Munich Personal RePEc Archive

Semiparametric Efficient Adaptive Estimation of the PTTGARCH model

Ciccarelli, Nicola

Tilburg University

2016

Online at https://mpra.ub.uni-muenchen.de/72021/
MPRA Paper No. 72021, posted 15 Jun 2016 16:31 UTC

Semiparametric Efficient Adaptive Estimation of the PTTGARCH model

Nicola Ciccarelli[*]

Tilburg University

Abstract

Financial data sets exhibit conditional heteroskedasticity and asymmetric volatility. In this paper we derive a semiparametric efficient adaptive estimator of a conditional heteroskedasticity and asymmetric volatility GARCH-type model (i.e., the PTTGARCH(1,1) model). Via kernel density estimation of the unknown density function of the innovation and via the Newton-Raphson technique applied on the $\sqrt{n}$-consistent quasi-maximum likelihood estimator, we construct a more efficient estimator than the quasi-maximum likelihood estimator. Through Monte Carlo simulations, we show that the semiparametric estimator is adaptive for parameters included in the conditional variance of the model with respect to the unknown distribution of the innovation.

JEL Classification: C14; C22.

Keywords: Semiparametric adaptive estimation; Power-transformed and threshold GARCH.

1 Introduction

Stock prices and other asset prices exhibit conditional heteroskedasticity, that is, volatility shocks are clustered in some periods while other periods are characterized by low volatility. Engle (1982) proposed the ARCH model to allow for the presence of conditional heteroskedasticity in asset prices. Bollerslev (1986) generalized Engle (1982)'s idea with the GARCH(p,q) model. Additionally, Black (1976), Christie (1982), Engle (1990) and Engle and Ng (1993) show that stock market prices are affected by asymmetric volatility. Essentially, the latter authors show that stock prices tend to have higher volatility in the case of negative news than in the case of positive news, leading to an asymmetric evolution of volatility through time. In order to capture the presence of volatility asymmetry, many models were proposed starting with the Taylor-Schwert GARCH model [see Taylor (1986) and Schwert (1989) for more details], the AGARCH(1,1) model proposed by Engle (1990) and the EGARCH model proposed by Nelson (1991). One of the most recent asymmetric volatility models -i.e., the PTTGARCH(p,q) model- was introduced by PWT (2008); the PTTGARCH(p,q) model is a very flexible model and, under certain

[*]CentER and Department of Econometrics and Operations Research, Tilburg University, P.O. Box 90153, 5000 LE Tilburg, The Netherlands: Phone: +31 13 466 3254, E-mail: ciccarelli.nicola@gmail.com.

conditions, it includes several ARCH/GARCH models such as Bollerslev (1986)'s GARCH(p,q) model. Engle and Ng (1993) and Awartani and Corradi (2005) compare the standard GARCH model with several asymmetric GARCH-type models, and they conclude that asymmetric GARCH-type models are more appropriate than the symmetric GARCH(1,1) model to forecast financial markets volatility.

Building upon these results we construct a semiparametric efficient adaptive estimator for the PTTGARCH model, and, through a series of Monte Carlo simulations, we show that this estimator regains most of the efficiency loss of the inefficient quasi-maximum likelihood estimator whenever the true innovation is not distributed as a standard normal random variable or as a Gaussian mixture random variable. For simplicity we restrict our attention to the PTTGARCH(1,1) model; anyhow, the results can be generalized to the PTTGARCH(p,q) case.

The paper is organized along the following lines. Section 2 contains two main results. First, we show that the quasi-maximum likelihood estimator is consistent and asymptotically normal in the case of the PTTGARCH(1,1) model. Second, we show that the log-likelihood ratio of the PTTGARCH(1,1) model satisfies the Local Asymptotic Normality (LAN) condition. Through the Convolution theorem, the semiparametric estimator -which is built upon the quasi-maximum likelihood estimator- is shown to achieve the semiparametric lower bound for the variance. In Section 3 we show how to compute the theoretical asymptotic variance of the maximum likelihood (ML) and quasi-maximum likelihood (QML) estimators in the case of the PTTGARCH(1,1) model. The theoretical asymptotic variance of the maximum likelihood (ML) and quasi-maximum likelihood (QML) estimators are obtained through simulation, and can be used to check whether Monte Carlo simulations are appropriately set up. In Section 4 we report the results of Monte Carlo simulations; in this section we compare estimators in terms of their efficiency, and we check whether the asymptotic results of the ML and QML estimators from Monte Carlo experiments approximate well enough the theoretical asymptotic behavior of the ML and QML estimators. Section 5 concludes.

2 The Semiparametric Efficient Adaptive Estimator for the PTTGARCH(1,1) model

The power-transformed and threshold GARCH(p, q) model [henceforth PTTGARCH(p, q) model] was introduced by PWT (2008) [see equation (1.2) of PWT (2008), page 353], and it includes several conditional heteroskedasticity models. In this paper we use the PTTGARCH(1,1) model without power-transformation, that is, we set $\delta = 1$ and, for notational simplicity, we set $p = 1$ and $q = 1$ in equation (1.2) of PWT (2008). The PTTGARCH(1,1) model used in this paper incorporates conditional heteroskedasticity and asymmetric volatility. Let $\{Y_t\}$ denote an observed real-valued discrete-time stochastic process; the stochastic process $\{Y_t\}$ is conditionally heteroschedastic:

$$Y_t = \sqrt{h_t}\eta_t, \;\; \eta_t \sim iid(0, 1), \tag{2.1}$$

where the unobservable heteroskedasticity factors $\{h_t\}_{t\in\mathbb{Z}}$ follow a PTTGARCH(1,1) process:

$$h_t = \omega + \beta h_{t-1} + \alpha_+ (Y_{t-1}^+)^2 + \alpha_- (Y_{t-1}^-)^2, \tag{2.2}$$

where (and in the sequel) the notation [see Hwang and Kim (2004), page 296]

$$a_t^+ = \max(a_t, 0) \text{ and } a_t^- = \max(-a_t, 0) \text{ is used so that } a_t = a_t^+ - a_t^-,$$

and where $\omega > 0$, $\alpha_+ \geq 0$, $\alpha_- \geq 0$, $\beta \geq 0$, $h_1 = \frac{\omega}{1-\beta-\frac{1}{2}(\alpha_+)-\frac{1}{2}(\alpha_-)}$.

It is not possible to obtain a semiparametric estimation of the PTTGARCH(1,1) model in eq. (2.1)-(2.2) that is fully efficient since the score space is not orthogonal to the tangent space generated by the nuisance parameter (i.e., the distribution of the standardized innovation), thus a reparametrization of the PTTGARCH(1,1) model is needed to obtain a semiparametric efficient adaptive estimator. We set $\omega = 1$ and we introduce the location parameter μ and the scale parameter σ in the PTTGARCH(1,1) model so that the orthogonality relation (3.6) of DKW (1997) is satisfied.[1]

Let $\mu \in \mathbb{R}$, $\sigma > 0$, $\alpha_+ > 0$, $\alpha_- > 0$, $\beta > 0$ be parameters[2] and let $\{\eta_t : t \in \mathbb{Z}\}$ be an i.i.d. sequence of innovation errors with zero mean, unit variance and density f. Put $\xi_t = \mu + \sigma\eta_t$, thus ξ_t is a random variable with location μ, scale σ and density $\sigma^{-1} f(\{\xi_t - \mu\}/\sigma)$. Consider the reparametrized PTTGARCH(1,1) model with observations

$$Y_t = \sqrt{h_t}\xi_t = \mu\sqrt{h_t} + \sigma\sqrt{h_t}\eta_t, \qquad \eta_t \sim iid(0,1), \tag{2.3}$$

where the unobservable heteroskedasticity factors $\{h_t\}_{t\in\mathbb{Z}}$ follow a PTTGARCH(1,1) process:

$$h_t = 1 + \beta h_{t-1} + \alpha_+ (Y_{t-1}^+)^2 + \alpha_- (Y_{t-1}^-)^2 = 1 + h_{t-1}\left[\beta + \alpha_+ \left(\xi_{t-1}^+\right)^2 + \alpha_- \left(\xi_{t-1}^-\right)^2\right], \tag{2.4}$$

where the notation [see Hwang and Kim (2004), page 296]

$$a_t^+ = \max(a_t, 0) \text{ and } a_t^- = \max(-a_t, 0) \text{ is used so that } a_t = a_t^+ - a_t^-,$$

where $h_1 = \frac{1}{1-\frac{1}{2}(\alpha_+)-\frac{1}{2}(\alpha_-)-\beta}$, and where $\alpha_+ > 0$, $\alpha_- > 0$, $\beta > 0$, $\mu \in \mathbb{R}$ and $\sigma > 0$.[3] Observe that the Euclidean parameter $\theta = (\alpha_+, \alpha_-, \beta, \mu, \sigma)' \in \Theta \subset \mathbb{R}^5$ is identifiable.

[1] Similar reparametrizations are used in DK (1997) for the GARCH(1,1) model and in Sun and Stengos (2006) for the AGARCH(1,1) model.

[2] From this point onwards in this paper, unless explicitly stated otherwise, we restrict our attention to the case of strictly positive conditional variance parameters (i.e., $\alpha_+ > 0$, $\alpha_- > 0$, $\beta > 0$). We assume that parameters α_+ and α_- are strictly positive as it is needed for the consistency of the quasi-maximum likelihood estimator [see Assumption QML4 in Section 2.4], and, for simplicity, we assume that β is strictly positive.

[3] See Footnote 2.

We assume that the observed data Y_t are stationary given the starting value $h_1(\theta) = h_{01}$ initializing (2.4). The necessary and sufficient condition for a unique strictly stationary and ergodic solution for the (non-reparametrized) PTTGARCH(1,1) model is reported in Liu (2006) [Theorem 2.1, page 1324]. We adapt the necessary and sufficient condition given in Liu (2006) to the case of the reparametrized PTTGARCH(1,1) model.

Assumption 2.1. ***Strict Stationarity of the Reparametrized PTTGARCH(1,1) Model Based on Theorem 2.1 of Liu (2006)***

$$\mathrm{E}\ln\left[\beta + \alpha_+(\xi_t^+)^2 + \alpha_-(\xi_t^-)^2\right] < 0. \tag{2.5}$$

If $\mathrm{E}\Big[\alpha_+(\xi_{t-k}^+)^2 + \alpha_-(\xi_{t-k}^-)^2 + \beta)\Big] < 1$, then the necessary and sufficient condition for the strict stationarity of $\{h_t : t \in Z\}$ reported in eq. (2.5) is satisfied [see Assumption QML3 in Section 2.4 for the proof].

2.1 The Data Generating Process

We assume that there is an underlying probability distribution, or data generating process (DGP), for the observable Y_t and a true parameter vector $\theta_0 = (\alpha_{0+}, \alpha_{0-}, \beta_0, \mu_0, \sigma_0)'$ which is a characteristic of that DGP, with

$$Y_t = \mu_0\sqrt{h_t} + \sigma_0\sqrt{h_t}\eta_t, \ \eta_t \sim \mathrm{iid}(0, 1),$$

$$h_t = 1 + \beta_0 h_{t-1} + \alpha_{0+}(Y_{t-1}^+)^2 + \alpha_{0-}(Y_{t-1}^-)^2 = 1 + h_{t-1}\left[\beta_0 + \alpha_{0+}\left(\xi_{t-1}^+\right)^2 + \alpha_{0-}\left(\xi_{t-1}^-\right)^2\right],$$

where $a_t^+ = \max(a_t, 0)$ and $a_t^- = \max(-a_t, 0)$ is used so that $a_t = a_t^+ - a_t^-$ [see Hwang and Kim (2004), page 296], where $h_1 = \frac{1}{1-\frac{1}{2}(\alpha_+)-\frac{1}{2}(\alpha_-)-\beta}$, and where $\alpha_+ > 0$, $\alpha_- > 0$, $\beta > 0$, $\mu \in \mathbb{R}$ and $\sigma > 0$.

The probability density functions used for the innovation η_t are: Normal(0,1), Laplace(0,1), Balanced Mixture of two Normals, $\mathcal{N}(-2, 1)$ and $\mathcal{N}(+2, 1)$, t-student distributions with 5, 7, 9 degrees of freedom, Chi-squared distributions with 6 and 12 degrees of freedom; all densities are rescaled so that the innovation has zero mean and unit variance. The eight probability density functions for the standardized innovation η_t are reported in Table 1.

[Table 1]

2.2 The general form of the log-likelihood function of the PTTGARCH(1,1) model

The joint density function for a sample $\mathbf{Y} = (Y_1, ..., Y_T)$ is the product of the conditional densities of $\{Y_t : t = 2, ..., T\}$ and of the marginal density of Y_1:

$$\begin{aligned} f(\mathbf{Y}; \theta) = f(Y_1, ..., Y_T; \theta) = f(Y_T | Y_{T-1}, ..., Y_1; \theta) \times ... \\ \times f(Y_{T-1} | Y_{T-2}, ... Y_1; \theta) \times ... \times f(Y_2 | Y_1; \theta) \times f(Y_1; \theta) \end{aligned} \tag{2.6}$$

Equation (2.6) can be rewritten as:

$$f(\mathbf{Y};\theta)=\left(\prod_{t=2}^{T} f(Y_t|I_{t-1};\theta)\right) f(Y_1;\theta)$$

where $I_t=\{Y_t,...,Y_1\}$ denotes the information available at time t, and Y_1 denotes the initial value. The log-likelihood function may then be expressed as:

$$\ln L(\theta|\mathbf{Y})=\sum_{t=2}^{T}\ln f(Y_t|I_{t-1};\theta)+\ln f(Y_1;\theta) \tag{2.7}$$

The first term contained in eq. (2.7) is the conditional log-likelihood function while the second term in eq. (2.7) is the marginal log-likelihood for the initial value. The marginal density of Y_1 is dropped from the conditional likelihood function because its contribution is negligible when the number of observations is large, thus eq. (2.7) can be rewritten as:

$$\ln L(\theta|\mathbf{Y})=\sum_{t=2}^{T}\ln f(Y_t|I_{t-1};\theta) \tag{2.8}$$

Given eq. (2.3), for any $t=2,...,T$, the conditional density of Y_t given $I_{t-1}=Y_{t-1},...,Y_1$ is:

$$f(Y_t|I_{t-1};\theta)=\frac{1}{\sigma\sqrt{h_t}}f(\eta_t)=\frac{1}{\sigma\sqrt{h_t}}f\left(\frac{Y_t-\mu\sqrt{h_t}}{\sigma\sqrt{h_t}}\right) \tag{2.9}$$

Using (2.8) and (2.9), the conditional log-likelihood function is:

$$\ln L(\theta|\mathbf{Y})=\sum_{t=2}^{T}\ln\left(\frac{1}{\sigma\sqrt{h_t}}f\left(\frac{Y_t-\mu\sqrt{h_t}}{\sigma\sqrt{h_t}}\right)\right) \tag{2.10}$$

2.3 The log-likelihood function of the PTTGARCH(1,1) model for each type of innovation

Using the probability density functions reported in Table 1 and using eq. (2.10) we can obtain the conditional log-likelihood function for each type of (zero mean and unit variance) innovation η_t. The conditional log-likelihood function for each type of (zero mean and unit variance) innovation is reported in Table 2.

[Table 2]

2.4 The Asymptotic Properties of the QML Estimator in the Case of the PTTGARCH(1,1) Model

Hamadeh and Zakoïan (2011) prove that, under assumptions QML0-QML6 reported below, the quasi-maximum likelihood estimator is consistent and asymptotically normal for the generic PTTGARCH(p,q) model in their Theorem 2.1 and Theorem 2.2, respectively. In this section we show that the assumptions needed for consistency and asymptotic normality of the quasi-maximum likelihood estimator are satisfied in the case of the reparametrized PTTGARCH(1,1) model studied in this paper. The consistency of the QML estimator is crucial to construct the semiparametric efficient adaptive estimator; see Section 2.6 for more details. *Assumptions QML0-QML6 reported below correspond to Assumptions A0-A6 of Hamadeh and Zakoïan (2011).*

Assumption QML0. *η_t is a sequence of independent and identically distributed (i.i.d.) random variables with $E|\eta_t|^r < \infty$ for some $r > 0$.*

First, η_t is generated as an i.i.d. random variable. Second, $E|\eta_t|^r$ can be rewritten as $E\left(\sqrt{\eta_t^2}\right)^r$; choosing $r = 2$ we obtain $E\left(\sqrt{\eta_t^2}\right)^r = E\left(\sqrt{\eta_t^2}\right)^2 = E\left(\eta_t^2\right)$. Since the innovation has zero mean and unit variance by construction, then $E\left(\eta_t^2\right) = 1 < \infty$. Thus, Assumption QML0 is satisfied.

Assumption QML1. *$\theta_0 \in \Theta$ and Θ is compact.*

The vector of the true parameters in the PTTGARCH(1.1) model is:

$$\theta_0 = (\alpha_{0+}, \alpha_{0-}, \beta_0, \mu_0, \sigma_0)'$$

and belongs to a parameter space $\Theta \subset]0, \infty[^4 \times \mathbb{R}$ with Θ compact since Θ is a bounded and closed set. Therefore, Assumption QML1 is satisfied.

Assumption QML2. *$E\eta_t^2 = 1$ and $P[\eta_t > 0] \in (0, 1)$. If $P(\eta_t \in \Gamma) = 1$ for a set Γ, then Γ has a cardinal $|\Gamma| > 2$.*

Assumption QML2 is satisfied by construction: the (rescaled) innovation is created such that $E\eta_t^2 = 1$ and such that the (rescaled) innovation includes both positive and negative values, thus $P[\eta_t > 0] \in (0, 1)$. The cardinality of the set Γ is $|\Gamma| > 2$ since the innovation η_t is generated as an i.i.d. random variable with a continuous probability density function.

Assumption QML3. *$\gamma(A_0) < 0$ and for all $\theta \in \Theta$, $\sum_{j=1}^{p} \beta_j < 1$.*

Assumption QML3 is satisfied if $E\left[\alpha_+(\xi_{t-k}^+)^2 + \alpha_-(\xi_{t-k}^-)^2 + \beta\right] < 1$. In order to show that Assumption QML3 is satisfied in the case of our model, we use the notation of Example 2.1 (GARCH(1,1)) of Francq and Zakoïan (2010) [page 34], and we use other results presented in Chapter 2.2 of Francq and Zakoïan (2010).

Proof. In the PTTGARCH(1,1) model the matrix A_t can be written as [see Francq and Zakoïan (2010), page 34]

$$A_t = ((\xi_t^+)^2, (\xi_t^-)^2, 1)' \, (\alpha_+, \alpha_-, \beta) \, .$$

We thus have

$$A_t A_{t-1} ... A_1 = \prod_{k=1}^{t-1} (\alpha_+(\xi_{t-k}^+)^2 + \alpha_+(\xi_{t-k}^-)^2 + \beta) A_t.$$

It follows that

$$\ln \|A_t A_{t-1} ... A_1\| = \sum_{k=1}^{t-1} \ln(\alpha_+(\xi_{t-k}^+)^2 + \alpha_+(\xi_{t-k}^-)^2 + \beta) + \ln \|A_t\|.$$

The top Lyapunov exponent $\gamma(A_0)$ is defined as [see eq. (2.23) in Francq and Zakoïan (2010)]

$$\gamma(A_0) = \lim_{t\to\infty} a.s. \frac{1}{t} \ln \|A_t A_{t-1} ... A_1\|.$$

By the strong law of large numbers, the top Lyapunov exponent can be written as [see Francq and Zakoïan (2010), page 34]

$$\gamma(A_0) = E \ln \left(\alpha_+(\xi_{t-k}^+)^2 + \alpha_-(\xi_{t-k}^-)^2 + \beta\right).$$

By the Jensen Inequality we obtain

$$\gamma(A_0) = E \ln \left(\alpha_+(\xi_{t-k}^+)^2 + \alpha_-(\xi_{t-k}^-)^2 + \beta\right) \leq \ln \left[E\left(\alpha_+(\xi_{t-k}^+)^2 + \alpha_-(\xi_{t-k}^-)^2 + \beta\right)\right].$$

Using the assumption $E\left[\alpha_+(\xi_{t-k}^+)^2 + \alpha_-(\xi_{t-k}^-)^2 + \beta\right] < 1$, we finally obtain

$$\gamma(A_0) = E \ln \left(\alpha_+(\xi_{t-k}^+)^2 + \alpha_-(\xi_{t-k}^-)^2 + \beta\right) \leq \ln E\left(\alpha_+(\xi_{t-k}^+)^2 + \alpha_-(\xi_{t-k}^-)^2 + \beta\right) < 0$$

The second part of Assumption QML3 reduces in PTTGARCH(1,1) model to $\beta < 1$ and this assumption is satisfied since $\gamma(A_0) < 0$ as shown above and $\alpha_+ > 0$ and $\alpha_- > 0$ [see the restrictions below eq. (2.4)]. $\square$

Let $\mathcal{A}_{\theta+}(z) = \sum_{i=1}^q \alpha_{i+} z^i$, $\mathcal{A}_{\theta-}(z) = \sum_{i=1}^q \alpha_{i-} z^i$ and $\mathcal{B}_\theta(z) = 1 - \sum_{j=1}^p \beta_j z^j$. In the case of PTTGARCH(1,1) model the previous equations simplify to $\mathcal{A}_{\theta+}(z) = \alpha_+ z$, $\mathcal{A}_{\theta-}(z) = \alpha_- z$ and $\mathcal{B}_\theta(z) = 1 - \beta z$.

Assumption QML4. *If $p > 0$, $\mathcal{B}_{\theta_0}(z)$ has no common root with $\mathcal{A}_{\theta_0+}(z)$ and $\mathcal{A}_{\theta_0-}(z)$. Moreover $\mathcal{A}_{\theta_0+}(1) + \mathcal{A}_{\theta_0-}(1) \neq 0$ and $\alpha_{0+} + \alpha_{0-} + \beta_0 \neq 0$.*

Assumption QML4 is satisfied. $\mathcal{A}_{\theta_0+}(1) + \mathcal{A}_{\theta_0-}(1) \neq 0$ is verified since $\alpha_{0+} > 0$ and $\alpha_{0-} > 0$. The root of $\mathcal{A}_{\theta_0+}(z)$ is:

$$\mathcal{A}_{\theta_0+}(z) = 0 = \alpha_{0+} z \Rightarrow z = 0.$$

The root of $\mathcal{A}_{\theta_0-}(z)$ is:

$$\mathcal{A}_{\theta_0+}(z) = 0 = \alpha_{0-} z \Rightarrow z = 0.$$

and $B_{\theta_0}(0) \neq 0$. If $\beta_0 \neq 0$, the unique root of $B_{\theta_0}(z)$ is:

$$B_{\theta_0}(z) = 0 = 1 - \beta_0 z \Rightarrow z = 1/\beta_0.$$

The unique root of $B_{\theta_0}(z)$ is $1/\beta_0 > 0$ (if $\beta_0 = 0$, the polynomial does not admit any root) and because the coefficients α_{0+} and α_{0-} are positive, $\mathcal{A}_{\theta_0+}(1/\beta_0) \neq 0$ and $\mathcal{A}_{\theta_0-}(1/\beta_0) \neq 0$.

Given the above results, $\alpha_{0+} + \alpha_{0-} + \beta_0 \neq 0$.

Theorem 2.1.[Hamadeh and Zakoïan (2011)] *Let* $(\hat{\theta}_n^{QML})$ *be a sequence of QML estimators satisfying (2.1) of Hamadeh and Zakoïan (2011). Then, under assumptions QML0-QML4,* $\hat{\theta}_n^{QML} \to \theta_0$, *a.s. as* $n \to \infty$.

Proof. See Hamadeh and Zakoïan (2011), Proof of Theorem 2.1, pages 497-499. □

Two additional assumptions are needed for the asymptotic normality of the QML estimator.

Assumption QML5. $\theta_0 \in \dot{\Theta}$, *where* $\dot{\Theta}$ *denotes the interior of* Θ.

Assumption QML5 is satisfied: if $\hat{\theta}_n$ is consistent, it also belongs to the interior of Θ, for large n.

Assumption QML6. $\kappa_\eta := E\eta_t^4 < \infty$.

Assumption QML6 is satisfied since the kurtosis of each (zero mean and unit variance) innovation η_t is finite; see Table 3 for the kurtosis of each innovation.

[Table 3]

Theorem 2.2.[Hamadeh and Zakoïan (2011)] *Under assumptions QML0-QML6,*

$$\sqrt{n}\left(\hat{\theta}_n^{QML} - \theta_0\right) \Rightarrow \mathcal{N}(0, (\kappa_\eta - 1)J^{-1}),$$

where $J = E_{\theta_0}\left(\frac{\partial^2 \ell_t(\theta_0)}{\partial\theta\partial\theta'}\right) = 4E_{\theta_0}\left(\frac{1}{h_t^2(\theta_0)}\frac{\partial h_t(\theta_0)}{\partial\theta}\frac{\partial h_t(\theta_0)}{\partial\theta'}\right)$.

Proof. See Hamadeh and Zakoïan (2011), Proof of Theorem 2.2, pages 499-502. □

2.5 Local Asymptotic Normality and the Convolution Theorem

Let $\mu \in \mathbb{R}$, $\sigma > 0$, $\alpha_+ > 0$, $\alpha_- > 0$, $\beta > 0$ be parameters[4] and let $\{\eta_t : t \in \mathbb{Z}\}$ be an i.i.d. sequence of innovation errors with zero mean, unit variance and density f. Put $\xi_t = \mu + \sigma\eta_t$, thus ξ_t is a random variable with location μ, scale σ and density $\sigma^{-1}f(\{\xi_t - \mu\}/\sigma)$.

Consider the reparametrized PTTGARCH(1,1) model with observations

$$Y_t = \sqrt{h_t}\xi_t = \mu\sqrt{h_t} + \sigma\sqrt{h_t}\eta_t, \qquad \eta_t \sim iid(0,1), \tag{2.11}$$

[4] See Footnote 2.

where the unobservable heteroskedasticity factors $\{h_t\}_{t\in\mathbb{Z}}$ follow a PTTGARCH(1,1) process:

$$h_t = 1 + \beta h_{t-1} + \alpha_+ (Y_{t-1}^+)^2 + \alpha_- (Y_{t-1}^-)^2 = 1 + h_{t-1}\left[\beta + \alpha_+\left(\xi_{t-1}^+\right)^2 + \alpha_-\left(\xi_{t-1}^-\right)^2\right], \tag{2.12}$$

where the notation [see Hwang and Kim (2004), page 296]

$$a_t^+ = \max(a_t, 0) \text{ and } a_t^- = \max(-a_t, 0) \text{ is used so that } a_t = a_t^+ - a_t^-,$$

where $h_1 = \frac{1}{1-\frac{1}{2}(\alpha_+)-\frac{1}{2}(\alpha_-)-\beta}$, and where $\alpha_+ > 0$, $\alpha_- > 0$, $\beta > 0$, $\mu \in \mathbb{R}$ and $\sigma > 0$. Observe that the Euclidean parameter $\theta = (\alpha_+, \alpha_-, \beta, \mu, \sigma)' \in \Theta \subset \mathbb{R}^5$ is identifiable. The necessary and sufficient condition for a unique strictly stationary solution $\{h_t : t \in \mathbb{Z}\}$ is reported in Assumption 2.1.

In this section we show that the log-likelihood ratio of the PTTGARCH(1,1) model in (2.11)-(2.12) satisfies the *Local Asymptotic Normality* condition defined in Theorem 2.5.1 below. This condition was introduced by Le Cam (1960), and it requires that, in large samples, the log-likelihood ratio is approximately quadratic in a small neighborhood of the true parameter [Linton (1993)]. The LAN condition is necessary to establish the properties of the semiparametric estimator presented in Section 2.6; Fabian and Hannan (1982) show that if the log-likelihood ratio satisfies the LAN condition, then the *Local Asymptotic Minimax* bound is achieved by estimators equivalent to the MLE [Linton (1993)].

Le Cam (1960), Swensen (1985) and Roussas (1979) give conditions under which the log-likelihood ratio of a general stochastic process satisfies the LAN condition [Linton (1993)]; these conditions have been verified for stationary ARMA processes by Kreiss (1987), and DKW (1997) generalized the results of Kreiss (1987) to classes of non-linear time-series (location-scale) models [Wellner et al. (2006)]. In this paper, following DKW (1997), we show that the LAN property holds in the case of the reparametrized PTTGARCH(1,1) model as the reparametrized PTTGARCH(1,1) model is a non-linear time-series (location-scale) model that satisfies conditions in DKW (1997). Following DKW (1997), we also give the relevant assumptions such that semiparametric efficient adaptive estimation is possible for $\pi = (\alpha_+, \alpha_-, \beta)$.

In order to derive asymptotic results, we assume that the observed data Y_1,...,Y_n are stationary given the starting value $h_1(\theta) = h_{01}$ initializing equation (2.12); this is insured by Assumption 2.1. The parameter vector θ can be estimated based on h_{01},Y_1,...,Y_n, in the presence of the infinite-dimensional nuisance parameter f. We fix the nuisance parameter f in this section,[5] and we show that the resulting parametric model satisfies the LAN (Local Asymptotic Normality) property. Finally, we derive a bound on the asymptotic performance of regular estimators of θ (ie., the Convolution Theorem).[6]

[5]On the contrary, in section 2.6 the nuisance parameter $f(\eta_t)$ is unknown and is estimated using the kernel density estimator.

[6]See Newey (1990) for the definition of regular estimator.

Observe that if we also fix α_+, α_-, β, then our model is the location-scale model for i.i.d. random variables [DK (1997)]; thus, the location-scale model is a parametric submodel of our time-series model and it is reasonable to assume that this submodel is regular [DK (1997)],[7] i.e.:

Assumption 2.2. *f is an absolutely continuous Lebesgue density function with first order derivative f' and finite Fisher information for location*

$$I_l(f) = \int \left\{\frac{f'}{f}\right\}^2 f(\eta)d\eta < \infty, \tag{2.13}$$

and finite Fisher information for scale:

$$I_s(f) = \int \left\{1 + \eta\frac{f'}{f}\right\}^2 f(\eta)d\eta < \infty. \tag{2.14}$$

Furthermore, the random variable η has location zero and scale one.

In order to derive the local asymptotic normality results of an estimator, we need a sequence of non-stochastic values $\theta_n = (\alpha_{n+}, \ \alpha_{n-}, \beta_n, \mu_n, \sigma_n)'$ and $\tilde{\theta}_n = (\tilde{\alpha}_{n+}, \tilde{\alpha}_{n-}, \ \tilde{\beta}_n, \tilde{\mu}_n, \tilde{\sigma}_n)'$ near the true parameter vector $\theta_0 = (\alpha_0, \beta_0, \gamma_0, \mu_0, \sigma_0)'$ such that $|\theta_n - \theta_0|= \mathrm{O}(n^{-1/2})$, $|\tilde{\theta}_n - \theta_0|= \mathrm{O}(n^{-1/2})$, and even

$$\lambda_n = \sqrt{n}\left(\tilde{\theta}_n - \theta_n\right) \to \lambda, \text{ as } n \to \infty. \tag{2.15}$$

In the rest of this section expectations, convergences, etc. are implicitly taken under θ_n and f.

In order to obtain a uniform LAN theorem we consider the log-likelihood ratio Λ_n of h_{01}, Y_1,..., Y_n for $\tilde{\theta}_n$ with respect to θ_n under θ_n (and f fixed). To derive the joint density of Y_1,..., Y_n, using $h_1(\theta) = h_{01}$, we calculate the residuals and conditional variances recursively for $t = 1, 2, ...$

$$\xi_t(\theta) = \frac{Y_t}{\sqrt{h_t(\theta)}}, \tag{2.16}$$

$$\eta_t(\theta) = \frac{\xi_t(\theta) - \mu}{\sigma}, \tag{2.17}$$

$$h_{t+1}(\theta) = 1 + \beta h_t(\theta) + \alpha_+(Y_t^+)^2 + \alpha_-(Y_t^-)^2. \tag{2.18}$$

The density of Y_1,..., Y_n under θ_n, conditionally on h_{01}, is

$$\prod_{t=1}^n \sigma_n^{-1}h_{nt}^{-1/2}f(\sigma_n^{-1}\left\{h_{nt}^{-1/2}Y_t - \mu_n\right\}) = \prod_{t=1}^n \sigma_n^{-1}h_{nt}^{-1/2}f(\{\xi_{nt} - \mu_n\}/\sigma_n) = \prod_{t=1}^n \sigma_n^{-1}h_{nt}^{-1/2}f(\eta_{nt}), \tag{2.19}$$

where $h_{nt} = h_t(\theta_n)$, $\xi_{nt} = \xi_t(\theta_n)$ and $\eta_{nt} = \eta_t(\theta_n)$.

To emphasize the link between the present time-series model and the i.i.d. location-scale model we introduce

[7] See Newey (1990) for the definition of regular model.

the notation $\tilde{h}_{nt} = h_t(\tilde{\theta}_n)$,

$$l\{\mu, \sigma\}(x) = \log\, f(\{x-\mu\}/\sigma) - \log\, \sigma,$$

$$\begin{pmatrix} M_{nt} \\ S_{nt} \end{pmatrix} = n^{1/2}\sigma_n^{-1} h_{nt}^{-1/2} \begin{pmatrix} \tilde{\mu}_n \tilde{h}_{nt}^{1/2} - \mu_n h_{nt}^{1/2} \\ \tilde{\sigma}_n \tilde{h}_{nt}^{1/2} - \sigma_n h_{nt}^{1/2} \end{pmatrix}, \tag{2.20}$$

and $\tilde{\eta}_{nt} = \eta_t(\tilde{\theta}_n)$. Denoting the log-likelihood ratio for the initial value h_{01} as Λ_n^s, we can write the log-likelihood ratio Λ_n as

$$\Lambda_n = \log\, \left\{ \prod_{t=1}^{n} \tilde{\sigma}_n^{-1} \tilde{h}_{nt}^{-1/2} f(\tilde{\eta}_{nt}) / \prod_{t=1}^{n} \sigma_n^{-1} h_{nt}^{-1/2} f(\eta_{nt}) \right\} + \Lambda_n^s = ...$$

$$= \sum_{t=1}^{n} \{\ell\{(\mu_n, \sigma_n) + \sigma_n n^{-1/2}(M_{nt}, S_{nt})\}(\xi_{nt}) - \ell\{\mu_n, \sigma_n\}(\xi_{nt})\} + \Lambda_n^s = ...$$

$$= \sum_{t=1}^{n} \{\ell\{(0,1) + n^{-1/2}(M_{nt}, S_{nt})\}(\eta_{nt}) - \ell\{0,1\}(\eta_{nt})\} + \Lambda_n^s. \tag{2.21}$$

Equation (2.21) is similar to the log-likelihood ratio statistic for the i.i.d. location-scale model but the deviations M_{nt} and S_{nt} are random in equation (2.21); in the following discussion we will apply the results of DKW (1997) which allow for such random sequences.

To eliminate asymptotically the dependence of the log-likelihood ratio statistic on the initial value h_{01}, we will use the following regularity condition [see Assumption A in DKW (1997)].

Assumption 2.3. *The density $\overline{f}_\theta$ of the initial value h_{01} satisfies, under θ_n,*

$$\Lambda_n^s = \log\{\overline{f}_{\tilde{\theta}_n}(h_{01}) / \overline{f}_{\theta_n}(h_{01})\} \xrightarrow{P} 0 \text{ as } n \to \infty. \tag{2.22}$$

In order to make an expansion of Λ_n, we introduce the notation $\dot{\ell}_{nt}$ for the five-dimensional conditional score at time t. Denote the three-dimensional vector derivative of the conditional variance by

$$H_t(\theta) = \frac{\partial h_t(\theta)}{\partial\, (\alpha_+, \alpha_-, \beta)} = \beta H_{t-1}(\theta) + \begin{pmatrix} (Y_{t-1}^+)^2 \\ (Y_{t-1}^-)^2 \\ h_{t-1}(\theta) \end{pmatrix}, \tag{2.23}$$

with $H_1(\theta) = 0_3$. Define the (5×2)-derivative matrix $W_t(\theta)$ -which is motivated by differentiation of (M_{nt}, S_{nt}) with respect to $\tilde{\theta}_n$ at θ_n- by

$$W_t(\theta) = \sigma^{-1} \begin{pmatrix} \frac{1}{2} h_t^{-1}(\theta) H_t(\theta)(\mu, \sigma) \\ I_2 \end{pmatrix}, \tag{2.24}$$

where I_2 is the (2×2) identity matrix. Denote the location-scale score by

$$\psi_t(\theta) = -\begin{pmatrix} \frac{f'(\eta_t(\theta))}{f(\eta_t(\theta))} \\ 1 + \eta_t(\theta)\frac{f'(\eta_t(\theta))}{f(\eta_t(\theta))} \end{pmatrix} = -\begin{pmatrix} \ell'(\eta_t(\theta)) \\ 1 + \eta_t(\theta)\ell'(\eta_t(\theta)) \end{pmatrix}, \tag{2.25}$$

where we denote $\ell' = \frac{f'}{f}$, and put

$$\dot{\ell}_t(\theta) = W_t(\theta)\psi_t(\theta).$$

Using the previous notation, the conditional score at time t can be denoted by $\dot{\ell}_{nt} = \dot{\ell}_t(\theta_n)$. An expansion of (2.21) shows that the log-likelihood ratio Λ_n can be written as

$$\Lambda_n = \lambda' n^{-1/2} \sum_{t=1}^{n} \dot{\ell}_{nt} - \frac{1}{2} n^{-1} \sum_{t=1}^{n} \{\lambda' \dot{\ell}_{nt}\}^2 + R_n. \tag{2.26}$$

In Appendix B we show that the following LAN theorem [Theorem 2.1 of DKW (1997)] holds for the parametric version of model (2.11),(2.12).

Theorem 2.5.1. ***LAN***

Suppose that Assumptions 2.1, 2.2, 2.3 are satisfied. Then the local log-likelihood ratio statistic Λ_n, as defined by (2.21) *and by* (2.26) *is asymptotically normal. More precisely, under θ_n,*

$$R_n \xrightarrow{P} 0,\ \Lambda_n \xrightarrow{D} N(-\frac{1}{2}\lambda' I(\theta_0)\lambda, \lambda' I(\theta_0)\lambda) \quad \text{as } n \to \infty, \tag{2.27}$$

where $I(\theta_0)$ is the probability limit of the averaged score products $\dot{\ell}_{nt}\dot{\ell}'_{nt}$.

Proof. See Appendix B. □

This LAN theorem implies that the sequences of probability measures $\{P_{\theta_n,f}\}$ and $\{P_{\tilde{\theta}_n,f}\}$ are contiguous. We can now apply the Convolution Theorem of Hájek (1970) [see Theorem 2.3.1 of BKRW (1993)].

Theorem 2.5.2. ***Convolution Theorem***

Under the assumptions of the LAN Theorem 2.5.1, let $\{T_n : n \in \mathbb{N}\}$ be a regular sequence of estimators of $q(\theta)$, where $q : \mathbb{R}^5 \to \mathbb{R}^k$ is differentiable with total differential matrix $\dot{q}$. As usual, regularity at $\theta = \theta_0$ means that there exists a random k-vector Z such that for all sequences $\{\theta_n : n \in \mathbb{N}\}$, with $n^{1/2}(\theta_n - \theta_0) = \mathrm{O}(1)$,

$$n^{1/2}\{T_n - q(\theta_n)\} \xrightarrow{D} Z \quad \text{as } n \to \infty, \tag{2.28}$$

where the convergence is under θ_n. Let $\tilde{\ell} = \dot{q}(\theta_0)I(\theta_0)^{-1}\dot{\ell}(\theta_0)$ be the efficient influence function, then, under θ_0,

$$\begin{pmatrix} n^{1/2}\{T_n - q(\theta_0) - n^{-1}\sum_{t=1}^{n}\tilde{\ell}_t\} \\ n^{-1/2}\sum_{t=1}^{n}\tilde{\ell}_t \end{pmatrix} \xrightarrow{D} \begin{pmatrix} \Delta_0 \\ Z_0 \end{pmatrix} \quad \text{as } n \to \infty, \tag{2.29}$$

where Δ_0 and Z_0 are independent and Z_0 is $N(0, \dot{q}(\theta_0)I(\theta_0)^{-1}\dot{q}(\theta_0)')$. Furthermore, $\{T_n : n \in \mathbb{N}\}$ is efficient if $\{T_n : n \in \mathbb{N}\}$ is asymptotically linear in the efficient influence function, i.e. if $\Delta_0 = 0$ (a.s.).

From the Convolution Theorem we obtain that a regular estimator $\hat{\theta}_n$ of θ satisfies, under θ_0, $\sqrt{n}(\hat{\theta}_n - \theta_0) \xrightarrow{D} \Delta_0 + Z_0$, that is, the limit distribution of $\hat{\theta}_n$ is the convolution of the random vector Δ_0 and a Gaussian random vector Z_0 with mean zero and variance equal to the inverse of the information matrix $I(\theta_0)$ (ie., the Cramer Rao lower bound).

2.6 The Semiparametric Efficient Adaptive Estimator

Our approach follows DK (1997), who derive a semiparametric efficient adaptive estimator for the GARCH(1,1) model while we derive a semiparametric efficient adaptive estimator for the (reparametrized) PTTGARCH(1,1) model.

In semiparametric models there exist methods to upgrade $\sqrt{n}$-consistent estimators to efficient ones using the Newton-Raphson technique as long as it is possible to accurately estimate the relevant score or influence functions. In Schick (1986) such a method based on "sample splitting" and Le Cam's "discretization" is described for i.i.d. models [DK (1997)]. DKW (1997) adapted Schick (1986)'s method to the case of time-series models [see Theorem 3.1 of DKW (1997)]. Following DK (1997), we apply the discretization of θ_n and the sample splitting method in our proofs.[8] If a $\sqrt{n}$-consistent estimator is available, it is trivial to derive the discretized local parameters near θ_0 since for any subsequence of $\{\hat{\theta}_n\}$, there are $\{\hat{\theta}_{n_k}\}$ satisfying the condition (2.15) [Sun and Stengos (2006)]. We assume the existence of a preliminary, $\sqrt{n}$-consistent estimator.

Assumption 2.4. *There exists a $\sqrt{n}$-consistent estimator $\hat{\theta}_n$ of θ_n (under θ_n and f).*

For our (reparametrized) PTTGARCH(1,1) model a natural candidate for such an initial estimator is the QML estimator; see Section 2.4 for the $\sqrt{n}$-consistency and asymptotic normality of the QML estimator in the case of the PTTGARCH(1,1) model.

We focus on efficient and adaptive estimation of the parameters α_+, α_-, β.[9] Theorem 3.1 of DKW (1997) contains the assumptions that are needed to obtain adaptive estimation in time-series models. In Appendix C we verify the conditions of Theorem 3.1 of DKW (1997), this yields the following theorem.

[8] The sample splitting method is introduced as a technical tool that simplifies the proofs.

[9] Alternatively, in view of eq. (2.24), note that the score $\dot{\ell}_{nt}$ satisfies the form discussed in Example 3.1 of DKW (1997).

Theorem 2.6.1. *Under Assumptions 2.1, 2.2, 2.3, 2.4 adaptive estimators of α_+, α_-, β do exist.*

Proof. See Appendix C. □

Let $\hat{\theta}_n = \left(\hat{\alpha}_{n+}, \hat{\alpha}_{n-}, \hat{\beta}_n, \hat{\mu}_n, \hat{\sigma}_n\right)'$ be the $\sqrt{n}$-consistent estimator of θ and compute $W_t(\hat{\theta}_n)$ via (2.23) and (2.24). Let $\hat{\eta}_{n1}, ..., \hat{\eta}_{nn}$ be the residuals computed from $h_1, Y_1, ..., Y_n$ and $\hat{\theta}_n$ using (2.17). Based on the estimated residuals $\hat{\eta}_{n1}, ..., \hat{\eta}_{nn}$, we estimate nonparametrically $f(\cdot)$ using the kernel density estimator with the standardized normal kernel $K(\cdot)$ and bandwidth b_n:[10]

$$\hat{f}_n(\cdot) = \frac{1}{n}\sum_{t=1}^{n} \frac{1}{b_n} K\left(\frac{\cdot - \hat{\eta}_{nt}}{b_n}\right),$$

and subsequently we estimate $\psi(\cdot)$ by $\hat{\psi}_n(\cdot)$;[11] here $b_n \to 0$ and $nb_n^4 \to \infty$. Our semiparametric estimator can be written as

$$\left(\hat{\alpha}_{n+}, \hat{\alpha}_{n-}, \hat{\beta}_n\right)' + (I_3, 0_{2\times 2}) \left(\frac{1}{n}\sum_{t=1}^{n} W_t(\hat{\theta}_n)\hat{\psi}_n(\hat{\eta}_{nt})\hat{\psi}_n(\hat{\eta}_{nt})' W_t(\hat{\theta}_n)'\right)^{-1} \times ...$$

$$\times \frac{1}{n}\sum_{t=1}^{n}\left\{W_t(\hat{\theta}_n) - \frac{1}{n}\sum_{s=1}^{n} W_s(\hat{\theta}_n)\right\}\hat{\psi}_n(\hat{\eta}_{nt}), \tag{2.30}$$

where $\hat{\theta}_n$ is the QML estimator. This is the semiparametric estimator used in the simulations of Section 4. We need the following two technical modifications in order to prove that the semiparametric estimator is adaptive.

Discretization

$\hat{\theta}_n$ is discretized by changing its value in $(0,\infty)\times$ $(0,\infty)\times$ $(0,\infty)\times$ $\mathbb{R}\times$ $(0,\infty)$ into one of the nearest points in the grid $\frac{c}{\sqrt{n}}(\mathbb{N}\times$ $\mathbb{N}\times$ $\mathbb{N}\times$ $\mathbb{Z}\times$ $\mathbb{N})$. This method gives the possibility to consider $\hat{\theta}_n$ to be non-random, and therefore independent of $\hat{\eta}_{nt}$, Y_t and h_1 [DK (1997)].

Sample splitting

The set of residuals $\hat{\eta}_{n1}, ..., \hat{\eta}_{nn}$ is split into two samples, which may be viewed as independent [DK (1997)]. For $\hat{\eta}_{nt}$ in the first sample, the second sample is used to estimate $\psi(\cdot)$ by $\hat{\psi}_{n2}(\cdot)$, and $\hat{\psi}_n(\hat{\eta}_{nt})$ in eq. (2.30) is replaced by $\hat{\psi}_{n2}(\hat{\eta}_{nt})$; for $\hat{\eta}_{nt}$ in the second sample, the first sample is used to estimate $\psi(\cdot)$ by $\hat{\psi}_{n1}(\cdot)$, and $\hat{\psi}_n(\hat{\eta}_{nt})$ in eq. (2.30) is replaced by $\hat{\psi}_{n1}(\hat{\eta}_{nt})$.

[10]The kernel function $K(\cdot)$ used in our study is the standardized normal kernel: $K\left(\frac{\cdot - \hat{\eta}_{nt}}{b_n}\right) = \frac{1}{\sqrt{2\pi}} \exp\left(-0.5\left(\frac{\cdot - \hat{\eta}_{nt}}{b_n}\right)^2\right)$.

[11]The kernel density derivative estimator can be written as [Wand and Jones (1995), page 49]: $\hat{f'}_n(\cdot) = \frac{1}{n}\sum_{t=1}^{n} \frac{1}{b_n^2} K'\left(\frac{\cdot - \hat{\eta}_{nt}}{b_n}\right)$, where $K'(\cdot)$ is the derivative of the standard normal density function. Finally, the location-scale score can be estimated using $\hat{\psi}_n(\cdot) = -\begin{bmatrix} \frac{\hat{f}'_n(\cdot)}{\hat{f}_n(\cdot)} \\ 1 + \hat{\eta}_t \frac{\hat{f}'_n(\cdot)}{\hat{f}_n(\cdot)}. \end{bmatrix}$, where $\hat{\eta}_t$ is the vector containing the estimated residuals $(\hat{\eta}_{n1}, ..., \hat{\eta}_{nn})$.

3 The Theoretical Asymptotic Variance of the ML and the QML Estimators

3.1 The Theoretical Asymptotic Variance of the ML Estimator

In this section $W_t(\cdot)$, $h_t(\cdot)$, $H_t(\cdot)$, $\psi_t(\cdot)$ are evaluated at the true parameters (θ_0). In the Monte Carlo experiments for the PTTGARCH(1,1) model, the parameter μ is set to zero and is not estimated [see Section 4], thus eq. (2.25) becomes

$$\psi_t(\theta_0) = -\left(1 + \eta_t(\theta_0)\frac{f'(\eta_t(\theta_0))}{f(\eta_t(\theta_0))}\right), \tag{3.1}$$

and eq. (2.24) becomes

$$W_t(\theta_0) = \begin{pmatrix} \frac{1}{2}h_t^{-1}(\theta_0)H_t(\theta_0) \\ \sigma_0^{-1} \end{pmatrix} = \begin{pmatrix} \frac{1}{2}h_t^{-1}(\theta_0)\frac{\partial h_t(\theta_0)}{\partial \alpha_+} \\ \frac{1}{2}h_t^{-1}(\theta_0)\frac{\partial h_t(\theta_0)}{\partial \alpha_-} \\ \frac{1}{2}h_t^{-1}(\theta_0)\frac{\partial h_t(\theta_0)}{\partial \beta} \\ \sigma_0^{-1} \end{pmatrix}, \tag{3.2}$$

where the three-dimensional vector of derivatives $H_t(\theta_0)$ -reported in eq. (2.23)- is evaluated at the true parameters (θ_0). The analytical solution for $\psi_t(\theta_0)$ for each type of innovation is reported in Appendix A. The Fisher information matrix can be written as

$$I(\theta_0) = I_s(f)E(W_t(\theta_0)W_t(\theta_0)'), \tag{3.3}$$

with $I_s(f)$ denoting the Fisher information for scale described by

$$I_s(f) = E\left[\psi_t(\theta_0)\psi_t(\theta_0)\right] = E\left[1 + \eta_t(\theta_0)\frac{f'(\eta_t(\theta_0))}{f(\eta_t(\theta_0))}\right]^2 = \int \left[1 + \eta_t(\theta_0)\frac{f'(\eta_t(\theta_0))}{f(\eta_t(\theta_0))}\right]^2 f(\eta_t(\theta_0))d\eta_t.$$

The Fisher information for scale is computed by numerical integration for each type of innovation and is reported in Table 4.

[Table 4]

The Fisher information matrix reported in eq. (3.3) can be written as

$$I(\theta_0) = I_s(f) \times ...$$

$$\times \left[E_{\theta_0} \begin{pmatrix} \frac{1}{4h_t(\theta_0)^2}\frac{\partial h_t(\theta_0)}{\partial\alpha_+}\frac{\partial h_t(\theta_0)}{\partial\alpha_+} & \frac{1}{4h_t(\theta_0)^2}\frac{\partial h_t(\theta_0)}{\partial\alpha_+}\frac{\partial h_t(\theta_0)}{\partial\alpha_-} & \frac{1}{4h_t(\theta_0)^2}\frac{\partial h_t(\theta_0)}{\partial\alpha_+}\frac{\partial h_t(\theta_0)}{\partial\beta} & \frac{0.5}{h_t(\theta_0)}\frac{\partial h_t(\theta_0)}{\partial\alpha_+} \\ \frac{1}{4h_t(\theta_0)^2}\frac{\partial h_t(\theta_0)}{\partial\alpha_-}\frac{\partial h_t(\theta_0)}{\partial\alpha_+} & \frac{1}{4h_t(\theta_0)^2}\frac{\partial h_t(\theta_0)}{\partial\alpha_-}\frac{\partial h_t(\theta_0)}{\partial\alpha_-} & \frac{1}{4h_t(\theta_0)^2}\frac{\partial h_t(\theta_0)}{\partial\alpha_-}\frac{\partial h_t(\theta_0)}{\partial\beta} & \frac{0.5}{h_t(\theta_0)}\frac{\partial h_t(\theta_0)}{\partial\alpha_-} \\ \frac{1}{4h_t(\theta_0)^2}\frac{\partial h_t(\theta_0)}{\partial\beta}\frac{\partial h_t(\theta_0)}{\partial\alpha_+} & \frac{1}{4h_t(\theta_0)^2}\frac{\partial h_t(\theta_0)}{\partial\beta}\frac{\partial h_t(\theta_0)}{\partial\alpha_-} & \frac{1}{4h_t(\theta_0)^2}\frac{\partial h_t(\theta_0)}{\partial\beta}\frac{\partial h_t(\theta_0)}{\partial\beta} & \frac{0.5}{h_t(\theta_0)}\frac{\partial h_t(\theta_0)}{\partial\beta} \\ \frac{0.5}{h_t(\theta_0)}\frac{\partial h_t(\theta_0)}{\partial\alpha_+} & \frac{0.5}{h_t(\theta_0)}\frac{\partial h_t(\theta_0)}{\partial\alpha_-} & \frac{0.5}{h_t(\theta_0)}\frac{\partial h_t(\theta_0)}{\partial\beta} & 1 \end{pmatrix} \right].$$

The (4x4)-dimensional matrix $E(W_t(\theta_0)W_t(\theta_0)')$ -reported in eq. (3.3)- is computed through simulation. The theoretical asymptotic variance matrix of the MLE can be written as

$$V_0 = \frac{[I(\theta_0)]^{-1}}{N}, \tag{3.4}$$

where N is the dimension of the sample (i.e. N=2000). In order to estimate the asymptotic variance of the MLE, the number of time periods used is N=2000 and 5000 replications are used.[12] The theoretical standard deviations $\hat{\sigma}_{\alpha_+}$, $\hat{\sigma}_{\alpha_-}$ and $\hat{\sigma}_{\beta}$ are obtained by taking the square root of (the diagonal elements of) V_0 [see eq. (3.4)]. The theoretical asymptotic standard deviations $\hat{\sigma}_{\alpha_+}$, $\hat{\sigma}_{\alpha_-}$ and $\hat{\sigma}_{\beta}$ are reported in Table 5 for each combination of $(\alpha_+, \alpha_-, \beta)$ [see Section 4].

3.2 The Theoretical Asymptotic Variance of the QML Estimator

All vectors and matrices used in this section are evaluated at the true parameters (θ_0). In the Monte Carlo experiments for the PTTGARCH(1,1) model, the parameter μ is set to zero and is not estimated, thus the score becomes a (4×1) vector $\dot{\ell}_t(\theta_0) = W_t(\theta_0)\psi_t(\theta_0)$, with $\psi_t(\theta_0)$ reported in eq. (3.1) and $W_t(\theta_0)$ reported in eq. (3.2).[13] The theoretical asymptotic variance of the QML estimator can be written as

$$V_0 = \frac{J(\theta_0)^{-1} I(\theta_0) J(\theta_0)^{-1}}{N},$$

where $I(\theta_0)$ is the Fisher information matrix evaluated at θ_0 and $J(\theta_0)$ is the negative of the expectation of the Hessian evaluated at θ_0:

$$I(\theta_0) = E_{\theta_0}\left(\frac{\partial l_t(\theta_0)}{\partial\theta}\frac{\partial l_t(\theta_0)}{\partial\theta}\right),$$

$$J(\theta_0) = -E_{\theta_0}\left(\frac{\partial^2 l_t(\theta_0)}{\partial\theta\partial\theta}\right).$$

[12] Firstly, the asymptotic variance of the MLE was estimated with 2500 replications and with 5000 replications; the two estimations gave slightly different results, thus the estimation of the asymptotic variance of the MLE with 5000 replications was adopted in order to obtain more precise estimates. Second, we obtained approximately the same (estimated) asymptotic variance of the MLE using a higher number of replications than 5000 replications, thus the choice of 5000 replications for the estimation of the asymptotic variance of the MLE was retained.

[13] $\dot{\ell}_t(\theta_0)$ is equivalent to $\frac{\partial l_t(\theta_0)}{\partial\theta}$.

The theoretical asymptotic variance matrix of the QMLE can be written as

$$V_0 = \frac{J(\theta_0)^{-1} I(\theta_0) J(\theta_0)^{-1}}{N} = \frac{1}{N}\frac{(\kappa_\eta - 1)}{2} \times ...$$

$$\times \left[E_{\theta_0} \begin{pmatrix} \frac{1}{2h_t(\theta_0)^2}\frac{\partial h_t(\theta_0)}{\partial \alpha_+}\frac{\partial h_t(\theta_0)}{\partial \alpha_+} & \frac{1}{2h_t(\theta_0)^2}\frac{\partial h_t(\theta_0)}{\partial \alpha_+}\frac{\partial h_t(\theta_0)}{\partial \alpha_-} & \frac{1}{2h_t(\theta_0)^2}\frac{\partial h_t(\theta_0)}{\partial \alpha_+}\frac{\partial h_t(\theta_0)}{\partial \beta} & \frac{1}{h_t(\theta_0)}\frac{\partial h_t(\theta_0)}{\partial \alpha_+} \\ \frac{1}{2h_t(\theta_0)^2}\frac{\partial h_t(\theta_0)}{\partial \alpha_-}\frac{\partial h_t(\theta_0)}{\partial \alpha_+} & \frac{1}{2h_t(\theta_0)^2}\frac{\partial h_t(\theta_0)}{\partial \alpha_-}\frac{\partial h_t(\theta_0)}{\partial \alpha_-} & \frac{1}{2h_t(\theta_0)^2}\frac{\partial h_t(\theta_0)}{\partial \alpha_-}\frac{\partial h_t(\theta_0)}{\partial \beta} & \frac{1}{h_t(\theta_0)}\frac{\partial h_t(\theta_0)}{\partial \alpha_-} \\ \frac{1}{2h_t(\theta_0)^2}\frac{\partial h_t(\theta_0)}{\partial \beta}\frac{\partial h_t(\theta_0)}{\partial \alpha_+} & \frac{1}{2h_t(\theta_0)^2}\frac{\partial h_t(\theta_0)}{\partial \beta}\frac{\partial h_t(\theta_0)}{\partial \alpha_-} & \frac{1}{2h_t(\theta_0)^2}\frac{\partial h_t(\theta_0)}{\partial \beta}\frac{\partial h_t(\theta_0)}{\partial \beta} & \frac{1}{h_t(\theta_0)}\frac{\partial h_t(\theta_0)}{\partial \beta} \\ \frac{1}{h_t(\theta_0)}\frac{\partial h_t(\theta_0)}{\partial \alpha_+} & \frac{1}{h_t(\theta_0)}\frac{\partial h_t(\theta_0)}{\partial \alpha_-} & \frac{1}{h_t(\theta_0)}\frac{\partial h_t(\theta_0)}{\partial \beta} & 2 \end{pmatrix} \right]^{-1}, \tag{3.5}$$

where N is the dimension of the sample (i.e., $N = 2000$) and where κ_η is the innovation's kurtosis. The kurtosis of each type of innovation is reported in Table 3. We compute the asymptotic variance of the QMLE through simulation using 5000 replications.[14] Taking the square root of (the diagonal elements of) V_0 [see eq. (3.5)], the theoretical standard deviations $\hat{\sigma}_{\alpha_+}$, $\hat{\sigma}_{\alpha_-}$ and $\hat{\sigma}_\beta$ are obtained. The theoretical asymptotic standard deviations $\hat{\sigma}_{\alpha_+}$, $\hat{\sigma}_{\alpha_-}$ and $\hat{\sigma}_\beta$ are reported in Table 5 for each combination of $(\alpha_+, \alpha_-, \beta)$ [see Section 4].

4 Monte Carlo Experiments

To enhance the interpretation and validity of the theoretical results of the previous sections we present Monte Carlo experiments in this section. The main purpose of Monte Carlo experiments is to evaluate the moderate sample properties of the QMLE, MLE and the semiparametric estimator. The sample size for Monte Carlo experiments is $N = 2000$. This choice is made since in the case of the PTTGARCH(1,1) model, two parameters (i.e., α_+, α_-) are estimated using approximately half of the sample size [see Section 2 for more details].[15]

The number of replications used in each Monte Carlo experiment is equal to 2500 as in DK (1997) and in recent studies [e.g., Pakel et al. (2011)]. Diaz-Emparanza (2002) studies the effect of the number of replications on the accuracy of Monte Carlo experiments; the author shows that 1000 (or less than 1000) replications yield a low level of accuracy of Monte Carlo experiments as the empirical distribution of estimated parameters obtained through Monte Carlo experiments is quite different from the theoretical distribution. Since we choose 2500 replications for our Monte Carlo experiments, the problems outlined in Diaz-Emparanza (2002) are unlikely to apply in our context.

[14] Firstly, the asymptotic variance of the QMLE was estimated with 2500 replications and with 5000 replications; the two estimations gave slightly different results, thus the estimation of the asymptotic variance of the QMLE with 5000 replications was adopted in order to obtain more precise estimates. Second, we obtained approximately the same (estimated) asymptotic variance of the QMLE using a higher number of replications than 5000 replications, thus the choice of 5000 replications for the estimation of the asymptotic variance of the QMLE was retained.

[15] Additionally, semiparametric estimators tend to yield strong efficiency improvements with respect to the QMLE in the case of moderate or large samples [see Section 4 of Sun and Stengos (2006) for an example]. This is due to the fact that the small sample properties of semiparametric estimators are worse than the large sample properties of semiparametric estimators due to inherent problems of choosing the bandwidth [see page 207-208 of DK (1997)].

Monte Carlo experiments are conducted using the data generating process described in Section 2.1. The conditional mean parameter μ is set to zero and is not included in the set of estimated parameters since our main interest is in the estimation of the conditional variance parameters $(\alpha_+, \alpha_-, \beta)$. The choice of setting $\mu = 0$ is not uncommon in studies that construct semiparametric estimators for conditional heteroskedasticity models; for example, DK (1997) set the conditional mean parameter equal to zero and Sun and Stengos (2006) set the conditional mean parameter to a value that is very close to zero. Three combinations for the set of parameters $(\alpha_+, \alpha_-, \beta, \sigma)$ are used in Monte Carlo experiments: $(\alpha_+, \alpha_-, \beta, \sigma) = (0.20, 0.40, 0.60, 1)$, $(\alpha_+, \alpha_-, \beta, \sigma) = (0.075, 0.15, 0.80, 1)$, $(\alpha_+, \alpha_-, \beta, \sigma) = (0.08, 0.11, 0.87, 1)$; we choose these combinations of parameters $(\alpha_+, \alpha_-, \beta, \sigma)$ to check the performance of the semiparametric estimator as persistence of the PTTGARCH process increases.[16]

The sample means and sample standard deviations of the estimates obtained from Monte Carlo experiments using the MLE, QMLE and the semiparametric estimator are reported in Table 5. The sample means and sample standard deviations of the estimates using the MLE are reported in order to confront the efficiency of the semiparametric estimator with the MLE, even though the consistency and the asymptotic normality of the MLE was not proved for the PTTGARCH(1,1) model. Additionally, the theoretical asymptotic standard deviations $\hat{\sigma}_{\alpha+}$, $\hat{\sigma}_{\alpha-}$, $\hat{\sigma}_{\beta}$, obtained with the estimation of the theoretical asymptotic variance of the MLE and QMLE [see Section 3], are added to Table 5 in order to check whether the asymptotic results of the MLE and QMLE from Monte Carlo experiments approximate well enough the theoretical asymptotic behavior of the MLE and QMLE. In the semiparametric part we used the standardized normal kernel with a bandwidth of $h = 0.40$; reasonable changes of the bandwidth, say $0.30 \leq h \leq 0.50$ do not alter the conclusions below.

This section is organized in the following way: **(1)** in order to check whether the asymptotic results of the Monte Carlo experiments approximate the theoretical asymptotic behavior of the MLE and QMLE, in Section 4.1 the theoretical asymptotic standard deviations $\hat{\sigma}_{\alpha+}$, $\hat{\sigma}_{\alpha-}$, $\hat{\sigma}_{\beta}$ of the MLE and QMLE are compared with the standard deviations of the MLE and QMLE obtained from Monte Carlo experiments. In order to have an easier comparison between theoretical asymptotic results and Monte Carlo results, the asymptotic relative efficiencies (AREs) of the MLE and QMLE are compared with the relative efficiencies (REs) of the MLE and QMLE in the same section. **(2)** The main results of this section are reported in Section 4.2. In this section, the performance of the MLE, QMLE and the semiparametric estimator is evaluated through the relative efficiencies (REs) of the estimators. **(3)** Some considerations and remarks are reported in Section 4.3.

[Table 5]

[16]Furthermore, we choose these combinations of parameters $(\alpha_+, \alpha_-, \beta, \sigma)$ to insure that the PTTGARCH process does not incorporate unstationary behavior.

4.1 Comparison of the Standard Deviations from the Monte Carlo Experiments with the Theoretical Asymptotic Standard Deviations

The standard deviations of the MLE $\hat{\sigma}_{\alpha_+}$, $\hat{\sigma}_{\alpha_-}$ and $\hat{\sigma}_{\beta}$ obtained from Monte Carlo experiments are close to the theoretical asymptotic standard deviations of the MLE [Table 5]. Also, the standard deviations of the QMLE $\hat{\sigma}_{\alpha_+}$, $\hat{\sigma}_{\alpha_-}$ and $\hat{\sigma}_{\beta}$ obtained from Monte Carlo experiments are close to the theoretical asymptotic standard deviations of the QMLE [see Table 5]. As a consequence, the asymptotic relative efficiencies (AREs) of the QMLE versus the MLE are close to the relative efficiencies (REs) of the QMLE versus the MLE [compare Table 6 with Table 7]. To conclude, the asymptotic results of the MLE and QMLE from Monte Carlo experiments seem to approximate well enough the theoretical asymptotic behavior of the MLE and QMLE.

[Table 6]

[Table 7]

4.2 The Relative Efficiencies of the Estimators

The performance of the three estimators used in Monte Carlo experiments is evaluated through the relative efficiencies (REs) of the estimators which are reported in Table 7.

In the case of the $\mathcal{N}(0,1)$ *innovation*, we only compare the semiparametric estimator with the MLE since the QMLE is equivalent to the MLE in this case. The semiparametric estimator is almost as efficient as the MLE.

For the *Laplace(0,1) innovation*, the semiparametric estimator performs significantly better than the QMLE [Table 7, row 6]; in addition, the performance gap between the semiparametric estimator and the MLE is small [Table 7, row 7]. Therefore, the semiparametric estimator yields a relevant efficiency gain with respect to the QMLE and is close to the maximum efficiency.

In the case of the *Gaussian Mixture innovation*, the performance of the semiparametric estimator is worse than the MLE [Table 7, row 10] and is approximately the same as the QMLE [Table 7, row 9].

For the t_5*-student innovation*, the performance of the semiparametric estimator is significantly higher than the QMLE [Table 7, row 12], especially in the case that the PTTGARCH process is characterized by high persistence (i.e., $(\alpha_+, \alpha_-, \beta, \sigma) = (0.08, 0.11, 0.87, 1)$). For the t_7*-student* and t_9*-student innovations*, the semiparametric estimator has significantly higher efficiency than the QMLE [see Table 7, row 15 for the t_7-student innovation and see Table 7, row 18 for the t_9-student innovation, respectively]. In the case of the t_7-student and t_9-student innovations, the efficiency of the semiparametric estimator is very close to the maximum efficiency that an estimator can attain (i.e., the efficiency of the MLE) [see Table 7, row 16 for the t_7-student innovation and Table 7, row 19 for the t_9-student innovation].

For the χ^2_6 *innovation*, the performance gap between the QMLE and the MLE is very large [Table 7, row 20] since in the case of the QMLE we assume that the innovation is distributed as a normal distribution with zero mean and unit variance while the real innovation is fat-tailed and skewed. The performance of the semiparametric estimator is much higher than the performance of the QMLE [Table 7, row 21], that is, the semiparametric estimator regains most of the (efficiency) loss caused by the inefficient QML estimator. Similar results are obtained in the case of the χ^2_{12} *innovation*: the QMLE has much lower efficiency than the MLE [Table 7, row 23], and the semiparametric estimator regains most of the efficiency loss caused by the QML estimator.

Finally, the semiparametric estimator recaptures most of the efficiency loss of the QMLE in the interesting case of high persistence (i.e., $(\alpha_+, \alpha_-, \beta, \sigma) = (0.08, 0.11, 0.87, 1)$), and the performance of the semiparametric estimator versus the QMLE (or MLE) seems not to be affected by the degree of persistence of the PTTGARCH process [compare the REs (relative efficiencies) presented in Column (7)-(9) of Table 7 with the REs presented in Column (1)-(3) of Table 7].

4.3 Concluding Remarks

In this section we have shown that Monte Carlo experiments approximate well enough the theoretical asymptotic behavior of the MLE and QMLE, that is, Monte Carlo experiments are appropriately set up. The main result of this section is that the use of the semiparametric estimator yields a large efficiency gain with respect to the QMLE for all innovations, except in the case of the Gaussian Mixture innovation;[17] in the case of the Gaussian Mixture innovation, and depending on the combination of $(\sigma_{\alpha_+}, \sigma_{\alpha_-}, \sigma_\beta)$, the semiparametric estimator has slightly inferior performance or approximately the same performance than the QMLE. Anyhow, this anomaly is also obtained in other studies on semiparametric estimators [e.g., DK (1997), Table 1]. In the case of the Laplace(0,1), t-student and χ^2 distributions, the semiparametric estimator recaptures most of the efficiency loss caused by the (inefficient) QML estimator. In empirical datasets, one often observes non-normal distributions for the innovation [see Table 2 of DK (1997)],[18] thus it seems worthwhile to apply semiparametric techniques in the case that the true innovation is unknown. Finally, the performance of the semiparametric estimator seems not be affected by the persistence of the PTTGARCH process.

[17] Also in the case of the $\mathcal{N}(0,1)$ innovation we have found that the semiparametric estimator is not more efficient than the QML estimator; this is not a surprising result since the QML estimator is the most efficient estimator in the case of the $\mathcal{N}(0,1)$ innovation.

[18] DK (1997) obtain significantly lower standard deviations for conditional variance parameters $\hat{\sigma}_\alpha$ and $\hat{\sigma}_\beta$ in the case of the semiparametric estimator than in the case of the QML estimator; from this evidence we can infer that the true innovation is not distributed as a Normal(0,1) random variable.

5 Conclusions

This paper derives a semiparametric efficient adaptive estimator of the PTTGARCH(1,1) model without imposing any additional restrictions on the innovation distribution other than some regularity conditions. We provide a series of Monte Carlo simulations to evaluate the moderate sample properties of the semiparametric estimator. Monte Carlo simulations show that the semiparametric estimator regains most of the efficiency loss of the (inefficient) QML estimator, especially in the case of fat-tailed (and skewed) distributions which are often observed in empirical datasets.

A Analytical Solution for the Location-Scale Score

In this section we report the analytical solution for $\psi_t(\theta_0) = -\left(1 + \eta_t(\theta_0)\ \frac{f'(\eta_t(\theta_0))}{f(\eta_t(\theta_0))}\right)$ for each zero mean and unit variance innovation η_t.

Normal(0,1) Innovation:

$$\psi_t(\theta_0) = -\left(1 + \eta_t(\theta_0)\ \frac{f'(\eta_t(\theta_0))}{f(\eta_t(\theta_0))}\right) = -\left(1 - \eta_t^2\right).$$

Laplace(0,1) Innovation:

$$\psi_t(\theta_0) = -\left(1 + \eta_t(\theta_0)\ \frac{f'(\eta_t(\theta_0))}{f(\eta_t(\theta_0))}\right) = -\left(1 - \frac{1}{\sqrt{0.5}}\sqrt{\eta_t^2}\right).$$

Gaussian Mixture Innovation:

$$\psi_t(\theta_0) = -\left(1 + \eta_t(\theta_0)\ \frac{f'(\eta_t(\theta_0))}{f(\eta_t(\theta_0))}\right) = -1 - \eta_t\sqrt{5} \times ...$$

$$... \times \left\{ \frac{\left[\exp\left(-0.5\left(\sqrt{5}\eta_t - 2\right)^2\right)\left[-\left(\sqrt{5}\eta_t - 2\right)\right] + \exp\left(-0.5\left(\sqrt{5}\eta_t + 2\right)^2\right)\left[-\left(\sqrt{5}\eta_t + 2\right)\right]\right]}{\left[\exp\left(-0.5\left(\sqrt{5}\eta_t - 2\right)^2\right) + \exp\left(-0.5\left(\sqrt{5}\eta_t + 2\right)^2\right)\right]} \right\}.$$

t_5-student Innovation:

$$\psi_t(\theta_0) = -\left(1 + \eta_t(\theta_0)\ \frac{f'(\eta_t(\theta_0))}{f(\eta_t(\theta_0))}\right) = -\left(1 - \frac{6\ \eta_t^2}{3 + \eta_t^2}\right).$$

t_7-student Innovation:

$$\psi_t(\theta_0) = -\left(1 + \eta_t(\theta_0)\ \frac{f'(\eta_t(\theta_0))}{f(\eta_t(\theta_0))}\right) = -\left(1 - \frac{8\ \eta_t^2}{5 + \eta_t^2}\right).$$

t_9-student Innovation:

$$\psi_t(\theta_0) = -\left(1 + \eta_t(\theta_0)\ \frac{f'(\eta_t(\theta_0))}{f(\eta_t(\theta_0))}\right) = -\left(1 - \frac{10\ \eta_t^2}{7 + \eta_t^2}\right).$$

χ_6^2 Innovation:

$$\psi_t(\theta_0) = -\left(1 + \eta_t(\theta_0)\ \frac{f'(\eta_t(\theta_0))}{f(\eta_t(\theta_0))}\right) = -\left(1 - \frac{\sqrt{12}\eta_t + 6\eta_t^2}{\sqrt{12}\eta_t + 6} 1_{\left\{\sqrt{12}\eta_t + 6 \geq 0\right\}}\right),$$

where $1_{\{\cdot\}}$ denotes the indicator function.

χ_{12}^2 Innovation:

$$\psi_t(\theta_0) = -\left(1 + \eta_t(\theta_0)\ \frac{f'(\eta_t(\theta_0))}{f(\eta_t(\theta_0))}\right) = -\left(1 - \frac{\sqrt{24}\eta_t + 12\eta_t^2}{\sqrt{24}\eta_t + 12} 1_{\left\{\sqrt{24}\eta_t + 12 \geq 0\right\}}\right),$$

where $1_{\{\cdot\}}$ denotes the indicator function.

B Proof of the LAN Theorem 2.5.1

Proof of the LAN Theorem 2.5.1. The reparametrizated PTTGARCH(1.1) model in (2.3), (2.4) fits into the general time-series framework of DKW (1997) since it is a general location-scale model in which the location-scale parameters only depend on the past. Therefore, in order to prove Theorem 2.5.1, it suffices to verify the conditions (2.3'), (A.1) and (2.4) of DKW (1997); we also prove (3.3') of DKW (1997) since we need to verify that condition in the proof of Theorem 2.6.1. Thus, with the notation introduced in Section 2.5 [see eq. (2.20), (2.24), (2.25)], and denoting the expectation under θ of the product $\psi(\theta)\psi(\theta)'$ as $I_{ls}(f)$, we need to show, under θ_0,

$$n^{-1}\sum_{t=1}^{n} W_t(\theta_0) I_{ls}(f) W_t(\theta_0)' \xrightarrow{P} I(\theta_0) > 0, \tag{B.1}$$

$$n^{-1}\sum_{t=1}^{n}|W_t(\theta_0)|^2\,1_{\left\{n^{-1/2}|W_t(\theta_0)|>\delta\right\}}\xrightarrow{P}0, \tag{B.2}$$

$$n^{-1}\sum_{t=1}^{n}W_t(\theta_0)\xrightarrow{P}W(\theta_0), \tag{B.3}$$

$$n^{-1}\sum_{t=1}^{n}|W_t(\theta_n)-W_t(\theta_0)|^2\xrightarrow{P}0, \tag{B.4}$$

and, under θ_n,

$$\sum_{t=1}^{n}\left|n^{-1/2}(M_{nt},S_{nt})'-W_t(\theta_n)'(\tilde{\theta}_n-\theta_n)\right|^2\xrightarrow{P}0, \tag{B.5}$$

for some positive definitive matrix $I(\theta_0)$ and some random matrix $W(\theta_0)$. If conditions B.2-B.5 are satisfied, and using Lemma A.1 of DKW (1997), we can show that the PTTGARCH model satisfies the LAN property.

Although $W_t(\theta_0)$ may not be strictly stationary and ergodic under θ_0, the following proposition shows that these variables can be approximated by a strictly stationary and ergodic sequence, and hence B.1, B.2, B.3 and B.4 hold in general.

Proposition B.1. *Let $h_t(\theta)$, $H_t(\theta)$ and $W_t(\theta)$ be given by* (2.18), (2.23), (2.24), *respectively, and let $h_{st}(\theta)$, $H_{st}(\theta)$ and $W_{st}(\theta)$ be their corresponding stationary solutions under θ, i.e.*

$$h_{st}(\theta)=\sum_{j=0}^{\infty}\prod_{k=1}^{j}\left\{\beta+\alpha_+(\xi_{t-k}^{+})^2+\alpha_-(\xi_{t-k}^{-})^2\right\},$$

$$H_{st}(\theta)=\sum_{i=0}^{\infty}\beta^i h_{s,t-1-i}(\theta)\begin{pmatrix}(\xi_{t-1-i}^{+})^2\\(\xi_{t-1-i}^{-})^2\\1\end{pmatrix},$$

$$W_{st}(\theta)=\sigma^{-1}\begin{pmatrix}\frac{1}{2}h_{st}^{-1}(\theta)H_{st}(\theta)(\mu,\sigma)\\I_2\end{pmatrix},$$

Then, under θ_0,

$$n^{-1}\sum_{t=1}^{n}|W_t(\theta_0)-W_{st}(\theta_0)|^2\to 0\ \textit{(a.s.), as } n\to\infty. \tag{B.6}$$

Proof. If we iterate $h_t(\theta)$, we can rewrite $h_t(\theta)$ as

$$h_t(\theta)=1+\beta h_{t-1}(\theta)+\alpha_+(Y_{t-1}^{+})^2+\alpha_-(Y_{t-1}^{-})^2=1+h_{t-1}\left\{\beta+\alpha_+(\xi_{t-1}^{+}(\theta))^2+\alpha_-(\xi_{t-1}^{-}(\theta))^2\right\}=$$

$$=\sum_{j=0}^{i-1}\prod_{k=1}^{j}\left\{\beta+\alpha_+(\xi_{t-k}^{+}(\theta))^2+\alpha_-(\xi_{t-k}^{-}(\theta))^2\right\}+h_{t-i}(\theta)\prod_{k=1}^{i}\left\{\beta+\alpha_+(\xi_{t-k}^{+}(\theta))^2+\alpha_-(\xi_{t-k}^{-}(\theta))^2\right\},\quad 0\le i\le t-1, \tag{B.7}$$

which in turn implies that

$$h_{t-i}(\theta)/h_t(\theta)\le\prod_{k=1}^{i}\left\{\beta+\alpha_+(\xi_{t-k}^{+}(\theta))^2+\alpha_-(\xi_{t-k}^{-}(\theta))^2\right\}^{-1},\quad 0\le i\le t-1. \tag{B.8}$$

Under θ, the calculated variables $\xi_t(\theta)=\left\{\xi_t^{+}(\theta),\xi_t^{-}(\theta)\right\}$ are the true innovations $\xi_t=\left\{\xi_t^{+},\xi_t^{-}\right\}$ in (B.7) and (B.8),[19] and $\xi_t\sim i.i.d.(\mu,\sigma^2)$ is a stationary and ergodic process by definition. Similarly, if we iterate $h_{st}(\theta)$, we

[19] As in Hwang and Kim (2004) we use the following notation: $a_t^{+}=\max(a_t,0)$, and $a_t^{-}=\max(-a_t,0)$, so that $a_t=a_t^{+}-a_t^{-}$.

can rewrite $h_{st}(\theta)$ as

$$h_{st}(\theta) = \sum_{j=0}^{i-1}\prod_{k=1}^{j}\left\{\beta + \alpha_+(\xi_{t-k}^+)^2 + \alpha_-(\xi_{t-k}^-)^2\right\} + h_{s,t-i}(\theta)\prod_{k=1}^{i}\left\{\beta + \alpha_+(\xi_{t-k}^+)^2 + \alpha_-(\xi_{t-k}^-)^2\right\}, \quad 0 \le i,$$

$$h_{s,t-i}(\theta)/h_{st}(\theta) \le \prod_{k=1}^{i}\left\{\beta + \alpha_+(\xi_{t-k}^+)^2 + \alpha_-(\xi_{t-k}^-)^2\right\}^{-1}, \quad 0 \le i,$$

thus, under θ, we obtain

$$|h_{st}(\theta)h_{t-i}(\theta) - h_t(\theta)h_{s,t-i}(\theta)| = |h_{s,t-i}(\theta) - h_{t-i}(\theta)|\sum_{j=0}^{i-1}\prod_{k=1}^{j}\left\{\beta + \alpha_+(\xi_{t-k}^+)^2 + \alpha_-(\xi_{t-k}^-)^2\right\}$$

$$\le h_t(\theta)\,|h_{s1}(\theta) - h_1(\theta)|\prod_{k=1}^{t-1-i}\left\{\beta + \alpha_+(\xi_{t-i-k}^+)^2 + \alpha_-(\xi_{t-i-k}^-)^2\right\}$$

$$= h_t(\theta)\,|h_{s1}(\theta) - h_1(\theta)|\prod_{k=1}^{i}\left\{\beta + \alpha_+(\xi_{t-k}^+)^2 + \alpha_-(\xi_{t-k}^-)^2\right\}^{-1}\prod_{k=1}^{t-1}\left\{\beta + \alpha_+(\xi_k^+)^2 + \alpha_-(\xi_k^-)^2\right\}, \quad 0 \le i \le t-1.$$

Denoting a generic constant which only depends on θ as C, we obtain, under θ,

$$|W_t(\theta) - W_{st}(\theta)| \le C\,|H_t(\theta)/h_t(\theta) - H_{st}(\theta)/h_{st}(\theta)| \le \ldots$$

$$\le C\sum_{i=0}^{t-2}\beta^i\left|\frac{h_{t-1-i}(\theta)}{h_t(\theta)} - \frac{h_{s,t-1-i}(\theta)}{h_{st}(\theta)}\right|\left\|\begin{pmatrix}(\xi_{t-1-i}^+)^2\\(\xi_{t-1-i}^-)^2\\1\end{pmatrix}\right\| + C\sum_{i=t-1}^{\infty}\beta^i\frac{h_{s,t-1-i}(\theta)}{h_{st}(\theta)}\left\|\begin{pmatrix}(\xi_{t-1-i}^+)^2\\(\xi_{t-1-i}^-)^2\\1\end{pmatrix}\right\| \le \ldots$$

$$\le C\,|h_{s1}(\theta) - h_1(\theta)|\prod_{k=1}^{t-1}\left\{\beta + \alpha_+(\xi_k^+)^2 + \alpha_-(\xi_k^-)^2\right\}\sum_{i=0}^{t-2}\prod_{k=1}^{i}\frac{\beta}{\beta + \alpha_+(\xi_{t-k}^+)^2 + \alpha_-(\xi_{t-k}^-)^2} + C\sum_{i=t-1}^{\infty}\prod_{k=1}^{i}\frac{\beta}{\beta + \alpha_+(\xi_{t-k}^+)^2 + \alpha_-(\xi_{t-k}^-)^2} \le \ldots$$

$$\le C\,|h_{s1}(\theta) - h_1(\theta)|\,(t-1)\prod_{k=1}^{t-1}\left\{\beta + \alpha_+(\xi_k^+)^2 + \alpha_-(\xi_k^-)^2\right\} + C\prod_{k=1}^{t-1}\frac{\beta}{\beta + \alpha_+(\xi_k^+)^2 + \alpha_-(\xi_k^-)^2}\sum_{i=-1}^{\infty}\prod_{k=0}^{i}\frac{\beta}{\beta + \alpha_+(\xi_{-k}^+)^2 + \alpha_-(\xi_{-k}^-)^2}.$$

Using Assumption 2.1 [see eq. (2.5)], the right-hand side tends to zero (a.s.), as $t \to \infty$. Cesaro's Means Theorem [Hardy (1991), page 96] completes the proof of the proposition. □

In the following proposition -which is parallel to Proposition A.2 in DK (1997)- we show that slight perturbations of the parameters yield solutions of (2.11) and (2.12) that are close.

Proposition B.2. *Let $h_t(\theta)$ and $H_t(\theta)$ be given by* (2.18) *and* (2.23), *respectively, and define*

$$Q_t(\theta) = H_t(\theta)/h_t(\theta) = \sum_{i=0}^{t-2}\beta^i\begin{pmatrix}(Y_{t-1-i}^+)^2\\(Y_{t-1-i}^-)^2\\h_{t-1-i}(\theta)\end{pmatrix}/h_t(\theta) = \sum_{i=0}^{t-2}\beta^i\frac{h_{t-1-i}(\theta)}{h_t(\theta)}\begin{pmatrix}(\xi_{t-1-i}^+(\theta))^2\\(\xi_{t-1-i}^-(\theta))^2\\1\end{pmatrix},$$

$$R_t(\theta,\tilde\theta) = h_t(\tilde\theta)/h_t(\theta) - 1 - \left(\tilde\alpha_+ - \alpha, \tilde\alpha_- - \alpha, \tilde\beta - \beta\right)Q_t(\theta).$$

Let θ_n and $\tilde\theta_n$ satisfy the conditions just above (2.15). *Put $Q_{nt} = Q_t(\theta_n)$ and $R_{nt} = R_t(\theta_n, \tilde\theta_n)$. Then, under θ_n,*

$$n^{-1}\sum_{t=1}^{n}|Q_{nt}|^2 = \mathrm{O}_p(1),\ n^{-1}\sum_{t=1}^{n}|Q_{nt}|^2\,1_{\left\{n^{-1/2}|Q_{nt}|>\delta\right\}} \to 0,\ \text{(a.s.) as } n \to \infty, \tag{B.9}$$

$$\sum_{t=1}^{n} R_{nt}^2 \to 0, \text{ (a.s.) as } n \to \infty. \tag{B.10}$$

Proof. Using (B.8) we obtain

$$Q_t(\theta) \leq \beta^{-1} \sum_{i=0}^{t-2} \prod_{k=1}^{i+1} \frac{\beta}{\beta + \alpha_+(\xi_{t-k}^+(\theta))^2 + \alpha_-(\xi_{t-k}^-(\theta))^2} \begin{pmatrix} (\xi_{t-1-i}^+(\theta))^2 \\ (\xi_{t-1-i}^-(\theta))^2 \\ 1 \end{pmatrix}.$$

The above inequality shows that, for n sufficiently large and under θ_n, $|Q_{nt}|$ can be bounded by the product of a constant depending on θ_0 only and the stationary sequence

$$S_t = \sum_{i=0}^{\infty} \prod_{k=1}^{i} \frac{\beta_0}{\beta_0 + \frac{1}{2}\alpha_{0+}(\xi_{t-k}^+)^2 + \frac{1}{2}\alpha_{0-}(\xi_{t-k}^-)^2}.$$

In view of the latter result, relations regarding Q_{nt} in (B.9) can be easily obtained.

To prove (B.10), note that an explicit relationship for the difference of $h_t(\tilde{\theta})$ and $h_t(\theta)$ is given by

$$h_t(\tilde{\theta}) - h_t(\theta) = \sum_{i=0}^{t-2} \tilde{\beta}^i h_{t-1-i}(\theta) \left(\tilde{\alpha}_+ - \alpha_+, \tilde{\alpha}_- - \alpha_-, \tilde{\beta} - \beta\right) \begin{pmatrix} (\xi_{t-1-i}^+(\theta))^2 \\ (\xi_{t-1-i}^-(\theta))^2 \\ 1 \end{pmatrix}.$$

Thus, the remainder term $R_t(\theta, \tilde{\theta})$ can be written as

$$R_t(\theta, \tilde{\theta}) = \sum_{i=0}^{t-2} (\tilde{\beta}^i - \beta^i) \frac{h_{t-1-i}(\theta)}{h_t(\theta)} \left(\tilde{\alpha}_+ - \alpha_+, \tilde{\alpha}_- - \alpha_-, \tilde{\beta} - \beta\right) \begin{pmatrix} (\xi_{t-1-i}^+(\theta))^2 \\ (\xi_{t-1-i}^-(\theta))^2 \\ 1 \end{pmatrix}.$$

Choose a constant $C > 1$ such that $E\left(\frac{C\beta_0}{\beta_0 + \frac{1}{2}\alpha_{0+}(\xi_1^+)^2 + \frac{1}{2}\alpha_{0-}(\xi_1^-)^2}\right) < 1$ is satisfied. By the mean value theorem, there exists a $\tilde{\beta}_{ni}$, which lies between $\tilde{\beta}_n$ and β_n such that, for n sufficiently large,

$$\left|\tilde{\beta}_n^i - \beta_n^i\right| = \left|\tilde{\beta}_n - \beta_n\right| i \tilde{\beta}_{ni}^{i-1} \leq \left|\tilde{\beta}_n - \beta_n\right| i c^{i-1} \beta_n^{i-1}, \quad i \geq 0.$$

Thus, we can bound R_{nt} by the product of a constant times n^{-1} and the stationary sequence

$$S_t^* = \sum_{i=0}^{\infty} i \prod_{k=1}^{i} \frac{c\beta_0}{\beta_0 + \frac{1}{2}\alpha_{0+}(\xi_{t-k}^+)^2 + \frac{1}{2}\alpha_{0-}(\xi_{t-k}^-)^2}.$$

The proof of (B.10) can be easily completed. □

Below, we prove that (B.2)-(B.5) hold. We first show that (B.2) and (B.3) hold; then, we show that (B.5) and (B.4) hold, respectively.

Define $I(\theta_0) = E_{\theta_0} W_{s1}(\theta_0) I_{ls}(f) W_{s1}(\theta_0)'$ and $W(\theta_0) = E_{\theta_0} W_{s1}(\theta_0)$; the existence of $I(\theta_0)$ and $W(\theta_0)$ can be obtained as in the proof of Proposition B.2 since $|W_{st}(\theta_0)|$ can be bounded by the product of S_t and a constant depending only on θ_0. The relations (B.2) and (B.3) hold if $W_t(\theta_0)$ is replaced by the stationary and ergodic sequence $W_{st}(\theta_0)$. Finally, Proposition B.1 implies that the previous relations are valid also in the case of $W_t(\theta_0)$.

We will use Proposition B.2 to prove (B.5). Writing $\lambda_n = (\lambda'_{1n}, \lambda'_{2n})$ with λ_{1n} (λ_{2n}) the first (latter) two components of λ_n, and defining

$$\chi(x) = \{-1 + 2\left(\sqrt{1+x} - 1\right)/x\}1_{\{x \geq -1\}},$$

we obtain

$$\sum_{t=1}^{n} \left| n^{-1/2}(M_{nt}, S_{nt})' - W_t(\theta_n)'(\tilde{\theta}_n - \theta_n) \right|^2 = ...$$

$$= \sigma_n^{-2} n^{-1} \sum_{t=1}^{n} |(\tilde{\mu}_n, \tilde{\sigma}_n)' \left\{ \frac{1}{2}\left(\lambda'_{1n} Q_{nt} + \sqrt{n} R_{nt}\right) \chi\left(n^{-1/2}\lambda'_{1n} Q_{nt} + R_{nt}\right) + \frac{1}{2}\sqrt{n} R_{nt} \right\} + n^{-1/2}\lambda_{2n}\frac{1}{2}\lambda'_{1n} Q_{nt}|^2.$$

Using the above equation, Proposition B.2 and Lemma 2.1 of DKW (1997) (with $Y_{nt} = \lambda'_{1n} Q_{nt}$, $X_{nt} = \lambda'_{1n} Q_{nt} + \sqrt{n} R_{nt}$, and the function $\phi = \chi^2$ as above) we obtain that (B.5) holds.

To prove that (B.4) holds, note that

$$|W_t(\theta_n) - W_t(\theta_0)|^2 \leq C\,|Q_t(\theta_n) - Q_t(\theta_0)|^2 + C\,|Q_t(\theta_0)|^2\,|\theta_n - \theta_0|^2,$$

and obtain contiguity of P_{θ_n} and P_{θ_0} from (B.2) and (B.5), and Theorem 2.1 of DKW (1997). The required result is obtained from

$$Q_t(\tilde{\theta}) - Q_t(\theta) = \sum_{i=0}^{t-2} (\tilde{\beta}^i - \beta^i)\frac{h_{t-1-i}(\tilde{\theta})}{h_t(\tilde{\theta})} \begin{pmatrix} (\xi^+_{t-1-i}(\tilde{\theta}))^2 \\ (\xi^-_{t-1-i}(\tilde{\theta}))^2 \\ 1 \end{pmatrix} + Q_t(\theta)\left\{(\theta_1 - \tilde{\theta}_1)' Q_t(\tilde{\theta}) + R_t(\tilde{\theta}, \theta)\right\} - ...$$

$$- \begin{pmatrix} 0 \\ 0 \\ \sum_{i=0}^{t-2} \tilde{\beta}^i \frac{h_{t-1-i}(\tilde{\theta})}{h_t(\tilde{\theta})} \left\{(\theta_1 - \tilde{\theta}_1)' Q_{t-1-i}(\tilde{\theta}) + R_{t-1-i}(\tilde{\theta}, \theta)\right\} \end{pmatrix}$$

in the same way as in the proofs of the propositions above. This completes the proofs of the theorems in Section 2.5. □

C Proof of Theorem 2.6.1

Proof of Theorem 2.6.1. It is sufficient to verify the conditions of DKW (1997). The conditions of DKW (1997) are composed by (B.2)-(B.5), which are verified in Appendix B, and by the existence of an estimator $\hat{\psi}_n(\cdot)$, based on $\eta_1, ..., \eta_n$, of $\psi(\cdot) = -(\ell'(\cdot), 1 + \cdot\ell'(\cdot))'$ [see equation (2.25)] which satisfies the consistency condition

$$\int \left| \hat{\psi}_n(x) - \psi(x) \right|^2 f(x)dx \xrightarrow{P} 0, \quad \text{under } f.$$

Such an estimator $\hat{\psi}_n(\cdot)$ exists in view of Proposition 7.8.1 of BKRW (1993) with k=0 or k=1. The estimator $\hat{\psi}_n(\cdot)$ used in Section 2.6 satisfies Proposition 7.8.1 of BKRW (1993). This completes the proof of Theorem 2.6.1. $\square$

References

Awartani, B.M.A., and Corradi, V., 2005. "Predicting the volatility of the S&P-500 stock index via GARCH models: the role of asymmetries," *International Journal of Forecasting, 21(1),* 167-183.

Bickel, P.J., 1982. "On adaptive estimation," *The Annals of Statistics, 10(3)*, 647-671.

Bickel, P.J., Klaassen, C.A.J., Ritov, Y., and Wellner, J.A., 1993. "Efficient and adaptive estimation for semiparametric models," *Johns Hopkins University Press*, Baltimore.

Black, F., 1976. "Studies of stock price volatility changes," in *Proceedings from the American Statistical Association, Business and Economics Statistics Section*, 177-181.

Bollerslev, T., 1986. "Generalized autoregressive conditional heteroskedasticity," *Journal of Econometrics, 31(3)*, 307-327.

Christie, A.A., 1982. "The stochastic behavior of common stock variances: value, leverage and interest rate effects," *Journal of Financial Economics, 10(4)*, 407-432.

Diaz-Emparanza, I., 2002. "Is a small Monte Carlo analysis a good analysis?" *Statistical Papers, 43(4)*, 567-577.

Drost, F.C., and Klaassen, C.A.J., 1997. "Efficient estimation in semiparametric GARCH models," *Journal of Econometrics, 81(1)*, 193-221.

Drost, F.C., Klaassen, C. A. J., and Werker, B. J. M., 1997. "Adaptive estimation in time series models," *The annals of statistics, 25(2)*, 786-817.

Engle, R.F., 1982. "Autoregressive Conditional Heteroscedasticity with Estimates of the Variance of United Kingdom Inflation," *Econometrica, 50(4),* 987-1007.

Engle, R.F., 1990. "Discussion: Stock Market Volatility and the Crash of 87," *Review of Financial Studies, 3(1)*, 103-106.

Engle, R.F., and Ng, V.K., 1993. "Measuring and Testing the Impact of News on Volatility," *The Journal of Finance, 48(5)*, 1749-1778.

Francq, C., and Zakoïan, J.M., 2010. "GARCH Models. Structure, statistical inference and financial applications," *John Wiley & Sons*, Chichester, West Sussex, United Kingdom.

Hájek, J., 1970. "A characterization of limiting distributions of regular estimates," *Zeitschrift fur Wahrscheinlichkeitstheorie und Verwandte Gebiete, 14(4)*, 323-330.

Hamadeh, T., and Zakoïan, J.M., 2011. "Asymptotic properties of LS and QML estimators for a class of nonlinear GARCH processes," *Journal of Statistical Planning and Inference, 141(1)*, 488-507.

Hardy, G.H., 1991. "Divergent Series," *Chelsea Publishing Company*, New York.

Hwang, S.Y., and Kim, T.Y., 2004. "Power transformation and threshold modeling for ARCH innovations with applications to tests for ARCH structure," *Stochastic Processes and their Applications 110(2),* 295-314.

Kreiss, J.P., 1987. "On adaptive estimation in stationary ARMA processes," *The Annals of Statistics, 15(1)*, 112-133.

Le Cam, L., 1960. "Locally Asymptotically Normal Families of Distributions. Certain Approximations to Families of Distributions and Their Use in the Theory of Estimation and Testing Hypotheses," *Berkeley & Los Angeles: University of California Press.*

Linton, O., 1993. "Adaptive Estimation in Arch Models", Cowles Foundation Discussion Paper No. 1054, *Yale University.*

Liu, J.C., 2006. "On the tail behaviors of Box-Cox transformed threshold GARCH(1,1) process," *Statistics & Probability Letters, 76(13)*, 1323-1330.

Nelson, D.B., 1991. "Conditional Heteroskedasticity in Asset Returns: A New Approach," *Econometrica, 59(2)*, 347-370.

Newey, W.K., 1990. "Semiparametric efficient bounds," *Journal of Applied Econometrics, 5(2)*, 99-135.

Pakel, C., Shephard, N., and Sheppard, K., 2011. "Nuisance parameters, composite likelihoods and a panel of GARCH models," *Statistica Sinica, 21(1)*, 307-329.

Pan, J., Wang, H., and Tong, H., 2008. "Estimation and tests for power-transformed and threshold GARCH models," *Journal of Econometrics, 142(1)*, 352-378.

Roussas, G.G., 1979. "Asymptotic distribution of the log likelihood function for stochastic processes," *Zeitschrift fÃijr Wahrscheinlichkeitstheorie und Verwandte Gebiete, 47(1)*, 31-46.

Schick, A., 1986. "On asymmetrically efficient estimation in semiparametric models," *The Annals of Statistics, 14(3)*, 1139-1151.

Schwert, G.W., 1989. "Why does stock market volatility change over time?", *Journal of Finance, 44(5)*, 1115-1153.

Sun, Y., and Stengos, T., 2006. "Semiparametric efficient adaptive estimation of asymmetric GARCH models," *Journal of Econometrics, 133(1)*, 373-386.

Swensen, A.R., 1985. "The Asymptotic Distribution of the Likelihood Ratio for Autoregressive Time Series with a Regression Trend," *Journal of Multivariate Analysis, 16(1)*, 54-70.

Taylor, S.J., 1986. "Modelling Financial Time Series," *John Wiley & Sons, Ltd.*

Wand, M.P., and Jones, M.C., 1995. "Kernel smoothing," *Chapman & Hall/CRC*, London.

Wellner, J.A., Klaassen, C.A.J., and Ritov, Y., 2006. "Semiparametric Models: a Review of Progress since BKRW (1993)," in Fan, J. and Koul, H.L., 2006. "Frontiers in Statistics", *World Scientific*.

Table 1: Probability density functions of the zero mean, unit variance innovation η_t

	Probability density functions
$\mathcal{N}(0,1)$	$f(\eta_t) = \frac{1}{\sqrt{2\pi}} \exp\left(-\frac{1}{2}\eta_t^2\right)$
$Laplace(0,1)$	$f(\eta_t) = \frac{1}{2\sqrt{0.5}} \exp\left(-\frac{\lvert\eta_t\rvert}{\sqrt{0.5}}\right)$
$Gaussian\ Mixture$	$f\left(\eta_t\right) = \sqrt{5}\left[0.5\left\{\frac{1}{\sqrt{2\pi}} \exp\left(-0.5\left(\sqrt{5}\eta_t - 2\right)^2\right)\right\}\right] + ...$
	$... + \sqrt{5}\left[+0.5\left\{\frac{1}{\sqrt{2\pi}} \exp\left(-0.5\left(\sqrt{5}\eta_t + 2\right)^2\right)\right\}\right]$
$t_5 - student$	$f(\eta_t) = \sqrt{5/3}\left[\frac{\Gamma(3)}{\Gamma(2.5)\sqrt{5\pi}}\left(1+\frac{\eta_t^2}{3}\right)^{-3}\right]$
$t_7 - student$	$f(\eta_t) = \sqrt{7/5}\left[\frac{\Gamma(4)}{\Gamma(3.5)\sqrt{7\pi}}\left(1+\frac{\eta_t^2}{5}\right)^{-4}\right]$
$t_9 - student$	$f(\eta_t) = \sqrt{9/7}\left[\frac{\Gamma(5)}{\Gamma(4.5)\sqrt{9\pi}}\left(1+\frac{\eta_t^2}{7}\right)^{-5}\right]$
χ_6^2	$f(\eta_t) = \sqrt{12}\left[\left(\frac{1}{2^3\Gamma(3)}\right)\left(\sqrt{12}\eta_t + 6\right)^2 \exp\left(-0.5\left(\sqrt{12}\eta_t + 6\right)\right)\right] 1_{\{\sqrt{12}\eta_t + 6 \geq 0\}}$
χ_{12}^2	$f(\eta_t) = \sqrt{24}\left[\left(\frac{1}{2^6\Gamma(6)}\right)\left(\sqrt{24}\eta_t + 12\right)^5 \exp\left(-0.5\left(\sqrt{24}\eta_t + 12\right)\right)\right] 1_{\left\{\sqrt{24}\eta_t + 12 \geq 0\right\}}$

Notes: $1_{\{.\}}$ denotes the indicator function.

Table 2: Conditional log-likelihood function for each (zero mean and unit variance) innovation η_t

	Conditional log-likelihood functions
$\mathcal{N}(0,1)$	$\ln L(\theta \vert \mathbf{Y}) = \sum_{t=2}^{T} \ln\left(\frac{1}{\sigma\sqrt{h_t}}\left[\frac{1}{\sqrt{2\pi}}\exp\left(-\frac{1}{2}\left(\frac{Y_t-\mu\sqrt{h_t}}{\sigma\sqrt{h_t}}\right)^2\right)\right]\right)$
$L(0,1)$	$\ln L(\theta \vert \mathbf{Y}) = \sum_{t=2}^{T} \ln\left(\frac{1}{\sigma\sqrt{h_t}}\left[\frac{1}{2\sqrt{0.5}}\exp\left(-\frac{\left\vert\frac{Y_t-\mu\sqrt{h_t}}{\sigma\sqrt{h_t}}\right\vert}{\sqrt{0.5}}\right)\right]\right)$
GM	$\ln L(\theta \vert \mathbf{Y}) = \sum_{t=2}^{T} \ln\left(\frac{1}{\sigma\sqrt{h_t}}\left[0.5\left(\frac{1}{\sqrt{2\pi}}\exp\left(-\frac{1}{2}\left(\frac{Y_t-\mu\sqrt{h_t}}{\sigma\sqrt{h_t}}-2\right)^2\right)\right)+0.5\left(\frac{1}{\sqrt{2\pi}}\exp\left(-\frac{1}{2}\left(\frac{Y_t-\mu\sqrt{h_t}}{\sigma\sqrt{h_t}}+2\right)^2\right)\right)\right]\right)$
t_5	$\ln L(\theta \vert \mathbf{Y}) = \sum_{t=2}^{T} \ln\left(\frac{\sqrt{1.6}}{\sigma\sqrt{h_t}}\left[\frac{\Gamma(3)}{\Gamma(2.5)\sqrt{5\pi}}\left(1+\frac{\left(\frac{Y_t-\mu\sqrt{h_t}}{\sigma\sqrt{h_t}}\right)^2}{3}\right)^{-3}\right]\right)$
t_7	$\ln L(\theta \vert \mathbf{Y}) = \sum_{t=2}^{T} \ln\left(\frac{\sqrt{1.4}}{\sigma\sqrt{h_t}}\left[\frac{\Gamma(4)}{\Gamma(3.5)\sqrt{7\pi}}\left(1+\frac{\left(\frac{Y_t-\mu\sqrt{h_t}}{\sigma\sqrt{h_t}}\right)^2}{5}\right)^{-4}\right]\right)$
t_9	$\ln L(\theta \vert \mathbf{Y}) = \sum_{t=2}^{T} \ln\left(\frac{\sqrt{9/7}}{\sigma\sqrt{h_t}}\left[\frac{\Gamma(5)}{\Gamma(4.5)\sqrt{7\pi}}\left(1+\frac{\left(\frac{Y_t-\mu\sqrt{h_t}}{\sigma\sqrt{h_t}}\right)^2}{7}\right)^{-5}\right]\right)$
χ^2_6	$\ln L(\theta \vert \mathbf{Y}) = \sum_{t=2}^{T} \ln\left(\frac{\sqrt{12}}{\sigma\sqrt{h_t}}\left[\left(\frac{1}{2^3\Gamma(3)}\right)\left(\sqrt{12}\frac{Y_t-\mu\sqrt{h_t}}{\sigma\sqrt{h_t}}+6\right)^2\exp\left(-0.5\left(\sqrt{12}\frac{Y_t-\mu\sqrt{h_t}}{\sigma\sqrt{h_t}}+6\right)\right)\right]1_{\left\{\sqrt{12}\frac{Y_t-\mu\sqrt{h_t}}{\sigma\sqrt{h_t}}+6\geq 0\right\}}\right)$
χ^2_{12}	$\ln L(\theta \vert \mathbf{Y}) = \sum_{t=2}^{T} \ln\left(\frac{\sqrt{24}}{\sigma\sqrt{h_t}}\left[\left(\frac{1}{2^6\Gamma(6)}\right)\left(\sqrt{24}\frac{Y_t-\mu\sqrt{h_t}}{\sigma\sqrt{h_t}}+12\right)^5\exp\left(-0.5\left(\sqrt{24}\frac{Y_t-\mu\sqrt{h_t}}{\sigma\sqrt{h_t}}+12\right)\right)\right]1_{\left\{\sqrt{24}\frac{Y_t-\mu\sqrt{h_t}}{\sigma\sqrt{h_t}}+12\geq 0\right\}}\right)$

Notes: $1_{\{.\}}$ denotes the indicator function.

Table 3: The kurtosis of each innovation η_t

	$\mathcal{N}(0,1)$	Laplace(0,1)	Gaussian mixture	t_5	t_7	t_9	χ^2_6	χ^2_{12}
$E\eta_t^4$	3	6	1.72	9	5	4.2	5	4

Table 4: Fisher information for scale for each type of innovation

	$\mathcal{N}(0,1)$	Laplace(0,1)	Gaussian mixture	t_5	t_7	t_9	χ^2_6	χ^2_{12}
$I_s(f)$	2	1	5.9134	1.25	1.40	1.50	6	3

Table 5: Sample Means and Sample Standard Deviations of Monte Carlo Estimates

		α_+	α_-	β				α_+	α_-	β				α_+	α_-	β			
		$\hat\alpha_+$	$\hat\alpha_-$	$\hat\beta$	$\hat\sigma_{\alpha_+}$	$\hat\sigma_{\alpha_-}$	$\hat\sigma_\beta$	$\hat\alpha_+$	$\hat\alpha_-$	$\hat\beta$	$\hat\sigma_{\alpha_+}$	$\hat\sigma_{\alpha_-}$	$\hat\sigma_\beta$	$\hat\alpha_+$	$\hat\alpha_-$	$\hat\beta$	$\hat\sigma_{\alpha_+}$	$\hat\sigma_{\alpha_-}$	$\hat\sigma_\beta$
$\mathcal{N}(0,1)$	ML=QML	0.200	0.400	0.598	0.042	0.069	0.038	0.075	0.152	0.799	0.026	0.038	0.041	0.086	0.119	0.872	0.031	0.039	0.026
	1 step	0.198	0.398	0.597	0.042	0.068	0.038	0.075	0.151	0.799	0.028	0.040	0.043	0.084	0.117	0.872	0.033	0.040	0.026
	As.Std.Dev.ML				0.041	0.066	0.036				0.024	0.036	0.035				0.025	0.030	0.020
$L(0,1)$	ML	0.198	0.400	0.596	0.052	0.084	0.046	0.075	0.151	0.794	0.028	0.041	0.042	0.079	0.108	0.866	0.028	0.034	0.026
	QML	0.198	0.402	0.595	0.059	0.096	0.051	0.075	0.152	0.793	0.031	0.047	0.050	0.082	0.112	0.867	0.032	0.039	0.033
	1 step	0.196	0.397	0.597	0.054	0.086	0.048	0.073	0.148	0.797	0.029	0.042	0.044	0.077	0.106	0.868	0.029	0.036	0.029
	As.Std.Dev.ML				0.052	0.082	0.045				0.028	0.041	0.040				0.027	0.033	0.023
	As.Std.Dev.QML				0.058	0.092	0.051				0.031	0.046	0.045				0.030	0.037	0.026
GM	ML	0.199	0.399	0.600	0.030	0.046	0.026	0.073	0.148	0.796	0.020	0.028	0.030	0.077	0.107	0.869	0.020	0.025	0.018
	QML	0.199	0.399	0.599	0.032	0.053	0.029	0.086	0.174	0.822	0.023	0.031	0.031	0.077	0.107	0.868	0.021	0.026	0.018
	1 step	0.199	0.399	0.599	0.032	0.053	0.028	0.076	0.155	0.810	0.024	0.033	0.031	0.077	0.107	0.869	0.023	0.027	0.020
	As.Std.Dev.ML				0.032	0.052	0.027				0.020	0.029	0.028				0.020	0.025	0.017
	As.Std.Dev.QML				0.033	0.054	0.028				0.020	0.030	0.029				0.021	0.026	0.017
t_5	ML	0.200	0.401	0.597	0.050	0.080	0.045	0.075	0.150	0.795	0.028	0.039	0.041	0.078	0.108	0.867	0.025	0.030	0.024
	QML	0.204	0.407	0.594	0.073	0.111	0.061	0.076	0.153	0.790	0.038	0.055	0.060	0.083	0.113	0.866	0.040	0.043	0.035
	1 step	0.199	0.398	0.596	0.064	0.098	0.057	0.073	0.148	0.794	0.033	0.048	0.054	0.073	0.101	0.870	0.038	0.040	0.026
	As.Std.Dev.ML				0.049	0.077	0.043				0.026	0.038	0.038				0.025	0.031	0.022
	As.Std.Dev.QML				0.077	0.122	0.068				0.041	0.061	0.059				0.040	0.049	0.035
t_7	ML	0.200	0.398	0.597	0.047	0.075	0.043	0.073	0.149	0.794	0.026	0.038	0.041	0.078	0.108	0.869	0.030	0.032	0.024
	QML	0.201	0.399	0.596	0.057	0.086	0.049	0.074	0.151	0.794	0.031	0.046	0.051	0.082	0.111	0.871	0.034	0.038	0.029
	1 step	0.198	0.394	0.598	0.049	0.078	0.044	0.072	0.146	0.797	0.028	0.040	0.043	0.074	0.104	0.870	0.034	0.034	0.024
	As.Std.Dev.ML				0.047	0.074	0.041				0.026	0.038	0.037				0.026	0.032	0.022
	As.Std.Dev.QML				0.055	0.088	0.049				0.031	0.045	0.044				0.030	0.037	0.026
t_9	ML	0.199	0.401	0.598	0.046	0.076	0.041	0.074	0.150	0.796	0.027	0.039	0.040	0.077	0.108	0.868	0.025	0.032	0.024
	QML	0.201	0.404	0.598	0.052	0.086	0.045	0.076	0.154	0.798	0.029	0.043	0.046	0.083	0.116	0.873	0.029	0.039	0.027
	1 step	0.198	0.398	0.598	0.047	0.077	0.043	0.073	0.148	0.797	0.027	0.040	0.041	0.075	0.104	0.870	0.026	0.035	0.024
	As.Std.Dev.ML				0.046	0.073	0.040				0.026	0.038	0.037				0.026	0.032	0.022
	As.Std.Dev.QML				0.050	0.080	0.044				0.028	0.041	0.041				0.028	0.035	0.024
χ^2_6	ML	0.195	0.394	0.598	0.030	0.052	0.029	0.072	0.148	0.797	0.019	0.035	0.024	0.077	0.108	0.866	0.017	0.027	0.017
	QML	0.203	0.410	0.597	0.054	0.118	0.060	0.077	0.161	0.793	0.028	0.074	0.070	0.082	0.117	0.867	0.030	0.065	0.036
	1 step	0.192	0.379	0.603	0.036	0.077	0.039	0.068	0.130	0.804	0.020	0.055	0.042	0.070	0.088	0.873	0.023	0.056	0.021
	As.Std.Dev.ML				0.022	0.046	0.024				0.012	0.028	0.024				0.012	0.022	0.013
	As.Std.Dev.QML				0.054	0.113	0.060				0.029	0.069	0.058				0.028	0.053	0.031
χ^2_{12}	ML	0.199	0.398	0.599	0.030	0.055	0.031	0.074	0.148	0.798	0.015	0.031	0.024	0.079	0.108	0.870	0.015	0.025	0.015
	QML	0.200	0.404	0.597	0.048	0.096	0.051	0.077	0.159	0.798	0.026	0.057	0.058	0.085	0.123	0.871	0.028	0.063	0.030
	1 step	0.195	0.390	0.599	0.037	0.069	0.039	0.070	0.139	0.799	0.020	0.042	0.040	0.072	0.094	0.869	0.025	0.042	0.021
	As.Std.Dev.ML				0.032	0.062	0.033				0.018	0.036	0.033				0.018	0.029	0.018
	As.Std.Dev.QML				0.048	0.093	0.049				0.026	0.054	0.049				0.026	0.043	0.027

Comparison of the MLE, QMLE, and the semiparametric estimator for the PTTGARCH(1,1) model with eight standardized innovation distributions. Number of observations N= 2000, true parameters $(\alpha_+, \alpha_-, \beta, \sigma) = (0.2, 0.4, 0.6, 1)$, $(0.075, 0.15, 0.8, 1)$, $(0.08, 0.11, 0.87, 1)$, respectively. The sample means of 2500 independent replications and their sample standard deviations are reported in the table. The theoretical asymptotic standard deviations of the MLE and QMLE for the parameters α_+, α_- and β is reported for each distribution of the innovation using 5000 replications [see "As.Std.Dev.ML" and "As.Std.Dev.QML", respectively].

Table 6: Asymptotic Relative Efficiencies (AREs) of the ML and QML Estimators

		$j=\alpha_+$	$j=\alpha_-$	$j=\beta$	$j=\alpha_+$	$j=\alpha_-$	$j=\beta$	$j=\alpha_+$	$j=\alpha_-$	$j=\beta$
		0.20	0.40	0.60	0.075	0.15	0.80	0.08	0.11	0.87
$L(0,1)$	$ARE_j = \frac{\hat{\sigma}_j^{QML}}{\hat{\sigma}_j^{ML}}$	1.115	1.122	1.133	1.107	1.122	1.125	1.111	1.121	1.130
GM	$ARE_j = \frac{\hat{\sigma}_j^{QML}}{\hat{\sigma}_j^{ML}}$	1.031	1.038	1.037	1.000	1.034	1.036	1.050	1.040	1.000
t_5	$ARE_j = \frac{\hat{\sigma}_j^{QML}}{\hat{\sigma}_j^{ML}}$	1.571	1.584	1.581	1.577	1.605	1.553	1.600	1.581	1.591
t_7	$ARE_j = \frac{\hat{\sigma}_j^{QML}}{\hat{\sigma}_j^{ML}}$	1.170	1.189	1.195	1.192	1.184	1.189	1.182	1.156	1.182
t_9	$ARE_j = \frac{\hat{\sigma}_j^{QML}}{\hat{\sigma}_j^{ML}}$	1.087	1.096	1.100	1.077	1.079	1.108	1.077	1.094	1.091
χ_6^2	$ARE_j = \frac{\hat{\sigma}_j^{QML}}{\hat{\sigma}_j^{ML}}$	2.455	2.457	2.5	2.417	2.464	2.417	2.333	2.410	2.385
χ_{12}^2	$ARE_j = \frac{\hat{\sigma}_j^{QML}}{\hat{\sigma}_j^{ML}}$	1.500	1.500	1.485	1.444	1.500	1.485	1.444	1.483	1.500

Notes: the asymptotic relative efficiencies (AREs) of the ML and QML estimators are obtained by dividing the QMLE theoretical asymptotic standard deviation of parameter j by the MLE theoretical asymptotic standard deviation of parameter j. The QMLE theoretical asymptotic standard deviation of parameter j and the MLE theoretical asymptotic standard deviation of parameter j are reported in Table 5 [see "As.Std.Dev.QML" and "As.Std.Dev.ML" in Table 5].

Table 7: Relative Efficiencies (REs) of the MLE, QMLE and the Semiparametric Estimator

		(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)
		$j=\alpha_+$	$j=\alpha_-$	$j=\beta$	$j=\alpha_+$	$j=\alpha_-$	$j=\beta$	$j=\alpha_+$	$j=\alpha_-$	$j=\beta$
		0.20	0.40	0.60	0.075	0.15	0.80	0.08	0.11	0.87
$\mathcal{N}(0,1)$	$RE_j^{AE/ML}=\frac{\hat{\sigma}_j^{AE}}{\hat{\sigma}_j^{ML}}$	1.000	1.015	1.000	1.077	1.053	1.049	1.065	1.026	1.000
$L(0,1)$	$RE_j^{QML/ML}=\frac{\hat{\sigma}_j^{QML}}{\hat{\sigma}_j^{ML}}$	1.135	1.143	1.109	1.107	1.146	1.190	1.143	1.147	1.269
	$RE_j^{QML/AE}=\frac{\hat{\sigma}_j^{QML}}{\hat{\sigma}_j^{AE}}$	1.093	1.116	1.063	1.069	1.119	1.136	1.103	1.083	1.138
	$RE_j^{ML-AE}=\frac{\hat{\sigma}_j^{AE}}{\hat{\sigma}_j^{ML}}$	1.038	1.023	1.043	1.036	1.024	1.048	1.036	1.059	1.115
GM	$RE_j^{QML/ML}=\frac{\hat{\sigma}_j^{QML}}{\hat{\sigma}_j^{ML}}$	1.067	1.152	1.115	1.150	1.107	1.033	1.050	1.040	1.000
	$RE_j^{QML/AE}=\frac{\hat{\sigma}_j^{QML}}{\hat{\sigma}_j^{AE}}$	1.000	1.000	1.036	0.958	0.939	1.000	0.913	0.963	0.900
	$RE_j^{AE/ML}=\frac{\hat{\sigma}_j^{AE}}{\hat{\sigma}_j^{ML}}$	1.067	1.152	1.077	1.200	1.179	1.033	1.150	1.080	1.111
t_5	$RE_j^{QML/ML}=\frac{\hat{\sigma}_j^{QML}}{\hat{\sigma}_j^{ML}}$	1.460	1.388	1.356	1.357	1.410	1.463	1.600	1.433	1.458
	$RE_j^{QML/AE}=\frac{\hat{\sigma}_j^{QML}}{\hat{\sigma}_j^{AE}}$	1.141	1.133	1.070	1.152	1.146	1.111	1.053	1.075	1.346
	$RE_j^{AE/ML}=\frac{\hat{\sigma}_j^{AE}}{\hat{\sigma}_j^{ML}}$	1.280	1.225	1.266	1.179	1.231	1.317	1.520	1.333	1.083
t_7	$RE_j^{QML/ML}=\frac{\hat{\sigma}_j^{QML}}{\hat{\sigma}_j^{ML}}$	1.212	1.147	1.139	1.192	1.211	1.243	1.133	1.188	1.210
	$RE_j^{QML/AE}=\frac{\hat{\sigma}_j^{QML}}{\hat{\sigma}_j^{AE}}$	1.163	1.103	1.114	1.107	1.150	1.186	1.000	1.118	1.208
	$RE_j^{AE/ML}=\frac{\hat{\sigma}_j^{AE}}{\hat{\sigma}_j^{ML}}$	1.043	1.040	1.023	1.077	1.053	1.049	1.133	1.063	1.000
t_9	$RE_j^{QML/ML}=\frac{\hat{\sigma}_j^{QML}}{\hat{\sigma}_j^{ML}}$	1.130	1.131	1.049	1.074	1.025	1.025	1.160	1.219	1.125
	$RE_j^{QML/AE}=\frac{\hat{\sigma}_j^{QML}}{\hat{\sigma}_j^{AE}}$	1.106	1.117	1.047	1.074	1.075	1.122	1.115	1.114	1.125
	$RE_j^{AE/ML}=\frac{\hat{\sigma}_j^{AE}}{\hat{\sigma}_j^{ML}}$	1.021	1.013	1.049	1.000	1.026	1.025	1.040	1.094	1.000
χ^2_6	$RE_j^{QML/ML}=\frac{\hat{\sigma}_j^{QML}}{\hat{\sigma}_j^{ML}}$	1.800	2.269	2.069	1.474	2.114	2.917	1.765	2.407	2.118
	$RE_j^{QML/AE}=\frac{\hat{\sigma}_j^{QML}}{\hat{\sigma}_j^{AE}}$	1.500	1.532	1.538	1.400	1.345	1.667	1.304	1.161	1.714
	$RE_j^{AE/ML}=\frac{\hat{\sigma}_j^{AE}}{\hat{\sigma}_j^{ML}}$	1.200	1.481	1.345	1.053	1.571	1.750	1.353	2.074	1.235
χ^2_{12}	$RE_j^{QML/ML}=\frac{\hat{\sigma}_j^{QML}}{\hat{\sigma}_j^{ML}}$	1.600	1.745	1.645	1.733	1.839	2.417	1.867	2.520	2.000
	$RE_j^{QML/AE}=\frac{\hat{\sigma}_j^{QML}}{\hat{\sigma}_j^{AE}}$	1.297	1.391	1.308	1.300	1.357	1.450	1.120	1.500	1.429
	$RE_j^{AE/ML}=\frac{\hat{\sigma}_j^{AE}}{\hat{\sigma}_j^{ML}}$	1.233	1.254	1.258	1.333	1.355	1.667	1.667	1.680	1.400

Notes: the relative efficiencies (REs) of the MLE, QMLE and the semiparametric estimator are obtained by dividing standard deviations of parameter j obtained with different estimators. The standard deviations for each parameter (using the MLE, QMLE and the semiparametric estimator) are in Table 5.