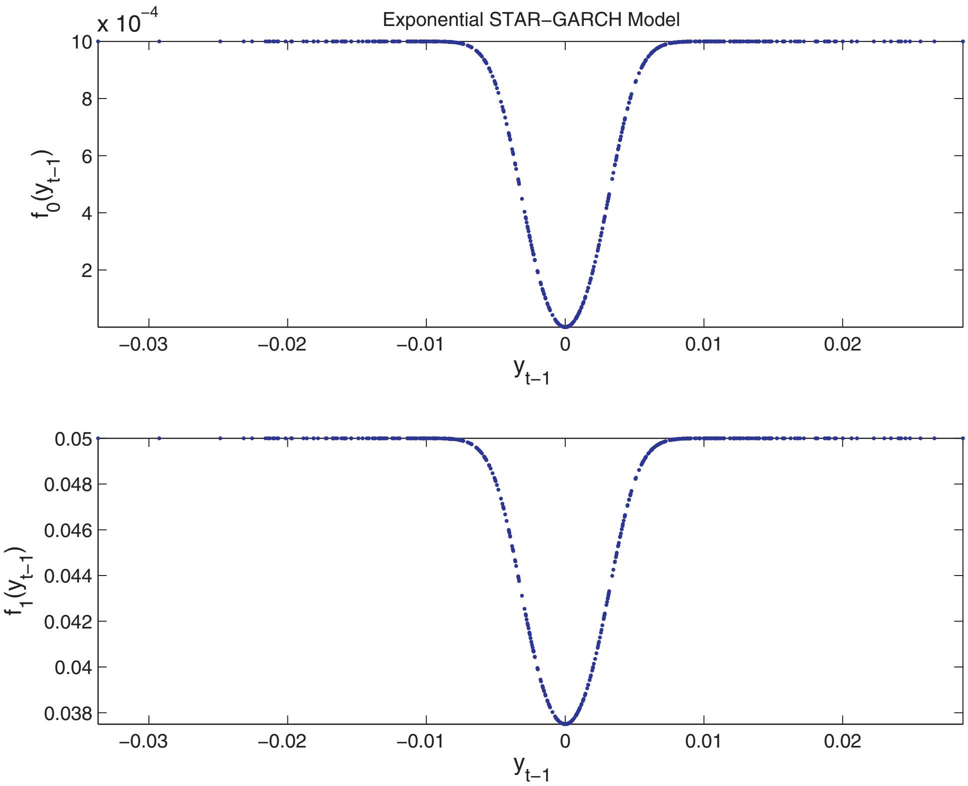

Fig. 3. Upper panel: $f_0(y_{t-1})$ versus y_{t-1} for one realization of the model described in Example 3. Lower panel: $f_1(y_{t-1})$ versus y_{t-1} for one realization of the model described in Example 3.

If the conditions of the above theorems are met, the processes $\{y_t\}$ and $\{h_t\}$ have the following causal expansions:

$$y_t = \lambda_{0,t-1} + \sum_{j=1}^{\infty}\prod_{k=0}^{j-1}\left[f_0(y_{t-1-j})f_1(y_{t-1-k}) + f_1(y_{t-1-k})\varepsilon_{t-j}\right], \tag{11}$$

$$h_t = \omega\left[1 + \sum_{j=1}^{\infty}\prod_{k=1}^{j}\left(\beta + \alpha\eta_{t-i}^2\right)\right]. \tag{12}$$

4. Parameter estimation and asymptotic theory

In this section, we discuss the estimation of general nonlinear autoregressive models with GARCH(1,1) errors. The STAR-GARCH models analyzed previously are just special cases of the general model.

Consider the following assumption.

Assumption 4. The $\mathbb{R}$-valued process $\{y_t, t \in \mathbb{Z}\}$ follows the following nonlinear autoregressive process with GARCH errors (NAR-GARCH):

$$y_t = g(\mathbf{y}_{t-1}; \boldsymbol{\lambda}) + \varepsilon_t, \tag{13}$$

$$\varepsilon_t = \eta_t\sqrt{h_t}, \tag{14}$$

$$h_t = \omega + \alpha\varepsilon_{t-1}^2 + \beta h_{t-1}, \tag{15}$$

where $\mathbf{y}_{t-1} = \left(y_{t-1}, \ldots, y_{t-p}\right)'$ and $\eta_t \sim \mathrm{IID}(0, 1)$.

Assumption 5. The nonlinear function $g(\mathbf{y}_{t-1}\,;\boldsymbol{\lambda})$ satisfies the following set of restrictions:

1. $g(\mathbf{y}_{t-1}\,;\boldsymbol{\lambda})$ is continuous in $\boldsymbol{\lambda}$ and measurable in $\mathbf{y}_{t-1}$.
2. $g(\mathbf{y}_{t-1}\,;\boldsymbol{\lambda})$ is parameterized such that the parameters are well defined.
3. $g(\mathbf{y}_{t-1}\,;\boldsymbol{\lambda})$ and ε_t are independent.
4. $\mathsf{E}|g(\mathbf{y}_{t-1}\,;\boldsymbol{\lambda})|^q<\infty$, $q=1, 2, 4$.
5. $\mathsf{E}\left\{\exp\left[g\left(\mathbf{y}_{t-1};\boldsymbol{\lambda}\right)\right]^q\right\}<\infty$, $q=1, 2, 4$.
6. $\mathsf{E}\left|\frac{\partial}{\partial\boldsymbol{\lambda}}g(\mathbf{y}_{t-1};\boldsymbol{\lambda})\right|^q<\infty$, $q=1, 2, 4$.
7. $\mathsf{E}\left|\frac{\partial^2}{\partial\boldsymbol{\lambda}\partial\boldsymbol{\lambda}'}g(\mathbf{y}_{t-1};\,\boldsymbol{\lambda})\right|^q<\infty$, $q=1, 2$.

Set $\boldsymbol{\psi}=\left(\boldsymbol{\lambda}', \boldsymbol{\pi}'\right)'$, where $\boldsymbol{\lambda}$ is the vector of parameters of the conditional mean, as defined in Section 2, and $\boldsymbol{\pi}=(\omega, \alpha, \beta)'$ is the vector of parameters of the conditional variance. As the distribution of η_t is unknown, the parameter vector $\boldsymbol{\psi}$ is estimated by the quasi-maximum likelihood (QML) method. Consider the following assumption.

Assumption 6. The true parameter vector, $\boldsymbol{\psi}_0\in\boldsymbol{\Psi}\subseteq\mathbb{R}^N$, is in the interior of $\boldsymbol{\Psi}$, a compact and convex parameter space, where $N=\mathsf{dim}(\boldsymbol{\lambda})+\mathsf{dim}(\boldsymbol{\pi})$ is the total number of parameters.

The quasi-log-likelihood function of the NAR-GARCH model is given by:

$$\mathcal{L}_T(\boldsymbol{\psi})=\frac{1}{T}\sum_{t=1}^{T}\ell_t(\boldsymbol{\psi})=\frac{1}{T}\sum_{t=1}^{T}-\frac{1}{2}\ln(2\pi)-\frac{1}{2}\ln(h_t)-\frac{\varepsilon_t^2}{2h_t}. \tag{16}$$

Note that the processes y_t and h_t, $t\leq 0$, are unobserved, and hence are only arbitrary constants. Thus, $\mathcal{L}_T(\boldsymbol{\psi})$ is a quasi-log-likelihood function that is not conditional on the true (y_0, h_0), making it suitable for practical applications. However, to prove the asymptotic properties of the QMLE, it is more convenient to work with the unobserved process $\left\{\left(\varepsilon_{u,t}, h_{u,t}\right):\ t=0, \pm1, \pm2, \ldots\right\}$.

The unobserved quasi-log-likelihood function conditional on $\mathcal{F}_0=(y_0, y_{-1}, y_{-2}, \ldots)$ is

$$\mathcal{L}_{u,T}(\boldsymbol{\psi})=\frac{1}{T}\sum_{t=1}^{T}\ell_{u,t}(\boldsymbol{\psi})=\frac{1}{T}\sum_{t=1}^{T}-\frac{1}{2}\ln(2\pi)-\frac{1}{2}\ln(h_{u,t})-\frac{\varepsilon_{u,t}^2}{2h_{u,t}}. \tag{17}$$

The main difference between $\mathcal{L}_T(\boldsymbol{\psi})$ and $\mathcal{L}_{u,T}(\boldsymbol{\psi})$ is that the former is conditional on any initial values, whereas the latter is conditional on an infinite series of past observations. In practical situations, the use of (17) is not possible.

Let

$$\widehat{\boldsymbol{\psi}}_T=\mathsf{argmax}\boldsymbol{\psi}\in\boldsymbol{\Psi}\mathcal{L}_T(\boldsymbol{\psi})=\mathsf{argmax}\boldsymbol{\psi}\in\boldsymbol{\Psi}\left(\frac{1}{T}\sum_{t=1}^{T}\ell_t(\boldsymbol{\psi})\right),$$

and

$$\widehat{\boldsymbol{\psi}}_{u,T}=\mathsf{argmax}\boldsymbol{\psi}\in\boldsymbol{\Psi}\mathcal{L}_{u,T}(\boldsymbol{\psi})=\mathsf{argmax}\boldsymbol{\psi}\in\boldsymbol{\Psi}\left(\frac{1}{T}\sum_{t=1}^{T}\ell_{u,t}(\boldsymbol{\psi})\right).$$

Define $\mathcal{L}(\boldsymbol{\psi})=\mathsf{E}\left[l_{u,t}(\boldsymbol{\psi})\right]$. In the following subsection, we discuss the existence of $\mathcal{L}(\boldsymbol{\psi})$ and the identifiability of the NAR-GARCH models. Then, in Section 4.2, we prove the consistency of $\widehat{\boldsymbol{\psi}}_T$ and $\widehat{\boldsymbol{\psi}}_{u,T}$. We first prove the strong consistency of $\widehat{\boldsymbol{\psi}}_{u,T}$, and then show that

$$\sup_{\boldsymbol{\psi}\in\boldsymbol{\Psi}}\left|\mathcal{L}_{u,T}(\boldsymbol{\psi})-\mathcal{L}_T(\boldsymbol{\psi})\right|\overset{a.s.}{\to}0,$$

so that the consistency of $\widehat{\boldsymbol{\psi}}_T$ follows. Asymptotic normality of both estimators is considered in Section 4.3. We prove the asymptotic normality of $\widehat{\boldsymbol{\psi}}_{u,T}$. The proof of $\widehat{\boldsymbol{\psi}}_T$ is straightforward.

4.1. Existence of the QMLE

The following theorem proves the existence of $\mathcal{L}(\boldsymbol{\psi})$. It is based on Theorem 2.12 in White (1994), which establishes that $\mathcal{L}(\boldsymbol{\psi})$ exists under certain conditions of continuity and measurability of the quasi-log-likelihood function.

Theorem 4. *Under Assumptions 1 and 2, $\mathcal{L}(\boldsymbol{\psi})$ exists, is finite, and is uniquely maximized at $\boldsymbol{\psi}_0$.*

4.2. Consistency

The following theorem states the sufficient conditions for strong consistency of the QMLE.

Theorem 5. *Under Assumptions 1–6, the QMLE of $\boldsymbol{\psi}$ is strongly consistent for $\boldsymbol{\psi}_0$, $\widehat{\boldsymbol{\psi}}\overset{a.s.}{\rightarrow}\boldsymbol{\psi}_0$.*

4.3. Asymptotic normality

First, we introduce the following matrices:

$$\mathbf{A}(\boldsymbol{\psi}_0)=\mathsf{E}\left[-\frac{\partial^2 l_{u,t}(\boldsymbol{\psi})}{\partial\boldsymbol{\psi}\partial\boldsymbol{\psi}'}\,|_{\boldsymbol{\psi}_0}\right],\qquad \mathbf{B}(\boldsymbol{\psi}_0)=\mathsf{E}\left[\frac{\partial\ell_{u,t}(\boldsymbol{\psi})}{\partial\boldsymbol{\psi}}\,|_{\boldsymbol{\psi}_0}\,\frac{\partial\ell_{u,t}(\boldsymbol{\psi})}{\partial\boldsymbol{\psi}'}\,|_{\boldsymbol{\psi}_0}\right],$$

and

$$\begin{aligned}\mathbf{A}_T(\boldsymbol{\psi})=\frac{1}{T}\sum_{t=1}^{T}\Bigg[&\frac{1}{2h_t}\left(\frac{\varepsilon_t^2}{h_t}-1\right)\frac{\partial^2 h_t}{\partial\boldsymbol{\psi}\partial\boldsymbol{\psi}'}-\frac{1}{2h_t^2}\left(2\frac{\varepsilon_t^2}{h_t}-1\right)\frac{\partial h_t}{\partial\boldsymbol{\psi}}\frac{\partial h_t}{\partial\boldsymbol{\psi}'}\\&+\left(\frac{\varepsilon_t}{h_t^2}\right)\left(\frac{\partial\varepsilon_t}{\partial\boldsymbol{\psi}}\frac{\partial h_t}{\partial\boldsymbol{\psi}'}+\frac{\partial h_t}{\partial\boldsymbol{\psi}}\frac{\partial\varepsilon_t}{\partial\boldsymbol{\psi}'}\right)+\frac{1}{h_t}\left(\frac{\partial\varepsilon_t}{\partial\boldsymbol{\psi}}\frac{\partial\varepsilon_t}{\partial\boldsymbol{\psi}'}+\varepsilon_t\frac{\partial^2\varepsilon_t}{\partial\boldsymbol{\psi}}\right)\Bigg],\end{aligned}\tag{18}$$

$$\begin{aligned}\mathbf{B}_T(\boldsymbol{\psi})&=\frac{1}{T}\sum_{t=1}^{T}\frac{\partial\ell_t(\boldsymbol{\psi})}{\partial\boldsymbol{\psi}}\frac{\partial\ell_t(\boldsymbol{\psi})}{\partial\boldsymbol{\psi}'}\\&=\frac{1}{T}\sum_{t=1}^{T}\left[\frac{1}{4h_t^2}\left(\frac{\varepsilon_t^2}{h_t}-1\right)^2\frac{\partial h_t}{\partial\boldsymbol{\psi}}\frac{\partial h_t}{\partial\boldsymbol{\psi}'}+\frac{\varepsilon_t^2}{h_t}\frac{\partial\varepsilon_t}{\partial\boldsymbol{\psi}}\frac{\partial\varepsilon_t}{\partial\boldsymbol{\psi}'}-\frac{\varepsilon_t}{2h_t^2}\left(\frac{\varepsilon_t^2}{h_t}-1\right)\left(\frac{\partial h_t}{\partial\boldsymbol{\psi}}\frac{\partial\varepsilon_t}{\partial\boldsymbol{\psi}'}+\frac{\partial\varepsilon_t}{\partial\boldsymbol{\psi}}\frac{\partial h_t}{\partial\boldsymbol{\psi}'}\right)\right].\end{aligned}\tag{19}$$

Consider the additional assumption:

Assumption 7. There exists no set Λ of cardinal 2 such that $\mathsf{Pr}[\eta_t\in\Lambda]=1$.

As in Francq and Zakoïan (2004), Assumption 7 is necessary for identifying reasons when the distribution of η_t is non-symmetric.

The following theorem states the asymptotic normality result.

Theorem 6. *Under Assumptions 1–7 and the additional assumption $\mathsf{E}\left[\varepsilon_t^4\right]=\mu_4<\infty$, then*

$$T^{1/2}(\widehat{\boldsymbol{\psi}}_T-\boldsymbol{\psi}_0)\overset{d}{\rightarrow}\mathsf{N}\left(\mathbf{0},\boldsymbol{\Omega}\right),\tag{20}$$

where $\boldsymbol{\Omega}=\mathbf{A}(\boldsymbol{\psi}_0)^{-1}\mathbf{B}(\boldsymbol{\psi}_0)\mathbf{A}(\boldsymbol{\psi}_0)^{-1}$. If the distribution of η_t is symmetric and $\mathsf{E}\left[\eta_t^4\right]=\kappa_4$, then

$$\mathbf{A}(\boldsymbol{\psi}_0)=\begin{pmatrix}\mathbf{A}_1&\mathbf{0}\\\mathbf{0}&\mathbf{A}_2\end{pmatrix},\qquad \mathbf{B}(\boldsymbol{\psi}_0)=\begin{pmatrix}\mathbf{B}_1&\mathbf{0}\\\mathbf{0}&\mathbf{B}_2\end{pmatrix},$$

Table 1
Simulation: Estimation results.

Parameter	True value	500 observations					
		Model 1		Model 2		Model 3	
		Mean	Std. Dev.	Mean	Std. Dev.	Mean	Std. Dev.
ϕ_{00}	−0.001	−0.0028	0.0190	−0.0026	0.0102	−0.0066	0.1005
ϕ_{10}	0.002	0.0071	0.0387	0.0038	0.0108	0.0077	0.1005
ϕ_{20}	0.001	0.0004	0.0421	–	–	–	–
ϕ_{01}	0.025	0.0350	0.198	0.0142	0.1323	0.0599	0.5974
ϕ_{11}	0.25	0.1342	0.3390	0.4171	0.1326	0.3872	0.5978
ϕ_{21}	0.001	0.0002	0.0531	–	–	–	–
γ_1	1000	1000	1.91e−7	1000	1.71e−8	1000	1.69e−8
γ_2	1000	1000	2.01e−7	–	–	–	–
c_1	−0.01	−0.0093	0.0115	−0.0101	0.0105	−0.0107	0.0127
c_2	0.01	0.0101	0.0099	–	–	–	–
ω	10e−5	16.78e−5	6.01e−4	17.43e−5	4.01e−4	17.83e−5	3.56e−4
α	0.05	0.0650	0.0600	0.0686	0.0578	0.0685	0.0577
β	0.85	0.6315	0.3347	0.6264	0.3399	0.6389	0.3400
Parameter	True value	1000 observations					
		Model 1		Model 2		Model 3	
		Mean	Std. Dev.	Mean	Std. Dev.	Mean	Std. Dev.
ϕ_{00}	−0.001	−0.0024	0.0091	−0.00212	0.0076	−0.0034	0.0144
ϕ_{10}	0.002	0.0015	0.0232	0.0032	0.0078	0.0045	0.0145
ϕ_{20}	0.001	0.0006	0.0236	–	–	–	–
ϕ_{01}	0.025	0.0204	0.0825	0.0465	0.8080	0.0371	0.6172
ϕ_{11}	0.25	0.171	0.273	0.3615	0.809	0.3261	0.6538
ϕ_{21}	0.001	0.0005	0.0181	–	–	–	–
γ_1	1000	1000	2.88e−9	1000	2.58e−8	1000	2.27e−8
γ_2	1000	1000	1.59e−8	–	–	–	–
c_1	−0.01	−0.0096	0.0004	−0.0106	7.67e−5	−0.0106	5.0e−5
c_2	0.01	0.0092	0.0001	0.0093	4.35e−5	–	–
ω	10e−5	15.09e−5	6.01e−5	16.46e−5	6.40e−5	16.51e−5	6.75e−5
α	0.05	0.0673	0.0368	0.0708	0.357	0.0707	0.0351
β	0.85	0.772	0.3223	0.7483	0.3143	0.7621	0.3063
Parameter	True value	5000 observations					
		Model 1		Model 2		Model 3	
		Mean	Std. Dev.	Mean	Std. Dev.	Mean	Std. Dev.
ϕ_{00}	−0.001	−0.0006	0.0022	−0.0015	0.00194	−0.0017	0.0026
ϕ_{10}	0.002	0.0028	0.0598	0.0024	0.00206	0.0023	0.0026
ϕ_{20}	0.001	0.0011	0.0089	–	–	–	–
ϕ_{01}	0.025	0.0338	0.0554	0.0211	0.0173	0.0235	0.2601
ϕ_{11}	0.25	0.2232	0.1399	0.2663	0.1732	0.2854	0.2602
ϕ_{21}	0.001	0.0009	0.0452	–	–	–	–
γ_1	1000	1000	1.30e−9	1000	5.60e−9	1000	1.907e−8
γ_2	1000	1000	8.23e−10	–	–	–	–
c_1	−0.01	−0.0098	4.06e−5	−0.0104	2.92e−5	−0.0102	1.25e−5
c_2	0.01	0.0126	0.0001	0.0098	7.80e−6	–	–
ω	1e−5	15.85e−5	5.66e−5	15.67e−5	5.90e−5	15.55e−5	5.84e−5
α	0.05	0.0598	0.0107	0.0596	0.0106	0.0594	0.0106
β	0.85	0.781	0.0616	0.7831	0.0642	0.7846	0.0638

The table shows the mean and the standard deviation of quasi-maximum likelihood estimator of the parameters of Models 1–3 over 1000 replications. We report the results with both 1000 and 5000 observations.

with

$$\mathbf{A}_1 = \mathsf{E}\left[\frac{1}{h_t^2}\frac{\partial h_t}{\partial \boldsymbol{\lambda}}\frac{\partial h_t}{\partial \boldsymbol{\lambda}'}\,|_{\boldsymbol{\psi}_0}\right] + \mathsf{E}\left[\frac{2}{h_t^2}\frac{\partial \varepsilon_t}{\partial \boldsymbol{\lambda}}\frac{\partial \varepsilon_t}{\partial \boldsymbol{\lambda}'}\,|_{\boldsymbol{\psi}_0}\right],$$

$$\mathbf{A}_2 = \mathsf{E}\left[\frac{1}{h_t^2}\frac{\partial h_t}{\partial \boldsymbol{\pi}}\frac{\partial h_t}{\partial \boldsymbol{\pi}'}\,|_{\boldsymbol{\psi}_0}\right],$$

$$\mathbf{B}_1 = (\kappa_4 - 1)\mathsf{E}\left[\frac{1}{h_t^2}\frac{\partial h_t}{\partial \boldsymbol{\lambda}}\frac{\partial h_t}{\partial \boldsymbol{\lambda}'}\,|_{\boldsymbol{\psi}_0}\right] + 4\mathsf{E}\left[\frac{1}{h_t^2}\frac{\partial \varepsilon_t}{\partial \boldsymbol{\lambda}}\frac{\partial \varepsilon_t}{\partial \boldsymbol{\lambda}'}\,|_{\boldsymbol{\psi}_0}\right],$$

and

$$\mathbf{B}_2 = (\kappa_4 - 1)\mathsf{E}\left[\frac{1}{h_t^2}\frac{\partial h_t}{\partial \boldsymbol{\pi}}\frac{\partial h_t}{\partial \boldsymbol{\pi}'}\,|_{\boldsymbol{\psi}_0}\right].$$

Furthermore, the matrices $\mathbf{A}(\boldsymbol{\psi}_0)$ *and* $\mathbf{B}(\boldsymbol{\psi}_0)$ *are consistently estimated by* $\mathbf{A}_T(\widehat{\boldsymbol{\psi}})$ *and* $\mathbf{B}_T(\widehat{\boldsymbol{\psi}})$*, respectively.*

Note that we allow the error term η_t to be non-Gaussian. However, we requite that $\epsilon_t = \sqrt{h_t}\eta_t$ to have finite fourth-order moment in order to achieve asymptotic normality of the estimates. In the case of very fat tailed errors, possibly without the existence of higher-order moments, the current results in this paper will not be valid anymore.

5. Monte Carlo simulations

In this section we report the results of a simulation study designed to evaluate the finite sample properties of the QMLE. We consider three different model specifications as described below:

- Model 1: MLSTAR(1)-GARCH(1,1) A three regime model where the transition variable is y_{t-1}, $\phi_{00} = -0.001$, $\phi_{10} = 0.002$, $\phi_{20} = 0.001$, $\phi_{01} = 0.025$, $\phi_{11} = 0.25$, $\phi_{21} = 0.001$, $\gamma_1 = 1000$, $\gamma_2 = 1000$, $c_1 = -0.01$, $c_2 = 0.01$, $\omega = 10^{-5}$, $\alpha = 0.05$, and $\beta = 0.85$.
- Model 2: GSTAR(1)-GARCH(1,1) A three regime model where the transition variable is y_{t-1}, $\phi_{00} = -0.001$, $\phi_{10} = 0.002$, $\phi_{01} = 0.025$, $\phi_{11} = 0.25$, $\gamma_1 = 1000$, $c_1 = -0.01$, $c_2 = 0.01$, $\omega = 10^{-5}$, $\alpha = 0.05$, and $\beta = 0.85$.
- Model 3: ESTAR(1)-GARCH(1,1) Consider a two regime model where the transition variable is y_{t-1}, $\phi_{00} = -0.001$, $\phi_{10} = 0.002$, $\phi_{01} = 0.025$, $\phi_{11} = 0.25$, $\gamma_1 = 1000$, $c_1 = -0.01$, $\omega = 10^{-5}$, $\alpha = 0.05$, and $\beta = 0.85$.

The results are illustrated in Table 1. The table shows the average and the standard deviation of the estimates over 1000 replications. As we can see from the table, the estimates are rather precise and improve as the sample size increases.

6. Concluding remarks

In this paper we have derived sufficient conditions for strict stationarity and geometric ergodicity of three different classes of first-order STAR-GARCH models. This is important in order to find the conditions under which the traditional LM linearity tests are valid. The asymptotic properties of the QMLE have also been considered. We have proved that the QMLE is consistent and asymptotically normal under weak conditions. These new results should be important for the estimation of STAR-GARCH models in financial econometrics.

Appendix A. Proofs of Theorems 1–3

The proofs of the theorems are based on Chan et al. (1985), and makes use of the results in Tweedie (1988).

Let $\mathbf{A}$ be a $k \times k$ matrix then $\rho(\mathbf{A})$ denotes the spectral radius of $\mathbf{A}$. That is, the maximum absolute eigenvalue of $\mathbf{A}$. Let $\mathfrak{A}$ be a bounded set of matrices and $\mathfrak{A}^k = \left\{\prod_{i=1}^{k} \mathbf{A}_i : \mathbf{A}_i \in \mathfrak{A}, i = 1, \ldots, k\right\}$, then $\rho_*(\mathfrak{A})$ denotes the joint spectral radius of the set $\mathfrak{A}$, that is

$$\rho_*(\mathfrak{A}) = \limsup_{k\to\infty} \left(\sup_{\mathbf{A}\in\mathfrak{A}^k} ||\mathbf{A}||\right)^{1/k}$$

For the purpose of the following proofs, consider a first-order STAR-GARCH models defined as:

$$y_t = f_0(y_{t-1}) + f_1(y_{t-1})y_{t-1} + \varepsilon_t, \tag{A.1}$$

$$\varepsilon_t = \eta_t \sqrt{h_t}, \tag{A.2}$$

and

$$h_t = \omega + \alpha\varepsilon_{t-1}^2 + \beta h_{t-1}, \tag{A.3}$$

where

$$\begin{aligned} f_0(y_{t-1}) &= \phi_{00} + \phi_{10} G(y_{t-1}; \gamma, c) \\ f_1(y_{t-1}) &= \phi_{01} + \phi_{11} G(y_{t-1}; \gamma, c) \end{aligned}$$

and $G(y_{t-1}; \gamma, c)$ is a twice differentiable function with the range equals to [0, 1]. Now, let $\mathbf{z}_t = (y_t, y_{t-1}, h_t)'$ then the STAR(1)-GARCH(1,1) model could have the following Markovian representation

$$\mathbf{z}_t = \mathbf{F}(\mathbf{z}_{t-1}, \eta_t) \tag{A.4}$$

where

$$\mathbf{F}(\mathbf{z}_{t-1}, \eta_t) = \begin{bmatrix} f_0(y_{t-1}) + f_1(y_{t-1})y_{t-1} \\ y_{t-1} \\ h(\mathbf{z}_{t-1}) \end{bmatrix} + \begin{bmatrix} h(\mathbf{z}_{t-1})^{1/2}\eta_t \\ 0 \\ 0 \end{bmatrix}. \tag{A.5}$$

The proof of ergodicity for STAR(1)-GARCH(1,1) is based on the results from Meitz and Saikkonen (2008), which provided sufficient conditions to verify ergodicity for the following process:

$$\begin{aligned} y_t &= f(y_{t-1}, \ldots, y_{t-p}) + h_t^{1/2}\eta_t \\ h_t &= g(u_{t-1}, h_{t-1}) \\ u_t &= y_t - f(y_{t-1}, \ldots, y_{t-p}) \end{aligned} \tag{A.6}$$

where f is a nonlinear function such that $f(y_{t-1}, \ldots, y_{t-p})$ defined a nonlinear autoregressive process of order p. h_t is a positive function of y_s such that $s < t$ and η_t is a sequence of iid(0, 1) random variables independent of $\{y_s : s < t\}$. Model (A.6) can be rewritten as a Markov chain such that

$$Z_t = F(Z_{t-1}, \eta_t)$$

where $Z_t = (y_t, y_{t-1}, \ldots, y_{t-p}, h_t)'$ and

$$F(Z_{t-1}, \eta_t) = \begin{pmatrix} f(y_{t-1}, \ldots, y_{t-p}) \\ y_{t-1} \\ \vdots \\ y_{t-p} \\ h_t(Z_{t-1}) \end{pmatrix} + \begin{pmatrix} h_t(Z_{t-1})^{1/2}\eta_t \\ 0 \\ \vdots \\ 0 \\ 0 \end{pmatrix}$$

Meitz and Saikkonen (2008) showed that the following conditions are sufficient to ensure geometric ergodicity for the Markov chain, Z_t.

Condition 1. η_t has a (Lebesgue) density which is positive and lower semicontinuous on $\mathbb{R}$. Furthermore, for some real $r \geq 1$, $\mathbb{E}(\eta_t^{2r}) < \infty$.

Condition 2. The function f is of the form

$$f(x) = a(x)'x + b(x), \quad x \in \mathbb{R}^p;$$

where the functions $a : \mathbb{R}^p \to \mathbb{R}^p$ and $b : \mathbb{R}^p \to \mathbb{R}$ are smooth and bounded.

Condition 3. Given $a(x)$ from the previous assumption, rewrite $a(x) = (a_1(x), a_2(x), \ldots, a_p(x))'$ and define the $(p+1) \times (p+1)$ matrix such that

$$A(x) = \begin{pmatrix} a_1(x) & a_2(x) & \ldots & a_p(x) & 0 \\ 1 & 0 & \ldots & 0 & 0 \\ \vdots & \vdots & \ddots & \vdots & \vdots \\ 0 & 0 & \ldots & 1 & 0 \end{pmatrix}.$$

Then there exists a matrix norm $||\bullet||$ induced by a vector norm such that $||A|| \leq \rho \ \forall A \in \mathfrak{A}$ where $\mathfrak{A} = \{A(x) : x \in \mathbb{R}^p\}$ and some $0 < \rho < 1$.

Condition 4.
a. The function $g : \mathbb{R} \times \mathbb{R}_+ \to \mathbb{R}_+$ is smooth and for some $\underline{g} > 0$, $\inf_{(u,x) \in \mathbb{R} \times \mathbb{R}_+} g(u, x) = \underline{g}$.
b. For all $x \in \mathbb{R}_+$, $g(u, x) \to \infty$ as $u \to \infty$.
c. $\exists h^* \in \mathbb{R}_+$ such that the sequence $h_k (k = 1, 2, \ldots)$ defined by $h_k = g(0, h_{k-1})$, $k = 1, 2, \ldots$ converges to h^* as $k \to \infty$ for all $h_0 \in \mathbb{R}_+$. If $g(u, x) \geq h^*$ for all $u \in \mathbb{R}$ and all $x \geq h^*$ it suffices that this convergence holds for all $h_0 \geq h^*$.
d. There exist nonnegative real numbers a and c, and a Borel measurable function $\psi : \mathbb{R} \to \mathbb{R}_+$ such that

$$g(x^{1/2}\eta_t, x) \leq (a + \psi(\eta_t))x + c$$

$\forall x \in \mathbb{R}_+$. Furthermore, $a + \psi(0) < 1$ and $E[(a + \psi(\eta_t))^r] < 1$ where the real number $r \geq 1$ is as in Assumption 1.
e. For each initial value $z_0 \in Z$, there exits a control sequence $e_1^{(0)}, \ldots, e_{p+2}^{(0)}$ such that the $(p+2) \times (p+2)$ matrix

$$\nabla F_{p+2}^{(0)} = \left[\frac{\partial}{\partial e_1} F_{p+2}(z_0, e_1^{(0)}, \ldots, e_{p+2}^{(0)}) : \ldots : \frac{\partial}{\partial e_{p+2}} F_{p+2}(z_0, e_1^{(0)}, \ldots, e_{p+2}^{(0)})\right]$$

is non-singular.

A.1. Proof of Theorem 1

It is sufficient to verify Conditions 1 to 5 in Meitz and Saikkonen (2008). Condition 1 is satisfied by Assumption (1) with $r = 1$. Define $f(y_{t-1}) = \lambda_{0,t-1} + \lambda_{1,t-1} y_{t-1}$ and let

$$\begin{aligned} a(x) &= \theta_0 + \theta_1 G(x;\gamma,c) \\ b(x) &= \phi_0 + \phi_1 G(x;\gamma,c) \\ g(u,x) &= \omega + \alpha u^2 + \beta x \end{aligned}$$

Hence, $f(x) = a(x)x + b(x)$ and hence Condition 2 is satisfied. Following Liebscher (2005), a sufficient condition to ensure Condition 3 is

$$\rho_*(\{\Phi_1, \Phi_2\}) < 1$$

where

$$\Phi_1 = \begin{pmatrix} \phi_0 & 0 \\ 1 & 0 \end{pmatrix}, \qquad \Phi_2 = \begin{pmatrix} \phi_0 + \phi_1 & 0 \\ 1 & 0 \end{pmatrix}$$

Let b_{ij} denotes the (i, j) element of the matrix B for $i, j = 1, 2$ such that $B = \prod_{i=1}^{k} A_i$, where $A_i \in \{\Phi_1, \Phi_2\}\ \forall\, i = 1, \ldots, k$. Given the structure of Φ_1 and Φ_2, it is easy to verify that $b_{12} = 0$ and $b_{22} = 0$ for all $k \in \mathbb{Z}_+$. This implies the eigenvalues of B are 0 and $\phi_0^l(\phi_0 + \phi_1)^m$ for some $l, m \in \mathbb{Z}_+$. Given the assumptions that $|\phi_0| < 1$, $|\phi_0 + \phi_1| < 1$ and $|\phi_0(\phi_0 + \phi_1)| < 1$, it is obvious that $\phi_0^l(\phi_0 + \phi_1)^m \to 0$ as $k \to \infty$. Hence, Condition 3 is satisfied.

Let $\underline{g} = \omega$, given that $\omega > 0$, $\alpha \geq 0$ and $\beta \geq 0$ then

$$\inf_{u,x \in \mathbb{R}\times\mathbb{R}_+} g(u,x) = \omega = \underline{g}.$$

In addition, $\forall x \in \mathbb{R}_+$, $g(u, x) \to \infty$ as $u \to \infty$. Since $\alpha + \beta < 1$, $\alpha > 0$ and $\beta > 0$ therefore $0 < \beta < 1$. Now, $h_k = g(0, h_{k-1}) = \omega + \beta h_{k-1}$ and for any nonnegative initial value $h_0 < \infty$, it is straightforward to show that

$$h_k = \frac{\omega(1-\beta^{k-1})}{1-\beta} + \beta^k h_0.$$

Hence, $h_k \to \frac{\omega}{1-\beta}$ as $k \to \infty$. Moreover, let $c = \omega$, $a = \beta$ and $\psi(\eta_t) = \alpha\eta_t^2$ then $g(x^{1/2}\eta_t, x) = (a + \psi(\eta_t))x + c$, with $a + \psi(0) = \beta < 1$. From Condition 1, $r = 1$ and therefore $\mathbb{E}(a + \psi(\eta_t))^r = \mathbb{E}(a + \psi(\eta_t)) = \alpha + \beta < 1$. Hence Condition 4 is satisfied.

To verify Condition 5, it is useful to note that $p = 1$ so that $\nabla F_{p+1}^{(0)} = \nabla F_3^{(0)}$ such that

$$\nabla F_3^{(0)} = \begin{pmatrix} \dfrac{\partial y_3}{\partial e_1} & \dfrac{\partial y_3}{\partial e_2} & h_3^{1/2} \\ \dfrac{\partial y_2}{\partial e_1} & h_2^{1/2} & 0 \\ \dfrac{\partial h_3}{\partial e_1} & \dfrac{\partial h_3}{\partial e_2} & 0 \end{pmatrix}$$

Let the control sequence be $\left(e_1^{(0)}, e_2^{(0)}, e_3^{(0)}\right) = (e_1, 0, 0)$ where $|e_1| < \infty$. Note that $h_i^{1/2} > 0$ for $i = 2, 3$. Evaluating $\nabla F_3^{(0)}$ at the specified control sequence gives

$$\begin{aligned} \frac{\partial h_3}{\partial e_1} &= \beta \frac{\partial h_2}{\partial e_1} > 0 \\ \frac{\partial h_3}{\partial e_2} &= 2\alpha e_2 h_2 = 0 \end{aligned}$$

and hence, there exists a control sequence such that $\nabla F_3^{(0)}$ is non-singular and therefore Conditions 1 to 5 are satisfied. This completes the proof. ■

A.2. Proof of Theorem 3

The proof of Theorem 3 follows the same lines as the one of Theorem 1. ■

A.3. Proof of Theorem 2

The proof of Theorem 2 follows the same lines as the one of Theorem 1. ■

Appendix B. Proofs of Theorems 4–6

B.1. Proof of Theorem 4

It is easy to see that $\mathbf{F}(\mathbf{z}_t)$, as in (10), is a continuous function in the parameter vector $\boldsymbol{\psi}$. Similarly, we can see that $\mathbf{F}(\mathbf{z}_t)$ is continuous in $\mathbf{z}_t$, and therefore is measurable, for each fixed value of $\boldsymbol{\psi}$.

Furthermore, under the restrictions in Assumption 2, and if the stationarity conditions of either Theorem 1, 2, or 3 are satisfied, then $\mathsf{E}\left[\sup_{\boldsymbol{\psi}\in\boldsymbol{\Psi}}\left|h_{u,t}\right|\right]<\infty$ and $\mathsf{E}\left[\sup_{\boldsymbol{\psi}\in\boldsymbol{\Psi}}\left|y_{u,t}\right|\right]<\infty$. By Jensen's inequality, $\mathsf{E}\left[\sup_{\boldsymbol{\psi}\in\boldsymbol{\Psi}}\left|\ln\left|h_{u,t}\right|\right|\right]<\infty$. Thus, $\mathsf{E}\left[\left|l_{u,t}(\boldsymbol{\psi})\right|\right]<\infty\ \forall\,\boldsymbol{\psi}\in\boldsymbol{\Psi}$.

Let $h_{0,t}$ be the true conditional variance and $\varepsilon_{0,t}=h_{0,t}^{1/2}\eta_t$. In order to show that $\mathcal{L}(\boldsymbol{\psi})$ is uniquely maximized at $\boldsymbol{\psi}_0$, rewrite the maximization problem as

$$\max_{\boldsymbol{\psi}\in\boldsymbol{\Psi}}\left[\mathcal{L}(\boldsymbol{\psi})-\mathcal{L}(\boldsymbol{\psi}_0)\right]=\max_{\boldsymbol{\psi}\in\boldsymbol{\Psi}}\left\{\mathsf{E}\left[\ln\left(\frac{h_{0,t}}{h_{u,t}}\right)-\frac{\varepsilon_t^2}{h_{u,t}}+1\right]\right\}. \tag{B.1}$$

Writing $\varepsilon_t=\varepsilon_t-\varepsilon_{0,t}+\varepsilon_{0,t}$, Eq. (B.1) becomes

$$\begin{aligned}\max_{\boldsymbol{\psi}\in\boldsymbol{\Psi}}\left[\mathcal{L}(\boldsymbol{\psi})-\mathcal{L}(\boldsymbol{\psi}_0)\right]&=\max_{\boldsymbol{\psi}\in\boldsymbol{\Psi}}\left\{\mathsf{E}\left[\ln\left(\frac{h_{0,t}}{h_{u,t}}\right)-\frac{h_{0,t}}{h_{u,t}}+1\right]-\mathsf{E}\left[\frac{[\varepsilon_t-\varepsilon_{0,t}]^2}{h_{u,t}}\right]-\mathsf{E}\left[\frac{2\eta_t h_{0,t}^{1/2}(\varepsilon_t-\varepsilon_{0,t})}{h_{u,t}}\right]\right\}\\&=\max_{\boldsymbol{\psi}\in\boldsymbol{\Psi}}\left\{\mathsf{E}\left[\ln\left(\frac{h_{0,t}}{h_{u,t}}\right)-\frac{h_{0,t}}{h_{u,t}}+1\right]-\mathsf{E}\left[\frac{[\varepsilon_t-\varepsilon_{0,t}]^2}{h_{u,t}}\right]\right\},\end{aligned} \tag{B.2}$$

where

$$\mathsf{E}\left[\frac{2\eta_t h_{0,t}^{1/2}(\varepsilon_t-\varepsilon_{0,t})}{h_{u,t}}\right]=0$$

by the Law of Iterated Expectations.

Note that, for any $x>0$, $m(x)=\ln(x)-x\leq 0$, so that

$$\mathsf{E}\left[\ln\left(\frac{h_{0,t}}{h_{u,t}}\right)-\frac{h_{0,t}}{h_{u,t}}\right]\leq 0.$$

Furthermore, $m(x)$ is maximized at $x=1$. If $x \neq 1$, $m(x) < m(1)$, implying that $\mathsf{E}[m(x)] \leq \mathsf{E}[m(1)]$, with equality only if $x=1$ a.s.. However, this will occur only if $\frac{h_{0,t}}{h_{u,t}} = 1$, a.s.. In addition,

$$\mathsf{E}\left[\frac{[\varepsilon_t - \varepsilon_{0,t}]^2}{h_{u,t}}\right] = 0$$

if and only if $\varepsilon_t = \varepsilon_{0,t}$. Hence, $\boldsymbol{\psi} = \boldsymbol{\psi}_0$. This completes the proof. ■

B.2. Proof of Theorem 5

Following White (1994), Theorem 3.5, $\widehat{\boldsymbol{\psi}}_{u,T} \xrightarrow{a.s.} \boldsymbol{\psi}_0$ if the following conditions hold:

(1) The parameter space $\boldsymbol{\Psi}$ is compact.
(2) $\mathcal{L}_{u,T}(\boldsymbol{\psi})$ is continuous in $\boldsymbol{\psi} \in \boldsymbol{\Psi}$. Furthermore, $\mathcal{L}_{u,T}(\boldsymbol{\psi})$ is a measurable function of y_t, $t=1, \ldots, T$, for all $\boldsymbol{\psi} \in \boldsymbol{\Psi}$.
(3) $\mathcal{L}(\boldsymbol{\psi})$ has a unique maximum at $\boldsymbol{\psi}_0$.
(4) $\lim_{T\to\infty} \sup_{\boldsymbol{\psi}\in\boldsymbol{\Psi}} \left|\mathcal{L}_{u,T}(\boldsymbol{\psi}) - \mathcal{L}(\boldsymbol{\psi})\right| = 0,\ a.s..$

Condition (1) holds by assumption. Theorem 4 shows that Conditions (2) and (3) are satisfied. By Lemma 1, Condition (4) is also satisfied. Thus, $\widehat{\boldsymbol{\psi}}_{u,T} \xrightarrow{a.s.} \boldsymbol{\psi}_0$.

Lemma 2 shows that

$$\lim_{T\to\infty} \sup_{\boldsymbol{\psi}\in\boldsymbol{\Psi}} \left|\mathcal{L}_{u,T}(\boldsymbol{\psi}) - \mathcal{L}_T(\boldsymbol{\psi})\right| = 0\, a.s.,$$

implying that $\widehat{\boldsymbol{\psi}}_T \xrightarrow{a.s.} \boldsymbol{\psi}_0$. This completes the proof. ■

B.3. Proof of Theorem 6

We start by proving asymptotic normality of the QMLE using the unobserved log-likelihood. When this is shown, the proof using the observed log-likelihood is immediate by Lemmas 2 and 4. According to Theorem 6.4 in White (1994), to prove the asymptotic normality of the QMLE we need the following conditions in addition to those stated in the proof of Theorem 5:

(5) The true parameter vector $\boldsymbol{\psi}_0$ is interior to $\boldsymbol{\Psi}$.
(6) The matrix

$$\mathbf{A}_T(\boldsymbol{\psi}) = \frac{1}{T}\sum_{t=1}^{T}\left(\frac{\partial^2 l_t(\boldsymbol{\psi})}{\partial\boldsymbol{\psi}\partial\boldsymbol{\psi}'}\right)$$

exists a.s. and is continuous in $\boldsymbol{\Psi}$.
(7) The matrix $\mathbf{A}_T(\boldsymbol{\psi}) \xrightarrow{a.s.} \mathbf{A}(\boldsymbol{\psi}_0)$, for any sequence $\boldsymbol{\psi}_T$, such that $\boldsymbol{\psi}_T \xrightarrow{a.s.} \boldsymbol{\psi}_0$.
(8) The score vector satisfies

$$\frac{1}{\sqrt{T}}\sum_{t=1}^{T}\left(\frac{\partial l_t(\boldsymbol{\psi})}{\partial\boldsymbol{\psi}}\right) \xrightarrow{d} \mathsf{N}(\mathbf{0}, \mathbf{B}(\boldsymbol{\psi}_0)).$$

Condition (5) is satisfied by assumption. Condition (6) follows from the fact that $l_t(\boldsymbol{\psi})$ is differentiable of order two on $\boldsymbol{\psi} \in \boldsymbol{\Psi}$, and the stationarity of the STAR-GARCH model. The non-singularity of $\mathbf{A}(\boldsymbol{\psi}_0)$ and $\mathbf{B}(\boldsymbol{\psi}_0)$ follows from

Lemma 4. Furthermore, Lemmas 3 and 5 implies that Condition (7) is satisfied. In Lemma 6, we prove that condition (8) is also satisfied. This completes the proof. ■

Appendix C. Lemmas

Lemma 1. *Suppose that y_t follows a STAR-GARCH model satisfying the restrictions in Assumptions 1 and 2, and the stationarity and ergodicity conditions are met. Then,*

$$\lim_{T\to\infty} \sup_{\boldsymbol{\psi}\in\boldsymbol{\Psi}} \left|\mathcal{L}_{u,T}(\boldsymbol{\psi}) - \mathcal{L}(\boldsymbol{\psi})\right| = 0, \ a.s..$$

Proof. Set $g(\mathbf{Y}_t, \boldsymbol{\psi}) = l_{u,t}(\boldsymbol{\psi}) - \mathsf{E}\left[l_{u,t}(\boldsymbol{\psi})\right]$, where $\mathbf{Y}_t = \left[y_t, y_{t-1}, y_{t-2}, \ldots\right]'$. Hence, $\mathsf{E}\left[g(\mathbf{Y}_t, \boldsymbol{\psi})\right] = 0$. It is clear that $\mathsf{E}\left[\sup_{\boldsymbol{\psi}\in\boldsymbol{\Psi}} |g(\mathbf{Y}_t, \boldsymbol{\psi})|\right] < \infty$ by Theorem 4. Furthermore, as $g(\mathbf{Y}_t, \boldsymbol{\psi})$ is strictly stationary and ergodic, then $\lim_{T\to\infty} \sup_{\boldsymbol{\psi}\in\boldsymbol{\Psi}} \left|T^{-1}\sum_{t=1}^{T} g(\mathbf{Y}_t, \boldsymbol{\psi})\right| = 0, \ a.s.$. This completes the proof. ■

Lemma 2. *Under the assumptions of* Lemma 1,

$$\lim_{T\to\infty} \sup_{\boldsymbol{\psi}\in\boldsymbol{\Psi}} \left|\mathcal{L}_{u,T}(\boldsymbol{\psi}) - \mathcal{L}_{T}(\boldsymbol{\psi})\right| = 0, a.s..$$

Proof. First, write

$$h_t = \sum_{i=0}^{t-1} \beta^i \left(\omega + \alpha\varepsilon_{t-1-i}^2\right) + \beta^t h_0$$

and

$$h_{u,t} = \beta^{t-1}\left(\omega + \alpha\varepsilon_{u,0}^2\right) + \sum_{i=0}^{t-2} \beta^i \left(\omega + \alpha\varepsilon_{t-1-i}^2\right) + \beta^t h_{u,0},$$

such that

$$\begin{aligned} \left|h_t - h_{u,t}\right| &= \left|\beta^{t-1}\alpha\left(\varepsilon_0^2 - \varepsilon_{u,0}^2\right) + \beta^t\left(h_0 - h_{u,0}\right)\right| \\ &\le \beta^{t-1}\alpha\left|\varepsilon_0^2 - \varepsilon_{u,0}^2\right| + \beta^t\left|h_0 - h_{u,0}\right|. \end{aligned}$$

Under the stationarity of the process, and if (R.2) in Assumption 2 and the log-moment condition hold, it is clear that $0<\beta<1$. Furthermore, $h_{u,0}$ and $\varepsilon_{0,u}^2$ are well defined, as

$$\mathsf{Pr}\left[\sup_{\boldsymbol{\psi}\in\boldsymbol{\Psi}} \left(h_{u,0} > K_1\right)\right] \to 0 \text{ as } K_1 \to \infty, \text{ and } \mathsf{Pr}\left[\sup_{\boldsymbol{\psi}\in\boldsymbol{\Psi}} \left(\varepsilon_{u,0}^2 > K_2\right)\right] \to 0 \text{ as } K_2 \to \infty.$$

Thus,

$$\sup_{\boldsymbol{\psi}\in\boldsymbol{\Psi}} \left|h_t - h_{u,t}\right| \le K_h\rho_1^t, \quad a.s.,$$

and

$$\sup_{\boldsymbol{\psi}\in\boldsymbol{\Psi}} \left|\varepsilon_0^2 - \varepsilon_{u,0}^2\right| \le K_\varepsilon\rho_2^t, \quad a.s.,$$

where K_h and K_ε are positive and finite constants, $0<\rho_1<1$, and $0<\rho_2<1$. Hence, as $h_t>\omega$ and $\log(x)\leq x-1$,

$$\begin{aligned}\sup_{\psi\in\Psi}\left|l_t-l_{u,t}\right| &\leq \sup_{\psi\in\Psi}\left[\varepsilon_t^2\left|\frac{h_{u,t}-h_t}{h_t h_{u,t}}\right|+\left|\log\left(1+\frac{h_t-h_{u,t}}{h_{u,t}}\right)\right|\right]\\ &\leq \sup_{\psi\in\Psi}\left(\frac{1}{\omega^2}\right)K_h\rho_1^t\varepsilon_t^2+\sup_{\psi\in\Psi}\left(\frac{1}{\omega}\right)K_h\rho_1^t, \quad , a.s..\end{aligned}$$

Following the same arguments as in the proof of Theorems 2.1 and 3.1 in Francq and Zakoïan (2004), it can be shown that

$$\lim_{T\to\infty}\sup_{\psi\in\Psi}\left|\mathcal{L}_{u,T}(\psi)-\mathcal{L}_T(\psi)\right|=0, \quad a.s..$$

This completes the proof. ■

Lemma 3. *Under the conditions of* Theorem 6,

$$\mathsf{E}\left[\left|\frac{\partial l_t(\psi)}{\partial\psi}\,|_{\psi_0}\right|\right]<\infty, \tag{C.1}$$

$$\mathsf{E}\left[\left|\frac{\partial l_t(\psi)}{\partial\psi}\,|_{\psi_0}\,\frac{\partial l_t(\psi)}{\partial\psi'}\,|_{\psi_0}\right|\right]<\infty, \tag{C.2}$$

and

$$\mathsf{E}\left[\left|\frac{\partial^2 l_t(\psi)}{\partial\psi\partial\psi'}\,|_{\psi_0}\right|\right]<\infty. \tag{C.3}$$

Proof. Set

$$\nabla_0 l_{u,t}\equiv\frac{\partial l_{u,t}(\psi)}{\partial\psi}\,|_{\psi_0}, \qquad \nabla_0 h_{u,t}\equiv\frac{\partial h_{u,t}}{\partial\psi}\,|_{\psi_0}, \qquad \nabla_0\varepsilon_t\equiv\frac{\partial\varepsilon_t}{\partial\psi}\,|_{\psi_0}$$

$$\nabla_0^2 l_{u,t}\equiv\frac{\partial^2 l_{u,t}(\psi)}{\partial\psi\partial\psi'}\,|_{\psi_0}, \qquad \nabla_0^2 h_{u,t}\equiv\frac{\partial^2 h_{u,t}}{\partial\psi\partial\psi'}\,|_{\psi_0}, \quad \text{and} \quad \nabla_0^2\varepsilon_t\equiv\frac{\partial^2\varepsilon_t}{\partial\psi\partial\psi'}\,|_{\psi_0}.$$

Then,

$$\nabla_0 l_{u,t}=\frac{1}{2h_{u,t}}\left(\frac{\varepsilon_t^2}{h_{u,t}}-1\right)\nabla_0 h_{u,t}-\frac{\varepsilon_t}{h_{u,t}}\nabla_0\varepsilon_t$$

and

$$\begin{aligned}\nabla_0^2 l_{u,t} &= \left(\frac{\varepsilon_t^2}{h_{u,t}}-1\right)\frac{1}{2h_{u,t}}\nabla_0^2 h_{u,t}-\frac{1}{2h_{u,t}^2}\left(2\frac{\varepsilon_t^2}{h_{u,t}}-1\right)\nabla_0 h_{u,t}\nabla_0 h_{u,t}'\\ &+\left(\frac{\varepsilon_t}{h_{u,t}^2}\right)\left(\nabla_0\varepsilon_t\nabla_0 h_{u,t}'+\nabla_0 h_{u,t}\nabla_0\varepsilon_t'\right)+\frac{1}{h_{u,t}}\left(\nabla_0\varepsilon_t\nabla_0\varepsilon_t'+\varepsilon_t\nabla_0^2\varepsilon_t\right).\end{aligned}$$

Set $\psi=\left(\lambda',\pi'\right)'$, where, as stated before, λ is the vector of parameters of the conditional mean and π is the vector of parameters of the conditional variance. As in the proof of Theorem 3.2 in Francq and Zakoïan (2004), the derivatives with respect to π are clearly bounded. We proceed by analyzing the derivatives with respect to λ. As $\varepsilon_t=y_t-f_0(y_{t-1};\lambda)-f_1(y_{t-1};\lambda)y_{t-1}$, we have

$$\frac{\partial\varepsilon_t}{\partial\lambda}=-\frac{\partial f_0(y_{t-1};\lambda)}{\partial\lambda}-\frac{\partial f_1(y_{t-1};\lambda)}{\partial\lambda}y_{t-1}, \tag{C.4}$$

$$\frac{\partial^2\varepsilon_t}{\partial\lambda\partial\lambda'}=-\frac{\partial^2 f_0(y_{t-1};\lambda)}{\partial\lambda\partial\lambda'}-\frac{\partial^2 f_1(y_{t-1};\lambda)}{\partial\lambda\partial\lambda'}y_{t-1}, \tag{C.5}$$

$$\frac{\partial h_{u,t}}{\partial \boldsymbol{\lambda}} = 2\alpha \sum_{i=0}^{\infty} \left(\beta^i \varepsilon_{t-1-i} \frac{\partial \varepsilon_{t-1-i}}{\partial \boldsymbol{\lambda}} \right), \tag{C.6}$$

and

$$\frac{\partial^2 h_{u,t}}{\partial \boldsymbol{\lambda} \partial \boldsymbol{\lambda}'} = 2\alpha \sum_{i=0}^{\infty} \beta^i \left(\varepsilon_{t-1-i} \frac{\partial^2 \varepsilon_{t-1-i}}{\partial \boldsymbol{\lambda} \partial \boldsymbol{\lambda}'} + \frac{\partial \varepsilon_{t-1-i}}{\partial \boldsymbol{\lambda}} \frac{\partial \varepsilon_{t-1-i}}{\partial \boldsymbol{\lambda}'} \right). \tag{C.7}$$

As the derivatives of the transition function are bounded, if the strict stationarity and ergodicity conditions hold, (C.4)–(C.7) are clearly bounded. Hence, the remainder of the proof follows from the proof of Theorem 3.2 (part (*i*)) in Francq and Zakoïan (2004). This completes the proof. ■

Lemma 4. *Under the conditions of* Theorem 6, $\mathbf{A}(\boldsymbol{\psi}_0)$ *and* $\mathbf{B}(\boldsymbol{\psi}_0)$ *are nonsingular and, when* η_t *has a symmetric distribution, are block-diagonal.*

Proof. First, note that (R1a)–(R1c) in Assumption 2 and Assumption 7 guarantee the minimality (identifiability) of the different specifications of the STAR models considered in this paper. Therefore, the results follow from the proof of Theorem 3.2 (part (*ii*)) in Francq and Zakoïan (2004). This completes the proof. ■

Lemma 5. *Under the conditions of* Theorem 6,

$$\text{(a)} \lim_{T\to\infty} \sup_{\boldsymbol{\psi}\in\boldsymbol{\Psi}} \left\| \frac{1}{T} \sum_{t=1}^{T} \left[\frac{\partial l_{u,t}(\boldsymbol{\psi})}{\partial \boldsymbol{\psi}} - \frac{\partial l_t(\boldsymbol{\psi})}{\partial \boldsymbol{\psi}} \right] \right\| = \mathbf{0}, \quad a.s.,$$

$$\text{(b)} \lim_{T\to\infty} \sup_{\boldsymbol{\psi}\in\boldsymbol{\Psi}} \left\| \frac{1}{T} \sum_{t=1}^{T} \left[\frac{\partial^2 l_{u,t}(\boldsymbol{\psi})}{\partial \boldsymbol{\psi} \partial \boldsymbol{\psi}'} - \frac{\partial^2 l_t(\boldsymbol{\psi})}{\partial \boldsymbol{\psi} \partial \boldsymbol{\psi}'} \right] \right\| = \mathbf{0}, \quad a.s, \quad \text{and}$$

$$\text{(c)} \lim_{T\to\infty} \sup_{\boldsymbol{\psi}\in\boldsymbol{\Psi}} \left\| \frac{1}{T} \sum_{t=1}^{T} \frac{\partial^2 l_{u,t}(\boldsymbol{\psi})}{\partial \boldsymbol{\psi} \partial \boldsymbol{\psi}'} - \mathsf{E} \left[\frac{\partial^2 l_{u,t}(\boldsymbol{\psi})}{\partial \boldsymbol{\psi} \partial \boldsymbol{\psi}'} \right] \right\| = \mathbf{0}, \quad a.s..$$

Proof. First, assume that h_0 and $h_{u,0}$ are fixed constants. It is easy to show that

$$\begin{aligned} \left| \frac{\partial h_t}{\partial \boldsymbol{\lambda}} - \frac{\partial h_{u,t}}{\partial \boldsymbol{\lambda}} \right| &= 2\alpha\beta^{t-1} \left| \varepsilon_0 \frac{\partial \varepsilon_0}{\partial \boldsymbol{\lambda}} - \varepsilon_{u,0} \frac{\partial \varepsilon_{u,0}}{\partial \boldsymbol{\lambda}} \right| \\ &\leq 2\alpha\beta^{t-1} \left(\left| \varepsilon_0 \frac{\partial \varepsilon_0}{\partial \boldsymbol{\lambda}} \right| + \left| \varepsilon_{u,0} \frac{\partial \varepsilon_{u,0}}{\partial \boldsymbol{\lambda}} \right| \right) < \infty, \end{aligned}$$

as $0 < \beta < 1$ and y_t is stationary and ergodic. Hence, following the same arguments as in the proof of Theorem 3.2 (part (*iii*)) in Francq and Zakoïan (2004), it is straightforward to show that

$$\lim_{T\to\infty} \sup_{\boldsymbol{\psi}\in\boldsymbol{\Psi}} \left\| \frac{1}{T} \sum_{t=1}^{T} \left[\frac{\partial l_{u,t}(\boldsymbol{\psi})}{\partial \boldsymbol{\lambda}} - \frac{\partial l_t(\boldsymbol{\psi})}{\partial \boldsymbol{\lambda}} \right] \right\| = \mathbf{0}.$$

Furthermore, as

$$\begin{aligned} \frac{\partial h_t}{\partial \omega} - \frac{\partial h_{u,t}}{\partial \omega} &= 0 \\ \frac{\partial h_t}{\partial \alpha} - \frac{\partial h_{u,t}}{\partial \alpha} &= \varepsilon_0^2 - \varepsilon_{u,0}^2 \\ \frac{\partial h_t}{\partial \beta} - \frac{\partial h_{u,t}}{\partial \beta} &= (t-1)\beta^{t-2} \left(\varepsilon_0^2 - \varepsilon_{u,0}^2 \right) + t\beta^{t-1}(h_0 - h_{u,0}), \end{aligned}$$

It is clear that

$$\lim_{T\to\infty}\sup_{\boldsymbol{\psi}\in\boldsymbol{\Psi}}\left\|\frac{1}{T}\sum_{t=1}^{T}\left[\frac{\partial l_{u,t}(\boldsymbol{\psi})}{\partial\boldsymbol{\pi}}-\frac{\partial l_t(\boldsymbol{\psi})}{\partial\boldsymbol{\pi}}\right]\right\|=\mathbf{0}.$$

The proof of part (a) is now complete. The proof of part (b) follows along similar lines. The proof of part (c) follows the same arguments as in the proof of Theorem 3.2 (part (v)) in Francq and Zakoïan (2004). This completes the proof. ■

Lemma 6. *Under the conditions of* Theorem 6,

$$\frac{1}{\sqrt{T}}\sum_{t=1}^{T}\frac{\partial l_t(\boldsymbol{\psi})}{\partial\boldsymbol{\psi}}\Big|_{\boldsymbol{\psi}_0}\xrightarrow{d}\mathsf{N}(\mathbf{0},\mathbf{B}(\boldsymbol{\psi}_0)).$$

Proof. Let $S_T=\sum_{t=1}^{T}\mathbf{c}'\nabla_0 l_{u,t}$, where $\mathbf{c}$ is a constant vector. Then S_T is a martingale with respect to $\mathcal{F}_t$, the filtration generated by all past observations of y_t. By the given assumptions, $\mathsf{E}[S_T]>0$. Using the central limit theorem of Stout (1974),

$$T^{-1/2}S_T\xrightarrow{d}\mathsf{N}(0,\mathbf{c}'\mathbf{B}(\boldsymbol{\psi}_0)\mathbf{c}).$$

By the Cramér-Wold device,

$$T^{-1/2}\sum_{t=1}^{T}\frac{\partial l_{u,t}(\boldsymbol{\psi})}{\partial\boldsymbol{\psi}}\Big|_{\boldsymbol{\psi}_0}\xrightarrow{d}\mathsf{N}\left(0,\mathbf{B}(\boldsymbol{\psi}_0)\right).$$

By Lemma 5,

$$T^{-1/2}\sum_{t=1}^{T}\left\|\frac{\partial l_{u,t}(\boldsymbol{\psi})}{\partial\boldsymbol{\psi}}\Big|_{\boldsymbol{\psi}_0}-\frac{\partial l_t(\boldsymbol{\psi})}{\partial\boldsymbol{\psi}}\Big|_{\boldsymbol{\psi}_0}\right\|\xrightarrow{a.s.}\mathbf{0}.$$

Thus,

$$T^{-1/2}\sum_{t=1}^{T}\frac{\partial l_t(\boldsymbol{\psi})}{\partial\boldsymbol{\psi}}\Big|_{\boldsymbol{\psi}_0}\xrightarrow{d}\mathsf{N}(0,\boldsymbol{B}_0).$$

This completes the proof. ■

References

An, H., Chen, S., 1997. A note on the ergodicity of non-linear autoregressive model. Stat. Probabil. Lett. 34, 365–372.

An, H., Huang, F., 1996. The geometrical ergodicity of nonlinear autoregressive models. Stat. Sinica 6, 943–956.

Bacon, D.W., Watts, D.G., 1971. Estimating the transition between two intersecting lines. Biometrika 58, 525–534.

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

Bhattacharya, R., Lee, C., 1995. On the geometric ergodicity of nonlinear autoregressive models. Stat. Probabil. Lett. 22, 311–315.

Bollerslev, T., 1986. Generalized autoregressive conditional heteroskedasticity. J. Economet. 21, 307–328.

Bougerol, P., Picard, N.M., 1992. Stationarity of GARCH processes and of some non-negative time series. J. Economet. 52, 115–127.

Boussama, F., 2000. Normalité Asymptotique de L'Estimateur du Pseudo-Maximum de Vraisemblance dún Modèle GARCH. C. R. Acad. Sci. Ser. I 331, 81–84.

Chan, F., McAleer, M., 2002. Maximum likelihood estimation of STAR and STAR-GARCH models: a Monte Carlo Analysis. J. Appl. Econom. 17, 509–534.

Chan, K.S., Petruccelli, J.D., Tong, H., Woolford, S.W., 1985. A multiple-threshold AR(1) model. J. Appl. Probabil. 22 (2), 267–279.

Chan, K.S., Tong, H., 1986. On estimating thresholds in autoregressive models. J. Time Series Anal. 7, 179–190.

Chen, G., Chen, M., 2001. On a class of nonlinear AR(p) models with nonlinear ARCH errors. Aust. N.Z. J. Stat. 43, 445–454.

Chen, R., Tsay, R.S., 1991. On the ergodicity of TAR(1) processes. Ann. Appl. Probab. 1, 613–634.

Chen, R., Tsay, R.S., 1993. Functional coefficient autoregressive models. J. Am. Stat. Assoc. 88, 298–308.

Cline, D., 2007. Stability of nonlinear stochastic recursions with application to nonlinear AR-GARCH models. Adv. Appl. Probab. 39, 462–491.

Cline, D., Pu, H., 1998. Verifying irreducibility and continuity of a nonlinear time series. Stat. Probabil. Lett. 40, 139–148.

Cline, D., Pu, H., 1999. Geometric ergodicity of nonlinear time series. Stat. Sinica 9, 1103–1118.

Cline, D., Pu, H., 2004. Stability and the Lyapounov exponent of threshold AR-ARCH models. Ann. Appl. Probab. 14, 1920–1949.

Comte, F., Lieberman, O., 2003. Asymptotic theory for multivariate GARCH processes. J. Multivar. Anal. 84, 61–84.

Engle, R., Kroner, K., 1995. Multivariate simultaneous generalized ARCH. Economet. Theory 11, 122–150.

Engle, R.F., 1982. Autoregressive conditional heteroskedasticity with estimates of the variance of UK inflations. Econometrica 50, 987–1007.

Ferrante, M., Fonseca, G., Vidoni, P., 2003. Geometric ergodicity, regularity of the invariant distribution and inference for a threshold bilinear Markov process. Stat. Sinica 13, 367–384.

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

Goldfeld, S.M., Quandt, R., 1972. Nonlinear Methods in Econometrics. North Holland, Amsterdam.

He, C., Teräsvirta, T., 1999a. Fourth moment structure of the GARCH(p, q) process. Economet. Theory 15, 824–846.

He, C., Teräsvirta, T., 1999b. Properties of moments of a family of GARCH processes. J. Economet. 92, 173–192.

Hwang, S., Woo, M.-J., 2001. Threshold ARCH(1) processes: asymptotic inference. Stat. Probabil. Lett. 53, 11–20.

Jeantheau, T., 1998. Strong consistency of estimators for multivariate ARCH models. Economet. Theory 14, 70–86.

Lee, C., 1998. Asymptotics of a class of pth-order nonlinear autoregressive processes. Stat. Probabil. Lett. 40, 171–177.

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

Li, W.K., Lam, K., 1995. Modelling asymmetry in stock returns by a threshold autoregressive conditional heteroscedastic model. The Statistician 44, 333–341.

Liebscher, E., 2005. Towards a unified approach for proving geometric ergodicity and mixing properties of nonlinear autoregressive processes. J. Time Series Anal. 26, 669–689.

Ling, S., 1999. On the probabilistic properties of a double threshold ARMA conditional heteroscedasticity model. J. Appl. Probabil. 36, 1–18.

Ling, S., McAleer, M., 2002a. Necessary and sufficient moment conditions for the GARCH(r, s) and asymmetric power GARCH(r, s) models. Economet. Theory 18, 722–729.

Ling, S., McAleer, M., 2002b. Stationarity and the existence of moments of a family of GARCH processes. J. Economet. 106, 109–117.

Ling, S., McAleer, M., 2003. Asymptotic theory for a vector ARMA-GARCH model. Economet. Theory 19, 280–310.

Liu, J., Li, W.K., Li, C.W., 1997. On a threshold autoregression with conditional heteroscedastic variances. J. Stat. Plan. Inference 62, 279–300.

Lu, Z., 1998. On the geometric ergodicity of a non-linear autoregressive model with an autoregressive conditional heteroscedastic term. Stat. Sinica 8, 1205–1230.

Lu, Z., Jang, Z., 2001. L1 geometric ergodicity of a multivariate nonlinear AR model with an ARCH term. Stat. Probabil. Lett. 51, 121–130.

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

Luukkonen, R., Saikkonen, P., Teräsvirta, T., 1988. Testing linearity against smooth transition autoregressive models. Biometrika 75, 491–499.

Masry, E., Tjostheim, D., 1995. Nonparametric estimation and identification of nonlinear ARCH time series. Economet. Theory 11, 258–289.

McAleer, M., 2005. Automated inference and learning in modeling financial volatility. Economet. Theory 21, 232–261.

McAleer, M., Chan, F., Marinova, D., 2007. An econometric analysis of asymmetric volatility: theory and application to patents. J. Economet. 139, 259–284.

Medeiros, M.C., Veiga, A., 2000. A hybrid linear-neural model for time series forecasting. IEEE Trans. Neural Netw. 11, 1402–1412.

Medeiros, M.C., Veiga, A., 2005. A flexible coefficient smooth transition time series model. IEEE Trans. Neural Netw. 16, 97–113.

Meitz, M., Saikkonen, P., 2008. Stability of nonlinear AR-GARCH models. J. Time Series Anal. 29, 453–475.

Mira, S., Escribano, A., 2000. Nonlinear time series models: consistency and asymptotic normality of NLS under new conditions. In: Barnett, W.A., Hendry, D., Hylleberg, S., Teräsvirta, T., Tjøsthein, D., Würtz, A. (Eds.), Nonlinear Econometric Modeling in Time Series Analysis. Cambridge University Press, pp. 119–164.

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

Saikkonen, P., 2007. Stability of mixtures of vector autoregressions with autoregressive conditional heteroskedasticity. Stat. Sinica 17, 221–239.

Stout, W.F., 1974. Almost Sure Convergence. Academic Press, New York.

Suarez-Fariñas, M., Pedreira, C., Medeiros, M., 2004. Local-global neural network: a new approach for nonlinear time series modeling. J. Am. Stat. Assoc. 99, 1092–1107.

Teräsvirta, T., 1994. Specification, estimation, and evaluation of smooth transition autoregressive models. J. Am. Stat. Assoc. 89, 208–218.

Tong, H., 1978. On a threshold model. In: Chen, C.H. (Ed.), Pattern Recognition and Signal Processing. Sijthoff and Noordhoff, Amsterdam.

Tong, H., 1990. Non-linear Time Series: A Dynamical Systems Approach. Oxford University Press, Oxford.

Tong, H., Lim, K., 1980. Threshold autoregression, limit cycles and cyclical data (with discussion). J. R. Stat. Soc. Ser. B 42, 245–292.

Tweedie, R., 1988. Invariant measure for Markov chains with no irreducibility assumptions. J. Appl. Probabil. 25A, 275–285.

Van Dijk, D., Teräsvirta, T., Franses, P., 2002. Smooth transition autoregressive models – a survey of recent developments. Economet. Rev. 21, 1–47.

White, H., 1994. Estimation, Inference and Specification Analysis. Cambridge University Press, New York, NY.