Structure and Asymptotic Theory for Nonlinear Models with GARCH Errors*

Felix Chan
School of Economics and Finance
Curtin University of Technology

Michael McAleer
Econometric Institute
Erasmus School of Economics
Erasmus University Rotterdam
and Tinbergen Institute
The Netherlands
and
Institute of Economic Research
Kyoto University
Japan

Marcelo C. Medeiros
Department of Economics
Pontifical Catholic University of Rio de Janeiro

EI 2010-79
Revised: December 2010
Preliminary and Incomplete

Abstract

Nonlinear time series models, especially those with regime-switching and conditionally heteroskedastic errors, have become increasingly popular in the economics and finance literature. However, much of the

*This paper 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.

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 first derive necessary conditions for strict stationarity and ergodicity of three different specifications of the first-order smooth transition autoregressions with heteroskedastic errors. This is important, among other reasons, to establish the conditions under which the traditional LM linearity tests based on Taylor expansions are valid. Second, we 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.
KEYWORDS: Nonlinear time series, regime-switching, smooth transition, STAR, GARCH, log-moment, moment conditions, asymptotic theory.

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, Saikkonen, and Teräsvirta (1988) and Teräsvirta (1994), has found a number of successful applications (see van Dijk, Teräsvirta, and Franses (2002) 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 time-varying symmetric 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 the 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), 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, Li, and Li (1997), Ling (1999), and, Cline (2007). Of these articles, those of Liu, Li, and Li (1997) and Ling (1999) are restrcited to threshold AR-GARCH models, whereas the one by Cline (2007) analyses 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 are difficult to verify. A threshold AR-GARCH model is the only example that is explicitly treated by the authors. Furthermore, conditional heteroskedasticity is driven by the observed series instead of the autoregressive errors as in the usual GARCH specification. Ferrante, Fonseca, and Vidoni (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 analyze 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 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 parameters for GARCH models, Lee and Hansen (1994) and Lumsdaine (1996) proved that the local Quasi-Maximum Likelihood Estimator (QMLE) was consistent and asymptotic 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, Horváth, and Kokoszka (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, Chan, and Marinova (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 the 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 structural and statistical properties developed in this paper can also be used to derive the distributions associated with various test statistics proposed in the nonlinear time series literature. These properties provide the foundation for developing more complete tests for important economic and financial hypotheses. For instance, the correlation between prices over time is often used as a test for the weak form of the Efficient Market Hypothesis (EMH), which assumes that prices follow a linear process. However, if prices follow a nonlinear process, such as a STAR-type process, the correlation between prices over time may appear insignificant in finite samples. Thus, formal tests of nonlinear dependence would also provide an important diagnostic for testing the EMH.

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 and an empirical illustration is shown in Section 6. Finally, Section 7 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, Pedreira, and Medeiros (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 van Dijk, Teräsvirta, and Franses (2002). Finally, the third specification is the Exponential STAR-GARCH (ESTAR-

GARCH) model, of which the Exponential STAR (ESTAR) Teräsvirta (1994) model 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*

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

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

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

where $\{\eta_t\}$ is a sequence of independently and identically distributed zero mean and unit variance random variables, $\eta_t \sim \mathsf{IID}(0,1)$ 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 equations (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

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

   where

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

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

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

   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

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

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* $\beta = 0.85$. *Figure 1 shows the scatter plot* $f_0(y_{t-1})$ *and* $f_1(y_{t-1})$ *versus* y_{t-1}. *One characteristic of such specification is that the linear parameters in each limiting regimes are allowed to be different.*

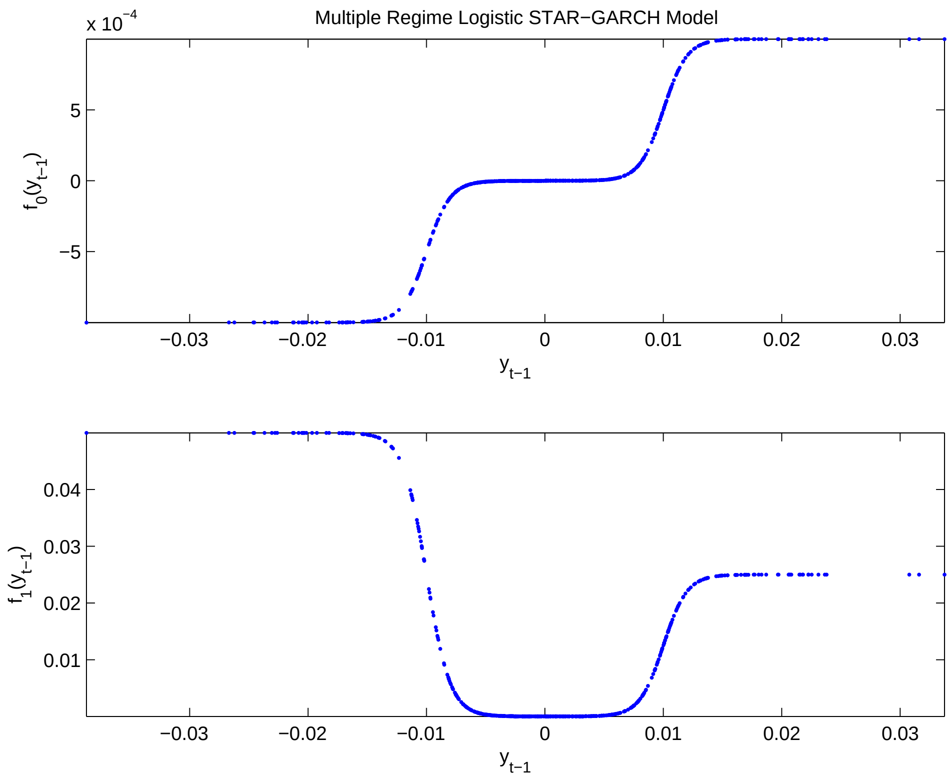

Figure 1: Upper panel: $f_0(y_{t-1})$ versus y_{t-1} for one realization of the model described in Example 1. Lower panel: $f_1(y_{t-1})$ versus y_{t-1} for one realization of the model described in Example 1.

EXAMPLE 2. *Consider a three regime GSTAR(1)-GARCH(1,1) model where the transition variable is* y_{t-1}, $\phi_{00} = -0.001$, $\phi_{10} = 0.002$, $\phi_{01} = 0.025$, $\phi_{11} = 0.0.25$, $\gamma = 100000$, $c_1 = -0.01$, $c_2 = 0.01$, $\omega = 10^{-5}$, $\alpha = 0.05$, *and* $\beta = 0.85$. *Figure 2 shows the scatter plot* $f_0(y_{t-1})$ *and* $f_1(y_{t-1})$ *versus* y_{t-1}. *Contrary to the MLSTAR model, the linear parameters in each limiting extreme regime are restricted to be equal. Furthermore,*

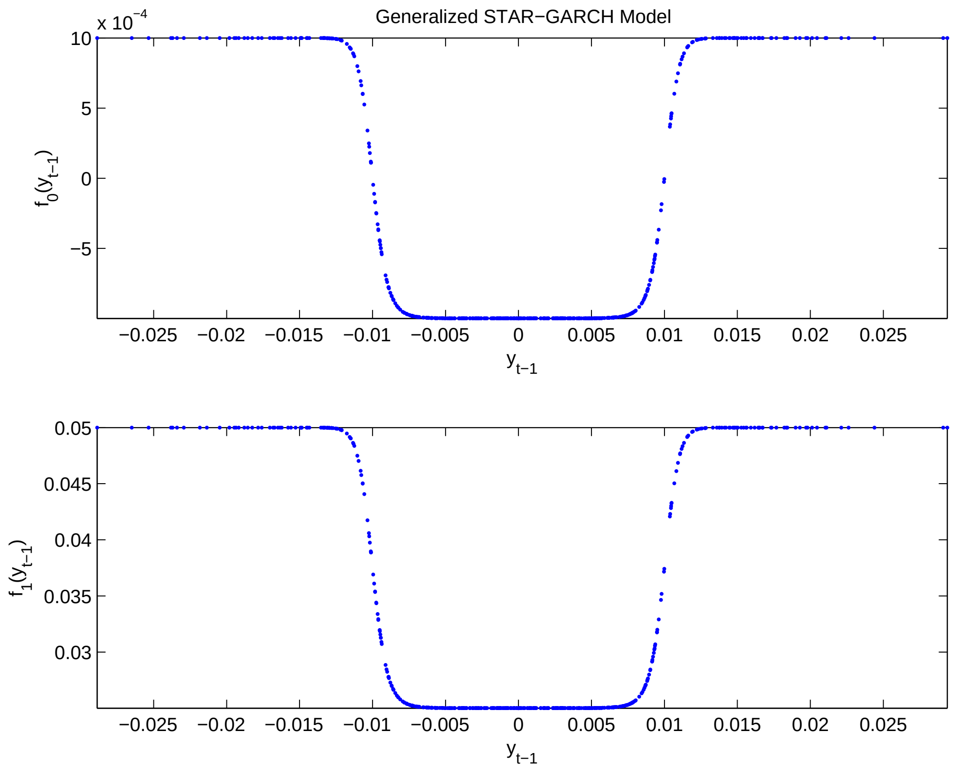

Figure 2: Upper panel: $f_0(y_{t-1})$ versus y_{t-1} for one realization of the model described in Example 2. Lower panel: $f_1(y_{t-1})$ versus y_{t-1} for one realization of the model described in Example 2.

EXAMPLE 3. *Consider a three regime ESTAR(1)-GARCH(1,1) model where the transition variable is* y_{t-1}, $\phi_{00} = -0.001$, $\phi_{10} = 0.002$, $\phi_{01} = 0.025$, $\phi_{11} = 0.0.25$, $\gamma = 100000$, $c = 0$, $\omega = 10^{-5}$, $\alpha = 0.05$, *and* $\beta = 0.85$. *Figure 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.*

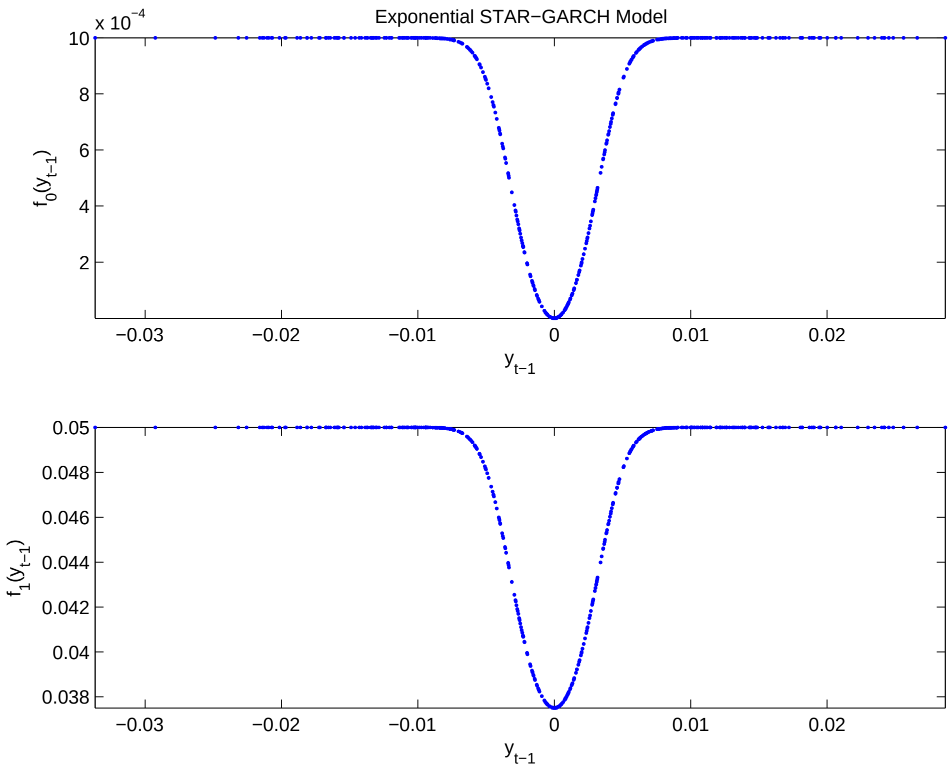

Figure 3: Upper panel: $f_0(y_{t-1})$ versus y_{t-1} for one realization of the model described in Example 3. Lower panel: $f_1(y_{t-1})$ versus y_{t-1} for one realization of the model described in Example 3.

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* $\mathsf{IID}(0,1)$ *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 Equation (1).*

ASSUMPTION 3 (Identifiability and Positiveness of the Variance). *The parameters of the model defined by (1)–(3) satisfy the following conditions: (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); (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 6 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

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

where

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

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 1–2, and if (R.1a) in Assumption 6 holds, the process $\{y_t, t \in \mathbb{Z}\}$ defined by equations (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 2 holds, the process $\{y_t, t \in \mathbb{Z}\}$ defined by equations (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 2 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 $\left|\overline{\phi}\right| < 1$. Furthermore, the process $\{\mathbf{z}_t, t \in \mathbb{Z}\}$ admits a unique causal expansion.*

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.

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 $\mathbf{y}_{t-1} = (y_{t-1}, \dots, y_{t-p})'$ and $\eta_t \sim \mathsf{IID}(0, 1)$.

ASSUMPTION 5. *The nonlinear function $g(\mathbf{y}_{t-1}; \boldsymbol{\lambda})$ satisfy 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 $varepsilon_t$ are independent.*
4. *$\mathbb{E}|g(\mathbf{y}_{t-1}; \boldsymbol{\lambda})|^q < \infty$,$q = 1, 2, 4$.*
5. *$\mathbb{E}\left\{\exp\left[g\left(\mathbf{y}_{t-1}; \boldsymbol{\lambda}\right)\right]^q\right\} < \infty$, $q = 1, 2, 4$.*
6. *$\mathbb{E}\left|\frac{\partial}{\partial \boldsymbol{\lambda}} g(\mathbf{y}_{t-1}; \boldsymbol{\lambda})\right|^q < \infty$, $q = 1, 2, 4$.*
7. *$\mathbb{E}\left|\frac{\partial^2}{\partial \boldsymbol{\lambda} \partial \boldsymbol{\lambda}'} g(\mathbf{y}_{t-1};\, \boldsymbol{\lambda})\right|^q < \infty$, $q = 1, 2$.*

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:

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

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 $\{(\varepsilon_{u,t}, h_{u,t}) : \ t = 0, \pm 1, \pm 2, \ldots\}$.

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

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

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

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

and

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

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 Subsection 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 Subsection 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 $\boldsymbol{\psi}_0$, $\widehat{\boldsymbol{\psi}} \stackrel{a.s.}{\to} \boldsymbol{\psi}_0$.*

4.3 Asymptotic Normality

First, we introduce the following matrices:

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

and

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

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

Consider the additional assumption:

ASSUMPTION 7. *There exists no set Λ of cardinal 2 such that $\Pr[\eta_t \in \Lambda] = 1$.*

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

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

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

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

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

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

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

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

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.*

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 bellow:

- Model 1: MLSTAR(1)-GARCH(1,1)

  A three regime model where the transition variable is y_{t-1}, $\phi_{00} = -0.001$, $\phi_{10} = 0.001$, $\phi_{20} = 0.001$, $\phi_{01} = -0.001$, $\phi_{11} = 0.001$, $\phi_{21} = 0.001$, $\gamma_1 = 1000$, $\gamma_2 = 1000$, $c_1 = -0.01$, $c_2 = 0.01$, $\omega = 10^{-5}$, $\alpha = 0.05$, and $\beta = 0.85$.

- Model 2: GSTAR(1)-GARCH(1,1) A three regime model where the transition variable is y_{t-1}, $\phi_{00} = -0.001$, $\phi_{10} = 0.002$, $\phi_{01} = 0.025$, $\phi_{11} = 0.0.25$, $\gamma = 100000$, $c_1 = -0.01$, $c_2 = 0.01$, $\omega = 10^{-5}$, $\alpha = 0.05$, and $\beta = 0.85$.

- Model 3: ESTAR(1)-GARCH(1,1) Consider a two regime model where the transition variable is y_{t-1}, $\phi_{00} = -0.001$, $\phi_{10} = 0.002$, $\phi_{01} = 0.025$, $\phi_{11} = 0.0.25$, $\gamma = 100000$, $c = 0$, $\omega = 10^{-5}$, $\alpha = 0.05$, and $\beta = 0.85$.

The results are illustrated in Table 1.

6 Empirical Illustration

7 Concluding Remarks

In this paper we have derived the necessary and sufficient conditions for strict stationarity and geometric ergodicity of three different classes of first-order STAR-GARCH models, and the sufficient

Table 1: SIMULATION: ESTIMATION RESULTS.

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 200 and 1000 observations.

		200 observations					
		Model 1		Model 2		Model 3	
Parameter	True Value	Mean	Std. Dev.	Mean	Std. Dev.	Mean	Std. Dev.
ϕ_{00}							
ϕ_{10}							
ϕ_{20}							
ϕ_{01}							
ϕ_{11}							
ϕ_{21}							
ω							
α							
β							

		1000 observations					
		Model 1		Model 2		Model 3	
Parameter	True Value	Mean	Std. Dev.	Mean	Std. Dev.	Mean	Std. Dev.
ϕ_{00}							
ϕ_{10}							
ϕ_{20}							
ϕ_{01}							
ϕ_{11}							
ϕ_{21}							
ω							
α							
β							

conditions for the existence of moments. 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 strongly 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, Petruccelli, Tong, and Woolford (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:

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

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

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

where

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

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:

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

where f is a nonlinear function such that $f(y_{t-1},...,y_{t-p})$ 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 $Z_t = (y_t, y_{t-1}, ..., y_{t-p}, h_t)'$ and

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

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

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

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 $a(x) = (a_1(x), a_2(x), ..., a_p(x))'$ and define the $(p+1)\times(p+1)$ matrix such that

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

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 $h_k (k = 1, 2, ...)$ defined by $h_k = g(0, h_{k-1})$, $k = 1, 2, ...$ 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 $e_1^{(0)}, ..., e_{p+2}^{(0)}$ such that the $(p+2) \times (p+2)$ matrix

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

is non-singular.

PROPOSITION 1. *Under Assumptions (?)-(?), the Model as defined in equations (**??**) - (**??**) is geometrically ergodic in the sense of* **?**.

Proof: It is sufficient to verify Conditions 1 to 5 in Meitz and Saikkonen (2008). Condition 1 is satified by Assumption (?) 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 **?**, 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 $A_i \in \{\Phi_1, \Phi_2\} \forall i = 1, ..., k$. 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

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

Let the control sequence be $(e_1^{(0)},e_2^{(0)},e_3^{(0)})=(e_1,0,0)$ 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

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

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.1 Proof of Theorem 1

Theorem 2.1 in Chan, Petruccelli, Tong, and Woolford (1985). ■

A.2 Proof of Theorem 2

■

A.3 Proof of Theorem 3

■

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 $\mathsf{E}\left[\sup_{\boldsymbol{\psi}\in\boldsymbol{\Psi}}|h_{u,t}|\right]<\infty$ and $\mathsf{E}\left[\sup_{\boldsymbol{\psi}\in\boldsymbol{\Psi}}|y_{u,t}|\right]<\infty$. By Jensen´s inequality, $\mathsf{E}\left[\sup_{\boldsymbol{\psi}\in\boldsymbol{\Psi}}|\ln|h_{u,t}||\right]<\infty$. Thus, $\mathsf{E}\left[|l_{u,t}(\boldsymbol{\psi})|\right]<\infty\ \forall\,\boldsymbol{\psi}\in\boldsymbol{\Psi}$.

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

$$\max_{\boldsymbol{\psi}\in\boldsymbol{\Psi}} \left[\mathcal{L}(\boldsymbol{\psi}) - \mathcal{L}(\boldsymbol{\psi}_0)\right] = \max_{\boldsymbol{\psi}\in\boldsymbol{\Psi}} \left\{ \mathsf{E}\left[\ln\left(\frac{h_{0,t}}{h_{u,t}}\right) - \frac{\varepsilon_t^2}{h_{u,t}} + 1\right]\right\}. \tag{B.7}$$

Writing $\varepsilon_t = \varepsilon_t - \varepsilon_{0,t} + \varepsilon_{0,t}$, equation (B.7) becomes

$$\begin{aligned}\max_{\boldsymbol{\psi}\in\boldsymbol{\Psi}} \left[\mathcal{L}(\boldsymbol{\psi}) - \mathcal{L}(\boldsymbol{\psi}_0)\right] &= \max_{\boldsymbol{\psi}\in\boldsymbol{\Psi}} \left\{ \mathsf{E}\left[\ln\left(\frac{h_{0,t}}{h_{u,t}}\right) - \frac{h_{0,t}}{h_{u,t}} + 1\right] - \mathsf{E}\left[\frac{\left[\varepsilon_t - \varepsilon_{0,t}\right]^2}{h_{u,t}}\right] \right. \\ &\qquad \left. - \mathsf{E}\left[\frac{2\eta_t h_{0,t}^{1/2}\left(\varepsilon_t - \varepsilon_{0,t}\right)}{h_{u,t}}\right]\right\} \qquad \text{(B.8)} \\ &= \max_{\boldsymbol{\psi}\in\boldsymbol{\Psi}} \left\{ \mathsf{E}\left[\ln\left(\frac{h_{0,t}}{h_{u,t}}\right) - \frac{h_{0,t}}{h_{u,t}} + 1\right] - \mathsf{E}\left[\frac{\left[\varepsilon_t - \varepsilon_{0,t}\right]^2}{h_{u,t}}\right]\right\},\end{aligned}$$

where

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

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,

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

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, $\widehat{\boldsymbol{\psi}}_{u,T} \overset{a.s.}{\to} \boldsymbol{\psi}_0$ 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, $\widehat{\boldsymbol{\psi}}_{u,T} \overset{a.s.}{\to} \boldsymbol{\psi}_0$.

Lemma 2 shows that

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

implying that $\widehat{\boldsymbol{\psi}}_T \overset{a.s.}{\to} \boldsymbol{\psi}_0$. 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 $\mathbf{A}_T(\boldsymbol{\psi}) \overset{a.s.}{\to} \mathbf{A}(\boldsymbol{\psi}_0)$, for any sequence $\boldsymbol{\psi}_T$, such that $\boldsymbol{\psi}_T \overset{a.s.}{\to} \boldsymbol{\psi}_0$.

(8) The score vector satisfies

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

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 below, we prove that condition (8) is also satisfied. This completes the proof. ■

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,*

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

PROOF. Set $g(\mathbf{Y}_t,\boldsymbol{\psi})=l_{u,t}(\boldsymbol{\psi})-\mathsf{E}\left[l_{u,t}(\boldsymbol{\psi})\right]$, where $\mathbf{Y}_t=[y_t,y_{t-1},y_{t-2},\ldots]'$. 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,*

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

PROOF. First, write

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

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

such that

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

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

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

Thus,

$$\sup_{\boldsymbol{\psi}\in\boldsymbol{\Psi}} |h_t - h_{u,t}| \leq K_h\rho_1^t,\ a.s., \text{and}$$

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

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 $\log(x) \leq x - 1$,

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

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

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

This completes the proof. ■

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

$$\mathsf{E}\left[\left|\left.\frac{\partial l_t(\boldsymbol{\psi})}{\partial\boldsymbol{\psi}}\right|_{\boldsymbol{\psi}_0}\right|\right] < \infty, \tag{C.9}$$

$$\mathsf{E}\left[\left|\left.\frac{\partial l_t(\boldsymbol{\psi})}{\partial\boldsymbol{\psi}}\right|_{\boldsymbol{\psi}_0}\left.\frac{\partial l_t(\boldsymbol{\psi})}{\partial\boldsymbol{\psi}'}\right|_{\boldsymbol{\psi}_0}\right|\right] < \infty,\ \textit{and} \tag{C.10}$$

$$\mathsf{E}\left[\left|\left.\frac{\partial^2 l_t(\boldsymbol{\psi})}{\partial\boldsymbol{\psi}\partial\boldsymbol{\psi}'}\right|_{\boldsymbol{\psi}_0}\right|\right] < \infty. \tag{C.11}$$

PROOF. Set

$$\nabla_0 l_{u,t} \equiv \left.\frac{\partial l_{u,t}(\boldsymbol{\psi})}{\partial \boldsymbol{\psi}}\right|_{\boldsymbol{\psi}_0}, \quad \nabla_0 h_{u,t} \equiv \left.\frac{\partial h_{u,t}}{\partial \boldsymbol{\psi}}\right|_{\boldsymbol{\psi}_0}, \quad \nabla_0 \varepsilon_t \equiv \left.\frac{\partial \varepsilon_t}{\partial \boldsymbol{\psi}}\right|_{\boldsymbol{\psi}_0},$$

$$\nabla_0^2 l_{u,t} \equiv \left.\frac{\partial^2 l_{u,t}(\boldsymbol{\psi})}{\partial \boldsymbol{\psi}\partial \boldsymbol{\psi}'}\right|_{\boldsymbol{\psi}_0}, \quad \nabla_0^2 h_{u,t} \equiv \left.\frac{\partial^2 h_{u,t}}{\partial \boldsymbol{\psi}\partial \boldsymbol{\psi}'}\right|_{\boldsymbol{\psi}_0}, \quad \text{and } \nabla_0^2 \varepsilon_t \equiv \left.\frac{\partial^2 \varepsilon_t}{\partial \boldsymbol{\psi}\partial \boldsymbol{\psi}'}\right|_{\boldsymbol{\psi}_0}.$$

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

$$\frac{\partial \varepsilon_t}{\partial \boldsymbol{\lambda}} = -\frac{\partial f_0(y_{t-1};\boldsymbol{\lambda})}{\partial \boldsymbol{\lambda}} - \frac{\partial f_1(y_{t-1};\boldsymbol{\lambda})}{\partial \boldsymbol{\lambda}}y_{t-1}, \tag{C.12}$$

$$\frac{\partial^2 \varepsilon_t}{\partial \boldsymbol{\lambda}\partial \boldsymbol{\lambda}'} = -\frac{\partial^2 f_0(y_{t-1};\boldsymbol{\lambda})}{\partial \boldsymbol{\lambda}\partial \boldsymbol{\lambda}'} - \frac{\partial^2 f_1(y_{t-1};\boldsymbol{\lambda})}{\partial \boldsymbol{\lambda}\partial \boldsymbol{\lambda}'}y_{t-1}, \tag{C.13}$$

$$\frac{\partial h_{u,t}}{\partial \boldsymbol{\lambda}} = 2\alpha\sum_{i=0}^{\infty}\left(\beta^i \varepsilon_{t-1-i}\frac{\partial \varepsilon_{t-1-i}}{\partial \boldsymbol{\lambda}}\right), \text{ and} \tag{C.14}$$

$$\frac{\partial^2 h_{u,t}}{\partial \boldsymbol{\lambda}\partial \boldsymbol{\lambda}'} = 2\alpha\sum_{i=0}^{\infty}\beta^i\left(\varepsilon_{t-1-i}\frac{\partial^2 \varepsilon_{t-1-i}}{\partial \boldsymbol{\lambda}\partial \boldsymbol{\lambda}'} + \frac{\partial \varepsilon_{t-1-i}}{\partial \boldsymbol{\lambda}}\frac{\partial \varepsilon_{t-1-i}}{\partial \boldsymbol{\lambda}'}\right). \tag{C.15}$$

As the derivatives of the transition function are bounded, if the strict stationarity and ergodicity conditions hold, (C.12)–(C.15) 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,*

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

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

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

PROOF.

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

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

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

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

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,*

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

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, $\mathsf{E}\left[S_T\right]>0$. Using the central limit theorem of Stout (1974),

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

By the Cramér-Wold device,

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

By Lemma 5,

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

Thus,

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

This completes the proof. ■

References

AN, H., AND S. CHEN (1997): "A note on the ergodicity of non-linear autoregressive model," *Statistics and Probability Letters*, 34, 365–372.

AN, H., AND F. HUANG (1996): "The geometrical ergodicity of nonlinear autoregressive models," *Statistica Sinica*, 6, 943–956.

BACON, D. W., AND D. G. WATTS (1971): "Estimating the Transition Between Two Intersecting Lines," *Biometrika*, 58, 525–534.

BERKES, I., L. HORVÁTH, AND P. KOKOSZKA (2003): "GARCH Processes: Structure and Estimation," *Bernoulli*, 9, 201–227.

BHATTACHARYA, R., AND C. LEE (1995): "On the geometric ergodicity of nonlinear autoregressive models," *Statistics and Probability Letters*, 22, 311–315.

BOLLERSLEV, T. (1986): "Generalized Autoregressive Conditional Heteroskedasticity," *Journal of Econometrics*, 21, 307–328.

BOUGEROL, P., AND N. M. PICARD (1992): "Stationarity of GARCH Processes and of Some Non-negative Time Series," *Journal of Econometrics*, 52, 115–127.

BOUSSAMA, F. (2000): "Normalité Asymptotique de L´Estimateur du Pseudo-Maximum de Vraisemblance dún Modèle GARCH," *Comptes Rendus de l´Academie des Sciences, Série I*, 331, 81–84.

CHAN, F., AND M. MCALEER (2002): "Maximum Likelihood Estimation of STAR and STAR-GARCH Models: A Monte Carlo Analysis," *Journal of Applied Econometrics*, 17, 509–534.

CHAN, K. S., J. D. PETRUCCELLI, H. TONG, AND S. W. WOOLFORD (1985): "A Multiple-Threshold AR(1) Model," *Journal of Applied Probability*, 22(2), 267–279.

CHAN, K. S., AND H. TONG (1986): "On Estimating Thresholds in Autoregressive Models," *Journal of Time Series Analysis*, 7, 179–190.

CHEN, G., AND M. CHEN (2001): "On a class of nonlinear AR(p) models with nonlinear ARCH errors," *Australian and New Zealand Journal of Statistics*, 43, 445–454.

CHEN, R., AND R. S. TSAY (1991): "On the Ergodicity of TAR(1) Processes," *Annals of Applied Probability*, 1, 613–634.

——— (1993): "Functional Coefficient AutoRegressive Models," *Journal of the American Statistical Association*, 88, 298–308.

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

CLINE, D., AND H. PU (1998): "Verifying irreducibility and continuity of a nonlinear time series," *Statistics and Probability Letters*, 40, 139–148.

——— (1999): "Geometric ergodicity of nonlinear time series," *Statistica Sinica*, 9, 1103–1118.

——— (2004): "Stability and the Lyapounov exponent of threshold AR-ARCH models," *Annals of Applied Probability*, 14, 1920–1949.

COMTE, F., AND O. LIEBERMAN (2003): "Asymptotic Theory for Multivariate GARCH Processes," *Journal of Multivariate Analysis*, 84, 61–84.

ENGLE, R., AND K. KRONER (1995): "Multivariate Simultaneous Generalized ARCH," *Econometric 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., G. FONSECA, AND P. VIDONI (2003): "Geometric Ergodicity, Regularity of the Invariant Distribution and Inference for a Threshold Bilinear Markov Process," *Statistica Sinica*, 13, 367–384.

FRANCQ, C., AND J.-M. ZAKOÏAN (2004): "Maximum Likelihood Estimation of Pure GARCH and ARMA-GARCH Processes," *Bernoulli*, 10, 605–637.

GOLDFELD, S. M., AND R. QUANDT (1972): *Nonlinear Methods in Econometrics*. North Holland, Amsterdam.

HE, C., AND T. TERÄSVIRTA (1999a): "Fourth Moment Structure of the GARCH(p, q) Process," *Econometric Theory*, 15, 824–846.

——— (1999b): "Properties of Moments of a Family of GARCH Processes," *Journal of Econometrics*, 92, 173–192.

HWANG, S., AND M.-J. WOO (2001): "Threshold ARCH(1) processes: asymptotic inference," *Statistics and Probability Letters*, 53, 11–20.

JEANTHEAU, T. (1998): "Strong Consistency of Estimators for Multivariate ARCH Models," *Econometric Theory*, 14, 70–86.

LEE, C. (1998): "Asymptotics of a class of pth-order nonlinear autoregressive processes," *Statistics and Probability Letters*, 40, 171–177.

LEE, S.-W., AND B. E. HANSEN (1994): "Asymptotic Theory for The GARCH(1,1) Quasi-Maximum Likelihood Estimator," *Econometric Theory*, 10, 29–52.

LI, W. K., AND K. LAM (1995): "Modelling Asymetry in Stock Returns by a Threshold Autoregressive Conditional Heterocedastic Model," *The Stastician*, 44, 333–341.

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. (1999): "On the Probabilitistic Properties of a Double Threshold ARMA Conditional Heteroscedasticity Model," *Journal of Applied Probability*, 36, 1–18.

LING, S., AND M. MCALEER (2002a): "Necessary and Sufficient Moment Conditions for the GARCH(r, s) and Asymmetric Power GARCH(r, s) Models," *Econometric Theory*, 18, 722–729.

——— (2002b): "Stationarity and The Existence of Moments of a Family of GARCH Processes," *Journal of Econometrics*, 106, 109–117.

——— (2003): "Asymptotic Theory for a Vector ARMA-GARCH Model," *Econometric Theory*, 19, 280–310.

LIU, J., W. K. LI, AND C. W. LI (1997): "On a Threshold Autoregression with Conditional Heteroscedastic Variances," *Journal of Statistical Planning and Inference*, 62, 279–300.

LU, Z. (1998): "On the geometric ergodicity of a non-linear autoregressive model with an autoregressive conditional heteroscedastic term," *Statistica Sinica*, 8, 1205–1230.

LU, Z., AND Z. JANG (2001): "L1 geometric ergodicity of a multivariate nonlinear AR model with an ARCH term," *Statistics and Probability Letters*, 51, 121–130.

LUMSDAINE, R. (1996): "Consistency and Asymptotic Normality of the Quasi-Maximum Likelihood Estimator in IGARCH(1,1) and Covariance Stationary GARCH(1,1) Models," *Econometrica*, 64, 575–596.

LUUKKONEN, R., P. SAIKKONEN, AND T. TERÄSVIRTA (1988): "Testing Linearity Against Smooth Transition Autoregressive Models," *Biometrika*, 75, 491–499.

MASRY, E., AND D. TJOSTHEIM (1995): "Nonparametric estimation and identification of nonlinear ARCH time series," *Econometric Theory*, 11, 258–289.

MCALEER, M. (2005): "Automated Inference and Learning in Modeling Financial Volatility," *Econometric Theory*, 21, 232–261.

MCALEER, M., F. CHAN, AND D. MARINOVA (2007): "An Econometric Analysis of Asymmetric Volatility: Theory and Application to Patents," *Journal of Econometrics*, 139, 259–284.

MEDEIROS, M. C., AND A. VEIGA (2000): “A Hybrid Linear-Neural Model for Time Series Forecasting,” *IEEE Transactions on Neural Networks*, 11, 1402–1412.

——— (2005): “A Flexible Coefficient Smooth Transition Time Series Model,” *IEEE Transactions on Neural Networks*, 16, 97–113.

MEITZ, M., AND P. SAIKKONEN (2008): “Stability of Nonlinear AR-GARCH Models,” *Journal of Time Series Analysis*, 29, 453–475.

MIRA, S., AND A. ESCRIBANO (2000): “Nonlinear Time Series Models: Consistency and Asymptotic Normality of NLS Under New Conditions,” in *Nonlinear Econometric Modeling in Time Series Analysis*, ed. by W. A. Barnett, D. Hendry, S. Hylleberg, T. Teräsvirta, D. Tjøsthein, and A. Würtz, pp. 119–164. Cambridge University Press.

NELSON, D. B. (1990): “Stationarity and Persistence in the GARCH(1,1) Model,” *Econometric Theory*, 6, 318–334.

SAIKKONEN, P. (2007): “Stability of mixtures of vector autoregressions with autoregressive conditional heteroskedasticity,” *Statistica Sinica*, 17, 221–239.

STOUT, W. F. (1974): *Almost Sure Convergence*. Academic Press, New York.

SUAREZ-FARIÑAS, M., C. PEDREIRA, AND M. MEDEIROS (2004): “Local-global neural network: A new approach for nonlinear time series modeling,” *Journal of the American Statistical Association*, 99, 1092–1107.

TERÄSVIRTA, T. (1994): “Specification, Estimation, and Evaluation of Smooth Transition Autoregressive Models,” *Journal of the American Statistical Association*, 89, 208–218.

TONG, H. (1978): “On a Threshold Model,” in *Pattern Recognition and Signal Processing*, ed. by C. H. Chen, Amsterdam. Sijthoff and Noordhoff.

——— (1990): *Non-linear Time Series: A Dynamical Systems Approach*. Oxford University Press, Oxford.

TONG, H., AND K. LIM (1980): “Threshold Autoregression, Limit Cycles and Cyclical Data (with Discussion),” *Journal of the Royal Statistical Society, Series B*, 42, 245–292.

TWEEDIE, R. (1988): “Invariant Measure for Markov Chains with No Irreducibility Assumptions,” *Journal of Applied Probability*, 25A, 275–285.

VAN DIJK, D., T. TERÄSVIRTA, AND P. FRANSES (2002): “Smooth Transition Autoregressive Models – A Survey of Recent Developments,” *Econometric Reviews*, 21, 1–47.

WHITE, H. (1994): *Estimation, Inference and Specification Analysis*. Cambridge University Press, New York, NY.