Fourier–type estimation of the power garch model with stable–paretian innovations

Francq, Christian and Meintanis, Simos

1 October 2012

Fourier–type estimation of the power garch model with stable–paretian innovations

CHRISTIAN FRANCQ[a], SIMOS G. MEINTANIS[b,c]

[a] *CREST and University Lille 3 (EQUIPPE), BP 60149, 59653 Villeneuve d'Ascq cedex, France*

[b] *Department of Economics, National and Kapodistrian University of Athens, Athens, Greece*

[c] *Centre for Business Mathematics and Informatics, North–West University, Potchefstroom, South Africa*

Abstract. We consider estimation for general power GARCH models under stable–Paretian innovations. Exploiting the simple structure of the conditional characteristic function of the observations driven by these models we propose minimum distance estimation based on the empirical characteristic function of corresponding residuals. Consistency of the estimators is proved, and we obtain a singular asymptotic distribution which is concentrated on a hyperplane. Efficiency issues are explored and finite–sample results are presented as well as applications of the proposed procedures to real data from the financial markets. A multivariate extension is also considered.

Keywords. GARCH model; Minimum distance estimation; Heavy–tailed distribution; Empirical characteristic function.

1 Introduction

Consider observations y_t from the so–called power GARCH model defined by

$$(1.1) \quad \begin{cases} y_t = c_t \varepsilon_t \\ c_t^\rho = \mu + \sum_{j=1}^p b_j |y_{t-j}|^\rho + \sum_{j=1}^q \gamma_j c_{t-j}^\rho, \quad \forall t = 1, 2, ..., \end{cases}$$

where $\{\varepsilon_t\}$ is a sequence of $i.i.d.$ random variables (with location zero and unit scale) independent of $\{c_t\}$, and $\rho, \mu, \{b_j\}_{j=1}^p$, and $\{\gamma_j\}_{j=1}^q$ denote unknown parameters. If the innovations $\{\varepsilon_t\}$ are standard normal and the power parameter ρ is set equal to two we obtain the classical Gaussian GARCH model. From the time of Mandelbrot (1963) and Fama (1965) however there is strong evidence that the distribution of financial returns could be heavy–tailed and possibly asymmetric, and many authors advocated the use of the stable–Paretian (SP) distribution instead of the normal distribution in financial modelling. For more recent evidence of stable–Paretian behavior of financial assets the reader is referred to the papers of Mittnik and Rachev (1993), Koutrouvelis and Meintanis (1999), Liu and Brorsen (1995a), Paolella (2001), Tsionas (2002), Akgül and Sayyan (2008), Tavares et al. (2008), Curto et al. (2009), and Xu et al. (2011), and the volumes by Adler et al. (1998), Rachev and Mittnik (2000), Rachev (2003) and Nolan, 2012).

Therefore one of the popular generalizations of model (1.1) is to assume that $\{\varepsilon_t\}$ follow a stable-Paretian distribution. We shall call this model SP power GARCH (SP–PGARCH) model. The most convenient way to introduce SP distributions is by means of their characteristic function (CF). Specifically if we assume that ε_t are zero–location SP random variables with unit scale, then their CF is given by

$$\text{(1.2)} \quad \varphi_\varepsilon(u) = e^{-|u|^\alpha\{1-\mathrm{i}\beta \operatorname{sgn}(u)\tan(\pi\alpha/2)\}}, \ \alpha \neq 1,$$

$$= e^{-|u|\{1+\mathrm{i}\beta\frac{2}{\pi}\operatorname{sgn}(u)\log|u|\}}, \quad \alpha = 1,$$

where $0 < \alpha \leq 2$, $-1 \leq \beta \leq 1$, and $\operatorname{sgn}(u) = 1, u > 0$, $\operatorname{sgn}(0) = 0$, and $\operatorname{sgn}(u) = -1, u < 0$. Note that α is a shape parameter often referred to as the 'tail index' and that the SP law reduces to the Gaussian distribution at $\alpha = 2$. On the other hand β measures skewness, and if $\beta = 0$ the corresponding SP law is symmetric. Besides the normal law, well known particular cases are the Cauchy distribution for $(\alpha, \beta) = (1, 0)$ and the Lévy distribution which corresponds to $(\alpha, \beta) = (1/2, 1)$. Several authors (see for instance Mittnik et al. 1999, Liu and Brorsen 1995b, and Bonato 2009) proposed maximum likelihood estimation of the SP–PGARCH model. However since the density of the SP law is generally not available in closed form various approximations are needed, and therefore likelihood methods may be characterized as computationally demanding.

In this paper we capitalize on the simplicity of eq. (1.2) and suggest to employ the

CF in order to estimate the parameters of SP–PGARCH models. The remainder of the paper is outlined as follows. In Section 2 we introduce the new estimation procedure. Section 3 is devoted to the asymptotic properties of the proposed method. The estimator is consistent under mild regularity conditions but, surprisingly, its asymptotic distribution is non standard, with a degenerate support concentrated on a hyperplane. Interestingly, the estimator of the SP parameter has a standard asymptotic distribution and enjoys an adaptiveness property with respect to the GARCH parameters. Optimality issues are also considered. The results of a Monte Carlo study for the finite–sample properties of the method are presented in Section 4. In Section 5 we consider an extension of the estimation procedure to multivariate SP–PGARCH models, while in Section 6 empirical applications are presented. Finally, we end in Section 7 with conclusions and discussion. An Appendix contains parts of the proofs.

2 CF estimation of the SP–PGARCH model

Consider the SP–PGARCH model whereby the observations y_t, $(t = 1, ..., T)$, are driven by equation (1.1) and the innovations ε_t have CF given by (1.2). We assume the standard positivity conditions $\mu > 0$, $\{b_j \geq 0,\ 1 \leq j \leq p\}$ and $\{\gamma_j \geq 0,\ 1 \leq j \leq q\}$.

Denote by $\boldsymbol{\theta} = (\rho, \mu, b_1, \ldots, b_p, \gamma_1, \ldots, \gamma_q)'$ the PGARCH parameter and by $\boldsymbol{\lambda} = (\alpha, \beta)'$ the SP parameter. We suggest to estimate the parameter $\boldsymbol{\vartheta} = (\boldsymbol{\theta}', \boldsymbol{\lambda}')')$, by minimum distance between the CF and a suitable empirical counterpart. Specifically, given the observations $(y_1, ..., y_T)$ and fixed initial values $(y_0, ..., y_{1-p})$ and $(\widetilde{c}_0, ..., \widetilde{c}_{1-q})$, the estimation method is defined as

(2.1) $$\widehat{\boldsymbol{\vartheta}}_T = \arg\min_{\boldsymbol{\vartheta} \in \Xi} \widetilde{\Delta}_T(\boldsymbol{\vartheta}),$$

where Ξ denotes the parameter space and

(2.2) $$\widetilde{\Delta}_T(\boldsymbol{\vartheta}) = \int_{-\infty}^{\infty} |\widetilde{\varphi}_T(u) - \varphi_\varepsilon(u)|^2 W(u) du,$$

with $W(\cdot)$ being a nonnegative weight function. In (2.2) $\widetilde{\varphi}_T(u) := \varphi_T(u; \widetilde{\varepsilon}_1, ..., \widetilde{\varepsilon}_T)$ is the empirical CF (ECF) defined by

(2.3) $$\varphi_T(u; x_1, ..., x_T) = \frac{1}{T} \sum_{t=1}^{T} e^{\mathrm{i}ux_t},$$

and computed from the residuals $\widetilde{\varepsilon}_t = y_t/\widetilde{c}_t$, with $\widetilde{c}_t$ being recursively defined for $t \geq 1$, by

$$(2.4) \quad \widetilde{c}_t^{\rho} = \mu + \sum_{j=1}^{p} \beta_j |y_{t-j}|^{\rho} + \sum_{j=1}^{q} \gamma_j \widetilde{c}_{t-j}^{\rho}.$$

Note that the introduction of the weight function $W(\cdot)$ in (2.2) is necessary in order to neutralize the periodic components in the ECF $\widetilde{\varphi}_T(u)$ and thus render the corresponding integral finite. For the moment we shall only assume that $W(\cdot)$ is symmetric, i.e., $W(u) = W(-u)$, and impose further conditions as they occur.

Estimation methods defined by (2.1) date back to Heathcote (1977), Thornton and Paulson (1997) and Bryant and Paulson (1979), for i.i.d. data. There is also work on ECF–based estimation for dependent data, but in a context different from the present one. The interested reader is referred to Kotchoni (2012), Carrasco et al. (2007), and Feuerverger (1990), and references therein.

Note that the ECF, $\widetilde{\varphi}_T(u)$, involves the PGARCH parameter $\boldsymbol{\theta}$ and the CF, $\varphi_{\varepsilon}(u)$, involves the SP parameter $\boldsymbol{\lambda}$. In the existing literature, the unknown parameter is only involved in the CF, the ECF being computed directly from the observations $y_1, \ldots, y_T$. The reason why we can not use the standard approach in our framework is that, for a PGARCH model, there exists no closed form for the CF of a vector of the form $(y_t, \ldots, y_{t-h})$, $h \geq 0$.

3 Asymptotic properties

Now consider the asymptotic properties of the estimator (2.1) of the parameter of the SP–PGARCH(p, q) model (1.1)-(1.2). Recall that the parameter vector is decomposed as $\boldsymbol{\vartheta} = (\boldsymbol{\theta}', \boldsymbol{\lambda}')'$ with $\boldsymbol{\theta} = (\rho, \mu, b_1, \ldots, b_p, \gamma_1, \ldots, \gamma_q)' \in \Theta$ and $\boldsymbol{\lambda} = (\alpha, \beta)' \in \Lambda$. The true parameter value is denoted by $\boldsymbol{\vartheta}_0 = (\boldsymbol{\theta}_0', \boldsymbol{\lambda}_0')'$ with $\boldsymbol{\theta}_0' = (\rho_0, \mu_0, b_{01}, \ldots, b_{0p}, \gamma_{01}, \ldots, \gamma_{0q})$ and $\boldsymbol{\lambda}_0' = (\alpha_0, \beta_0)$. Following the seminal paper of Bougerol and Picard (1992), it is easy to see that the necessary and sufficient condition for strict stationarity of (1.1) takes the form $\gamma(\boldsymbol{\theta}_0) < 0$, where $\gamma(\boldsymbol{\theta}_0)$ is the top-Lyapounov exponent of the model, as defined in Appendix A of Hamadeh and Zakoïan (2011).

Let $\mathcal{B}_{\boldsymbol{\theta}}(z) = \sum_{j=1}^{p} b_j z^j$ and $\mathcal{G}_{\boldsymbol{\theta}}(z) = 1 - \sum_{j=1}^{q} \gamma_j z^j$. By convention, $\mathcal{B}_{\boldsymbol{\theta}}(z) = 0$ if

$p = 0$ and $\mathcal{G}_{\boldsymbol{\theta}}(z) = 1$ if $q = 0$. To show the strong consistency, the following assumptions will be made.

A1: $\boldsymbol{\vartheta}_0 \in \Xi := \Theta \times \Lambda$ where Θ is a compact subset of $(0, \infty)^2 \times [0, \infty)^{p+q}$ and Λ is a compact subset of $(1, 2) \times [-1, 1]$.

A2: $\gamma(\boldsymbol{\theta}_0) < 0$ and $\forall \boldsymbol{\theta} \in \Theta$, $\quad \sum_{j=1}^{q} \gamma_j < 1$.

A3(j): $W(\cdot)$ is strictly positive over $\mathbb{R} \setminus \{0\}$, with $\int |u|^j W(u)du < \infty$.

A4: if $q > 0$, $\mathcal{B}_{\boldsymbol{\theta}_0}(z)$ and $\mathcal{G}_{\boldsymbol{\theta}_0}(z)$ have no common root, $\mathcal{B}_{\boldsymbol{\theta}_0}(1) \neq 0$, and $b_{0p} + \gamma_{0q} \neq 0$.

Assumption **A1** imposes standard positivity constraints on the PGARCH coefficients. It also puts restrictions on the value of the tail index α_0. DuMouchel (1983) showed that, in the case of a sample of stable distribution, the asymptotic distribution of the MLE is not standard when $\alpha_0 = 2$. Note also that when $\alpha_0 = 2$, the coefficient β_0 is not identifiable. We impose $\alpha_0 > 1$ because we need $E|\varepsilon_t| < \infty$. Assumption **A2** and the identifiability assumption **A4** are also required for the consistency of the QMLE of GARCH models.

It will be convenient to approximate the sequence $(\widetilde{c}_t)$ defined by (2.4) by an ergodic stationary sequence. Note that, under **A2**, the roots of $\mathcal{G}_{\boldsymbol{\theta}}(z)$ are outside the unit disk. Therefore, denote by $(c_t) = \{c_t(\boldsymbol{\theta})\}$ the strictly stationary, ergodic and nonanticipative solution of

$$(3.1) \quad c_t^{\rho} = \mu + \sum_{j=1}^{p} b_j |y_{t-j}|^{\rho} + \sum_{j=1}^{q} \gamma_j c_{t-j}^{\rho}, \qquad \forall t.$$

To show the identifiability of the conditional characteristic function (see Remark A.2), the following assumption will be needed.

A5: $E \sup_{\boldsymbol{\theta} \in \Theta} \frac{c_1(\boldsymbol{\theta}_0)}{c_1(\boldsymbol{\theta})}^{10} < \infty$.

For an ARCH(p) model, we have

$$\frac{c_t(\boldsymbol{\theta}_0)}{c_t(\boldsymbol{\theta})} = \left(\frac{\mu_0 + \sum_{j=1}^{p} b_{0j} y_{t-j}^{\rho}}{\mu + \sum_{j=1}^{p} b_j y_{t-j}^{\rho}} \right)^{1/\rho} \leq \left(\frac{\mu_0}{\mu} + \sum_{j=1}^{p} \frac{b_{0j}}{b_j} \right)^{1/\rho}.$$

Therefore, in the ARCH case, Assumption **A5** is satisfied when $\inf_{\theta\in\Theta}\min b_j > 0$. In the general case, it can be shown that Assumption **A5** is satisfied when Θ is sufficiently small (see (5.15) and (5.16) in Hamadeh and Zakoïan (2011), referred to as HZ hereafter).

Let

$$\Delta_T(\boldsymbol{\vartheta}) = \int_{-\infty}^{\infty} |\varphi_T(u) - \varphi_\varepsilon(u)|^2 W(u)du,$$

where $\varphi_T(u) = \varphi_T(u;\varepsilon_1,...,\varepsilon_T)$ is the ECF in (2.4) computed from the standardized innovations $\varepsilon_t = \varepsilon_t(\boldsymbol{\theta}) = y_t/c_t$. Note that $\Delta_T(\boldsymbol{\vartheta})$ is well defined under **A3**(0) because $\sup_u |\varphi_T(u) - \varphi_\varepsilon(u)|^2 < \infty$. We are now in a position to state our first result.

Theorem 3.1 *Let $(\widehat{\boldsymbol{\vartheta}}_T)$ be a sequence of CF estimators satisfying (2.1). Under the regularity conditions* **A1**, **A2**, **A3***(0),* **A3***(1),* **A4** *and* **A5**, *almost surely* $\widehat{\boldsymbol{\vartheta}}_T \to \boldsymbol{\vartheta}_0$, *as $T \to \infty$.*

Let K and ϱ be generic constants, whose values will be modified along the proofs, such that $K > 0$ and $0 < \varrho < 1$.

Proof of Theorem 3.1. In the appendix, we show the asymptotic irrelevance of the initial values by proving that

$$(3.2) \qquad \lim_{T\to\infty} \sup_{\boldsymbol{\vartheta}\in\Xi} |\Delta_T(\boldsymbol{\vartheta}) - \widetilde{\Delta}_T(\boldsymbol{\vartheta})| = 0, \quad a.s.$$

Let us show that the limiting criterion is minimal only at the true value, that is

$$(3.3) \quad \forall\boldsymbol{\vartheta} \neq \boldsymbol{\vartheta}_0, \quad \lim_{T\to\infty} \Delta_T(\boldsymbol{\vartheta}) > 0 \qquad \text{and} \qquad \lim_{T\to\infty} \Delta_T(\boldsymbol{\vartheta}_0) = 0 \quad a.s.$$

We will write

$$\varphi_T(u;\boldsymbol{\theta}) = \varphi_T(u) \qquad \text{and} \qquad \varphi(u;\boldsymbol{\lambda}) = \varphi_\varepsilon(u)$$

when it is necessary to emphasize that the empirical CF depends on $\boldsymbol{\theta}$ and the theoretical CF depends on $\boldsymbol{\lambda}$. The ergodic theorem shows that, almost surely,

$$\varphi_T(u,\boldsymbol{\theta}_0) \to Ee^{\mathrm{i}u\varepsilon_t(\boldsymbol{\theta}_0)} = \varphi(u,\boldsymbol{\lambda}_0) \quad \text{as} \quad T \to \infty.$$

The second convergence of (3.3) is thus a direct consequence of the dominated convergence theorem and **A3**(0). Using $W(\cdot) > 0$ and the continuity of the characteristic functions, the same arguments show that $\lim_{T\to\infty} \Delta_T(\boldsymbol{\vartheta}) = 0$ iff $\varphi(u, \boldsymbol{\lambda})$ is the characteristic function of

$$\varepsilon_1(\boldsymbol{\theta}) = \frac{c_1(\boldsymbol{\theta}_0)\varepsilon_1(\boldsymbol{\theta}_0)}{c_1(\boldsymbol{\theta})},$$

that is iff

$$\forall u, \qquad \varphi(u, \boldsymbol{\lambda}) = E\varphi\left(u\frac{c_1(\boldsymbol{\theta}_0)}{c_1(\boldsymbol{\theta})}, \boldsymbol{\lambda}_0\right).$$

Lemma A.1 of the appendix shows that, under **A5**, this is only possible if

$$\boldsymbol{\lambda} = \boldsymbol{\lambda}_0 \qquad \text{and} \qquad c_1(\boldsymbol{\theta}_0) = c_1(\boldsymbol{\theta}) \quad a.s.,$$

which is equivalent to $\boldsymbol{\vartheta} = \boldsymbol{\vartheta}_0$ by **A4** (see the proof of Theorem 3.1 in HZ). The proof of (3.3) is complete.

We need to show that the inequality in (3.3) holds locally uniformly, *i.e.* that there exists a neighborhood $V(\boldsymbol{\vartheta}^*)$ of $\boldsymbol{\vartheta}^* = (\boldsymbol{\theta}^{*'}, \boldsymbol{\lambda}^{*'})'$ such that

$$(3.4) \quad \liminf_{T\to\infty} \inf_{\boldsymbol{\vartheta}\in V(\boldsymbol{\vartheta}^*)} \Delta_T(\boldsymbol{\vartheta}) > 0 \qquad \text{if} \quad \boldsymbol{\vartheta}^* \neq \boldsymbol{\vartheta}_0.$$

Let $Ee^{iu\varepsilon_1}$ be the almost sure limit of $\varphi_T(u, \boldsymbol{\theta})$. Lemma A.3 shows that the convergence is actually uniform:

$$(3.5) \quad \forall u, \quad \sup_{\boldsymbol{\theta}\in\Theta} \left|\varphi_T(u, \boldsymbol{\theta}) - Ee^{\mathrm{i}u\varepsilon_1}\right| \to 0 \quad a.s.$$

Now note that

$$\Delta_T(\boldsymbol{\vartheta}) = a_T(\boldsymbol{\theta}) + b(\boldsymbol{\vartheta}) + d_T(\boldsymbol{\vartheta})$$

with

$$\begin{aligned}
a_T(\boldsymbol{\theta}) &= \int_{-\infty}^{+\infty} \left|\varphi_T(u, \boldsymbol{\theta}) - Ee^{\mathrm{i}u\varepsilon_1}\right|^2 W(u)du, \\
b(\boldsymbol{\vartheta}) &= \int_{-\infty}^{+\infty} \left|Ee^{\mathrm{i}u\varepsilon_1} - \varphi(u, \boldsymbol{\lambda})\right|^2 W(u)du, \\
d_T(\boldsymbol{\vartheta}) &= \int_{-\infty}^{+\infty} 2\mathrm{Re}\left\{\varphi_T(u, \boldsymbol{\theta}) - Ee^{\mathrm{i}u\varepsilon_1}\right\}\left\{Ee^{-\mathrm{i}u\varepsilon_1} - \overline{\varphi(u, \boldsymbol{\lambda})}\right\} W(u)du.
\end{aligned}$$

Using (3.5), **A3**(0) and the bound $\left|\varphi_T(u, \boldsymbol{\theta}) - Ee^{\mathrm{i}u\varepsilon_1}\right| \leq 2$, we show that $\sup_{\boldsymbol{\theta}\in\Theta} a_T(\boldsymbol{\theta}) \to 0$ *a.s.* By the Cauchy-Schwarz inequality, we similarly obtain $\sup_{\boldsymbol{\vartheta}\in\Xi} d_T(\boldsymbol{\vartheta}) \to 0$ *a.s.*

For any positive integer k, let $V_k(\boldsymbol{\vartheta}^*)$ be the open ball of center $\boldsymbol{\vartheta}^*$ and radius $1/k$. By Beppo Levi, as $k \to \infty$

$$\begin{aligned}\inf_{\boldsymbol{\vartheta}\in V_k(\boldsymbol{\vartheta}^*)} b(\boldsymbol{\vartheta}) &\geq \int_{-\infty}^{+\infty} \inf_{\boldsymbol{\vartheta}\in V_k(\boldsymbol{\vartheta}^*)} \left| Ee^{\mathrm{i}u\varepsilon_1} - \varphi(u,\boldsymbol{\lambda})\right|^2 W(u)du \\ &\uparrow \int_{-\infty}^{+\infty} \left| Ee^{\mathrm{i}u\varepsilon_1(\boldsymbol{\theta}^*)} - \varphi(u,\boldsymbol{\lambda}^*)\right|^2 W(u)du = \lim_{T\to\infty} \Delta_T(\boldsymbol{\vartheta}^*) > 0,\end{aligned}$$

where the last inequality comes from (3.3). It follows that there exists a neighborhood $V(\boldsymbol{\vartheta}^*)$ such that $\inf_{\boldsymbol{\vartheta}\in V(\boldsymbol{\vartheta}^*)} b(\boldsymbol{\vartheta}) > 0$. We then obtain (3.4) by noting that

$$\inf_{\boldsymbol{\vartheta}\in V(\boldsymbol{\vartheta}^*)} \Delta_T(\boldsymbol{\vartheta}) \geq \inf_{\boldsymbol{\vartheta}\in V(\boldsymbol{\vartheta}^*)} b(\boldsymbol{\vartheta}) - \sup_{\boldsymbol{\theta}\in\Theta} a_T(\boldsymbol{\theta}) - \sup_{\boldsymbol{\vartheta}\in\Theta} d_T(\boldsymbol{\vartheta}).$$

The conclusion follows from (3.2), (3.3) and (3.4) and a standard compactness argument. □

To show the asymptotic normality, the following additional assumption is needed.

A6: $\boldsymbol{\vartheta}_0 \in \overset{\circ}{\Xi}$, where $\overset{\circ}{\Xi}$ denotes the interior of Ξ.

We also need to introduce few additional notations. Let $g_t(u,\boldsymbol{\vartheta}) = e^{\mathrm{i}u\varepsilon_t} - \varphi_\varepsilon(u)$ and the vector of dimension $d = p + q + 3$

$$\boldsymbol{\Upsilon}_t = \int_{-\infty}^{\infty} \mathrm{Re}\left(\overline{g_t(u,\boldsymbol{\vartheta}_0)} E\frac{\partial g_1(u,\boldsymbol{\vartheta}_0)}{\partial\boldsymbol{\vartheta}}\right) W(u)du.$$

The next lemma, whose proof is in the appendix, shows the existence of $\mathbf{V} = \mathrm{Var}\,\boldsymbol{\Upsilon}_1$, as well as the existence of the matrix

$$\mathbf{G} = \int_{-\infty}^{\infty} \mathrm{Re}\left(E\overline{\frac{\partial g_1(u,\boldsymbol{\vartheta}_0)}{\partial\boldsymbol{\vartheta}}} E\frac{\partial g_1(u,\boldsymbol{\vartheta}_0)}{\partial\boldsymbol{\vartheta}'}\right) W(u)du.$$

Lemma 3.2 *Under the assumptions of Theorem 3.1, the matrices* $\mathbf{G}$ *and* $\mathbf{V}$ *are well defined and are singular.*

Proof. We have

$$\text{(3.6)} \qquad \frac{\partial g_t(u,\boldsymbol{\vartheta})}{\partial\boldsymbol{\theta}} = \frac{\partial e^{\mathrm{i}u\varepsilon_t}}{\partial\boldsymbol{\theta}} = -\mathrm{i}ue^{\mathrm{i}u\varepsilon_t}\varepsilon_t\frac{1}{c_t}\frac{\partial c_t}{\partial\boldsymbol{\theta}},$$

and

$$\text{(3.7)} \qquad \frac{\partial g_t(u,\boldsymbol{\vartheta}}{\partial\boldsymbol{\lambda}}) = -\frac{\partial\varphi(u,\boldsymbol{\lambda})}{\partial\boldsymbol{\lambda}} = \varphi(u,\boldsymbol{\lambda})\begin{pmatrix}\tau_1(u)\\ \tau_2(u)\end{pmatrix},$$

with

$$\tau_1(u) = |u|^{\alpha_0}\left\{\log|u| - \mathrm{i}\beta_0 \mathrm{sgn}\,(u)\left(\log|u|\tan\left(\frac{\pi\alpha_0}{2}\right) + \frac{\pi}{2}\frac{1}{\cos^2\left(\frac{\pi\alpha_0}{2}\right)}\right)\right\},$$
$$\tau_2(u) = -\mathrm{i}|u|^{\alpha_0}\mathrm{sgn}(u)\tan\left(\frac{\pi\alpha_0}{2}\right).$$

Since $E\,|\varepsilon_t(\boldsymbol{\theta}_0)| < \infty$, and $\varepsilon_t(\boldsymbol{\theta}_0)$ and $c_t(\boldsymbol{\theta})$ are independent, we have

$$E\left\|\frac{\partial g_t(u,\boldsymbol{\vartheta})}{\partial\boldsymbol{\theta}}\right\| \le E\left\|u\frac{c_t(\boldsymbol{\theta}_0)}{c_t(\boldsymbol{\theta})}\varepsilon_t(\boldsymbol{\theta}_0)\frac{1}{c_t}\frac{\partial c_t(\boldsymbol{\theta})}{\partial\boldsymbol{\theta}}\right\| \le K|u|E\frac{c_t(\boldsymbol{\theta}_0)}{c_t(\boldsymbol{\theta})}\left\|\frac{1}{c_t}\frac{\partial c_t(\boldsymbol{\theta})}{\partial\boldsymbol{\theta}}\right\|.$$

It can be shown (see (5.16) in HZ) that for all $r > 0$ there exists a neighborhood $V(\boldsymbol{\theta}_0)$ of $\boldsymbol{\theta}_0$ such that

$$(3.8)\quad E\sup_{\boldsymbol{\theta}\in V(\boldsymbol{\theta}_0)}\left(\frac{c_t(\boldsymbol{\theta}_0)}{c_t(\boldsymbol{\theta})}\right)^r < \infty.$$

In view of (5.20) in HZ and its extension Page 506, we also have

$$(3.9)\quad E\sup_{\theta\in\Theta}\left\|\frac{1}{c_t}\frac{\partial c_t(\boldsymbol{\theta})}{\partial\boldsymbol{\theta}}\right\|^r < \infty \quad \text{for all } r > 0.$$

The Hölder inequality then entails that for some neighborhood $V(\boldsymbol{\vartheta}_0)$ of $\boldsymbol{\vartheta}_0$, we have

$$(3.10)\quad E\sup_{\boldsymbol{\vartheta}\in V(\boldsymbol{\vartheta}_0)}\left\|\frac{\partial g_1(u,\boldsymbol{\vartheta})}{\partial\boldsymbol{\vartheta}}\right\| < \infty,$$

where the norm of a complex vector denotes the sum of the norms of its real and imaginary parts. By Lebesgue's dominated convergence theorem, we thus have

$$E\frac{\partial g_1(u,\boldsymbol{\vartheta})}{\partial\boldsymbol{\theta}} = \frac{\partial E e^{\mathrm{i}u\varepsilon_1}}{\partial\boldsymbol{\theta}} = \frac{\partial}{\partial\boldsymbol{\theta}}E\varphi\left(\frac{c_t(\theta_0)}{c_t(\theta)}u,\boldsymbol{\lambda}_0\right) = -uE\varphi'\left(\frac{c_t(\theta_0)}{c_t(\theta)}u\right)\frac{c_t(\theta_0)}{c_t(\theta)}\frac{1}{c_t(\theta)}\frac{\partial c_t}{\partial\boldsymbol{\theta}},$$

where

$$\varphi'(u) = -\varphi(u,\boldsymbol{\lambda}_0)\alpha_0|u|^{\alpha_0-1}\left\{1 - \mathrm{i}\beta_0\mathrm{sgn}\,(u)\tan\left(\frac{\pi\alpha_0}{2}\right)\right\}.$$

We then have

$$(3.11)\quad E\frac{\partial g_1(u,\boldsymbol{\vartheta}_0)}{\partial\boldsymbol{\theta}} = \varphi(u,\boldsymbol{\lambda}_0)\boldsymbol{\tau}_3(u),$$

with

$$\boldsymbol{\tau}_3(u) = \tau_3(u)E\frac{1}{c_t}\frac{\partial c_t}{\partial\boldsymbol{\theta}}(\boldsymbol{\theta}_0), \quad \tau_3(u) = \mathrm{sgn}\,(u)\,\alpha_0|u|^{\alpha_0}\left\{1 - \mathrm{i}\beta_0\mathrm{sgn}\,(u)\tan\left(\frac{\pi\alpha_0}{2}\right)\right\}.$$

In view of (3.7) and (3.11), each component of $E\frac{\partial}{\partial\boldsymbol{\vartheta}}g_1(u,\boldsymbol{\vartheta}_0)$ is a bounded function of u (since $\varphi(u,\boldsymbol{\lambda}_0)$ tends to zero at an exponential rate as $|u|\to\infty$). The existence of $\mathbf{G}$ thus follows from **A3**(0). Since $|g_t(u,\boldsymbol{\vartheta}_0)|\le 2$, the existence of $\mathbf{V}$ also follows.

Let us show that $\mathbf{G}$ is singular. This is equivalent to prove that there exists $\boldsymbol{a}\neq\mathbf{0}$ such that $\boldsymbol{a}'E\frac{\partial g_1(u,\boldsymbol{\vartheta}_0)}{\partial\boldsymbol{\vartheta}}=0$ for all u (see Theorem 2 in Bryant and Paulson, 1979). Letting $\boldsymbol{a}=(\boldsymbol{a}_1',a_2,a_3)'$ with $\boldsymbol{a}_1\in\mathbb{R}^{p+q+2}$, and using (3.7) and (3.11), we have

$$\boldsymbol{a}'E\frac{\partial g_1(u,\boldsymbol{\vartheta}_0)}{\partial\boldsymbol{\vartheta}}=\varphi(u,\boldsymbol{\lambda}_0)\left\{\boldsymbol{a}_1'E\frac{1}{c_t}\frac{\partial c_t}{\partial\boldsymbol{\theta}}(\boldsymbol{\theta}_0)\tau_3(u)+a_2\tau_1(u)+a_3\tau_2(u)\right\}.$$

Since $|\varphi(u,\boldsymbol{\lambda}_0)|>0$ and since the functions $\tau_1(u)$, $\tau_2(u)$ and $\tau_3(u)$ are linearly independent, $\boldsymbol{a}'E\frac{\partial g_1(u,\boldsymbol{\vartheta}_0)}{\partial\boldsymbol{\vartheta}}=0$ for all u if and only if

$$a_2=a_3=0\quad\text{and}\quad\boldsymbol{a}_1'E\frac{1}{c_t}\frac{\partial c_t}{\partial\boldsymbol{\theta}}(\boldsymbol{\theta}_0)=0.$$

This can obviously be achieved by choosing $\boldsymbol{a}_1\neq\mathbf{0}$. Note that the rank of $\mathbf{G}$ is equal to 3. The singularity of $\mathbf{V}$ is shown similarly. □

We are now in a position to give the asymptotic distribution of the CF estimators.

Theorem 3.3 *Under the assumptions of Theorem 3.1,* **A3***(4) and* **A6***,* $\mathbf{G}\sqrt{T}(\widehat{\boldsymbol{\vartheta}}_T-\boldsymbol{\vartheta}_0)$ *converges in law to the* $\mathcal{N}(\mathbf{0},\mathbf{V})$ *distribution as* $T\to\infty$.

Proof of Theorem 3.3. In the appendix, it is shown that there exists a neighborhood $V(\boldsymbol{\vartheta}_0)$ of $\boldsymbol{\vartheta}_0$ such that

$$\text{(3.12)}\quad \lim_{T\to\infty}\sqrt{T}\sup_{\boldsymbol{\vartheta}\in V(\boldsymbol{\vartheta}_0)}\left\|\frac{\partial\Delta_T(\boldsymbol{\vartheta})}{\partial\boldsymbol{\vartheta}}-\frac{\partial\widetilde{\Delta}_T(\boldsymbol{\vartheta})}{\partial\boldsymbol{\vartheta}}\right\|=0,\quad a.s.$$

Assumption **A6** and the consistency of $\widehat{\boldsymbol{\vartheta}}_T$ entail that $\partial\widetilde{\Delta}_T(\widehat{\boldsymbol{\vartheta}}_T)/\partial\boldsymbol{\vartheta}=0$, at least for T large enough. In view of (3.12) and (A.3), we thus have

$$\begin{aligned}o_P(1)&=\sqrt{T}\frac{\partial\Delta_T(\widehat{\boldsymbol{\vartheta}}_T)}{\partial\boldsymbol{\vartheta}}\\ \text{(3.13)}\qquad &=2\int_{-\infty}^{+\infty}\frac{1}{\sqrt{T}}\sum_{t=1}^{T}\mathrm{Re}\left(\overline{g_t(u,\widehat{\boldsymbol{\vartheta}}_T)}\frac{1}{T}\sum_{s=1}^{T}\frac{\partial g_s(u,\widehat{\boldsymbol{\vartheta}}_T)}{\partial\boldsymbol{\vartheta}}\right)W(u)du.\end{aligned}$$

Using Taylor expansions, the ergodic theorem and the consistency of $\widehat{\boldsymbol{\vartheta}}_T$, it is shown in the appendix that

$$\frac{1}{\sqrt{T}}\sum_{t=1}^{T}\overline{g_t(u,\widehat{\boldsymbol{\vartheta}}_T)}=\frac{1}{\sqrt{T}}\sum_{t=1}^{T}\overline{g_t(u,\boldsymbol{\vartheta}_0)}+\left\{E\frac{\partial\overline{g_1(u,\boldsymbol{\vartheta}_0)}}{\partial\boldsymbol{\vartheta}}\right\}\sqrt{T}(\widehat{\boldsymbol{\vartheta}}_T-\boldsymbol{\vartheta}_0)+o_P(1)$$

and

$$\frac{1}{T}\sum_{s=1}^{T}\frac{\partial g_s(u,\widehat{\boldsymbol{\vartheta}}_T)}{\partial\boldsymbol{\vartheta}} = E\frac{\partial g_1(u,\boldsymbol{\vartheta}_0)}{\partial\boldsymbol{\vartheta}} + o_P(1).$$

Showing that the limit of (3.13) as $T\to\infty$ can be taken under the integral sign (see the appendix for details), we obtain

$$o_P(1) = \frac{1}{\sqrt{T}}\sum_{t=1}^{T}\Upsilon_t + \mathbf{G}\sqrt{T}(\widehat{\boldsymbol{\vartheta}}_T - \boldsymbol{\vartheta}_0). \tag{3.14}$$

The conclusion follows from the central limit theorem. □

Because the matrix $\mathbf{G}$ is singular, the previous theorem does not entail the asymptotic normality of the entire estimator $\sqrt{T}\left(\widehat{\boldsymbol{\vartheta}}_T - \boldsymbol{\vartheta}_0\right) = \sqrt{T}\left(\widehat{\boldsymbol{\theta}}'_T - \boldsymbol{\theta}'_0, \widehat{\boldsymbol{\lambda}}'_T - \boldsymbol{\lambda}'_0\right)'$. As a consequence of the following result, we have however asymptotic normality of $\sqrt{T}\left(\widehat{\boldsymbol{\lambda}}_T - \boldsymbol{\lambda}_0\right)$.

Theorem 3.4 *Under the assumptions of Theorem 3.3, we have* $\mathbf{G} = \boldsymbol{A}\boldsymbol{B}\boldsymbol{A}'$ *where*

$$\boldsymbol{A} = \begin{pmatrix} E\frac{1}{c_1}\frac{\partial c_1(\boldsymbol{\theta}_0)}{\partial\boldsymbol{\theta}} & \mathbf{0} \\ \mathbf{0} & \boldsymbol{I}_2 \end{pmatrix}$$

has full rank 3 and $\boldsymbol{B}$ *is an invertible* 3×3 *matrix. Moreover* $\sqrt{T}\boldsymbol{A}'(\widehat{\boldsymbol{\vartheta}}_T - \boldsymbol{\vartheta}_0)$ *converges in law to the* $\mathcal{N}\left(\mathbf{0}, \boldsymbol{\Sigma} := \boldsymbol{B}^{-1}\boldsymbol{U}\boldsymbol{B}^{-1}\right)$ *distribution, where* $\boldsymbol{B}$ *and* $\boldsymbol{U}$ *explicitly depend on* $\boldsymbol{\lambda}_0$ *and* $W(\cdot)$ *(see (3.15) and (3.16) below), but not on* $\boldsymbol{\theta}_0$.

Proof of Theorem 3.4. In view of (3.7) and (3.10), we have

$$E\frac{\partial}{\partial\boldsymbol{\vartheta}}g_1(u,\boldsymbol{\vartheta}_0) = \boldsymbol{A}\boldsymbol{\tau}(u), \qquad \boldsymbol{\tau}(u) = \varphi(u,\boldsymbol{\lambda}_0)\begin{pmatrix}\tau_3(u)\\ \tau_1(u)\\ \tau_2(u)\end{pmatrix}.$$

Since the 3 functions $\tau_i(u)$, $i=1,2,3$, are linearly independent, the matrix

$$\boldsymbol{B} := \mathrm{Re}\int\overline{\boldsymbol{\tau}(u)}\boldsymbol{\tau}(u)'W(u)du \tag{3.15}$$

is invertible (see Theorem 2 in Bryant and Paulson, 1979), and we have

$$\mathbf{G} := \mathrm{Re}\int E\frac{\partial\overline{g_1(u,\boldsymbol{\vartheta}_0)}}{\partial\boldsymbol{\vartheta}}E\frac{\partial g_1(u,\boldsymbol{\vartheta}_0)}{\partial\boldsymbol{\vartheta}'}W(u)du = \boldsymbol{A}\boldsymbol{B}\boldsymbol{A}'.$$

By Theorem 3.3 and the fact that $\boldsymbol{B}^{-1}(\boldsymbol{A}'\boldsymbol{A})^{-1}\boldsymbol{A}'\mathbf{G} = \boldsymbol{A}'$, we obtain the asymptotic normality with $\boldsymbol{\Sigma} = \boldsymbol{B}^{-1}(\boldsymbol{A}'\boldsymbol{A})^{-1}\boldsymbol{A}'\mathbf{V}\boldsymbol{A}(\boldsymbol{A}'\boldsymbol{A})^{-1}\boldsymbol{B}^{-1}$. Since

$$\boldsymbol{\Upsilon}_t = \boldsymbol{A}\int_{-\infty}^{\infty} \mathrm{Re}\left\{e^{\mathrm{i}u\epsilon_t(\boldsymbol{\theta}_0)}\overline{\boldsymbol{\tau}(u)}\right\} W(u)du - \boldsymbol{A}\int_{-\infty}^{\infty} \mathrm{Re}\left\{\varphi(u, \boldsymbol{\lambda}_0)\overline{\boldsymbol{\tau}(u)}\right\} W(u)du,$$

a computation similar to that of (17) in Bryant and Paulson (1979) gives $\mathbf{V} = \boldsymbol{A}\boldsymbol{U}\boldsymbol{A}'$ with

$$\begin{aligned}(3.16)\quad \boldsymbol{U} \;=\; & \frac{1}{2}\int_{-\infty}^{\infty}\int_{-\infty}^{\infty} \mathrm{Re}\left\{\varphi(u+v, \boldsymbol{\lambda}_0)\overline{\boldsymbol{\tau}(u)}\,\overline{\boldsymbol{\tau}(v)}'\right\} W(u)W(v)dudv \\ & +\frac{1}{2}\int_{-\infty}^{\infty}\int_{-\infty}^{\infty} \mathrm{Re}\left\{\varphi(u-v, \boldsymbol{\lambda}_0)\overline{\boldsymbol{\tau}(u)}\,\boldsymbol{\tau}(v)'\right\} W(u)W(v)dudv \\ & -\int_{-\infty}^{\infty} \mathrm{Re}\left\{\varphi(u, \boldsymbol{\lambda}_0)\overline{\boldsymbol{\tau}(u)}\right\} W(u)du \left(\int_{-\infty}^{\infty} \mathrm{Re}\left\{\varphi(v, \boldsymbol{\lambda}_0)\overline{\boldsymbol{\tau}(v)}\right\} W(v)dv\right)'.\end{aligned}$$

The conclusion follows. □

The following corollary of Theorem 3.4 shows that the matrices $\boldsymbol{B}$ and $\boldsymbol{\Sigma}$ have simple forms when $W(\cdot)$ is chosen to be even. This implies two interesting consequences for the ECF estimator of $\boldsymbol{\vartheta}_0$: i) adaptiveness of $\widehat{\boldsymbol{\lambda}}_T$ with respect to $\boldsymbol{\theta}_0$ and 2) singular asymptotic distribution for $\widehat{\boldsymbol{\theta}}_T$.

Corollary 3.5 Under the assumptions of Theorem 3.3, when $W(\cdot)$ is even the matrix $\boldsymbol{B}$ and $\boldsymbol{\Sigma}$ are of the form

$$\boldsymbol{B} = \begin{pmatrix} b_{11} & \mathbf{0} \\ \mathbf{0} & \boldsymbol{B}_{22} \end{pmatrix} \qquad \text{and} \qquad \boldsymbol{\Sigma} = \begin{pmatrix} 0 & \mathbf{0} \\ \mathbf{0} & \boldsymbol{S} \end{pmatrix},$$

where $\boldsymbol{S}$ is the asymptotic variance of the ECF estimator of $\boldsymbol{\lambda}_0$ based on an iid sequence (see Thornton and Paulson, 1977). The asymptotic distribution of $\sqrt{T}\left(\widehat{\boldsymbol{\theta}}_T - \boldsymbol{\theta}_0\right)$ is concentrated on the line

$$\Delta_c = \left\{\boldsymbol{x} \in \mathbb{R}^{p+q+2} : \boldsymbol{x}'E\frac{1}{c_t}\frac{\partial c_t}{\partial \boldsymbol{\theta}}(\boldsymbol{\theta}_0) = c\right\}$$

for some constant c.

Proof of Corollary 3.5. The block-diagonal form of $\boldsymbol{B}$ comes from the fact that, for $i = 1, 2$, $\mathrm{Re}\left\{\overline{\tau_3(u)}\tau_i(u)\right\}$ is an odd function of u. Since $\mathrm{Re}\left\{|\varphi(u, \boldsymbol{\lambda}_0)|^2\,\overline{\tau_3(u)}\right\}$ is an odd function, the first component of the vector $\int_{-\infty}^{\infty} \mathrm{Re}\left\{\varphi(u, \boldsymbol{\lambda}_0)\overline{\boldsymbol{\tau}(u)}\right\} W(u)du$ is equal to zero. Now note that $\tau_1(-v) = \overline{\tau_1(v)}$, $\tau_2(-v) = \overline{\tau_2(v)}$ and $\tau_3(-v) = -\overline{\tau_3(v)}$. It follows that for the two matrices defined by the double integrals of (3.16), the elements of the first row and first column are opposite. The form of Σ follows. Denoting by Var_{as} the variance of the asymptotic distribution, we thus have

$$\mathrm{Var}_{as}\left\{\sqrt{T}\left(\widehat{\boldsymbol{\theta}}_T - \boldsymbol{\theta}_0\right) E\frac{1}{c_t}\frac{\partial c_t}{\partial \boldsymbol{\theta}}(\boldsymbol{\theta}_0)\right\} = \boldsymbol{\Sigma}(1,1) = 0,$$

and the conclusion follows. $\square$

3.1 Efficient weight function

Theorem 3.4 shows that the asymptotic distribution of $\sqrt{T}\boldsymbol{A}'(\widehat{\boldsymbol{\vartheta}}_T - \boldsymbol{\vartheta}_0)$ is the same as that of $\sqrt{T}\boldsymbol{A}'(\breve{\boldsymbol{\vartheta}}_T - \boldsymbol{\vartheta}_0)$ where $\breve{\boldsymbol{\vartheta}}_T = \left(\breve{\boldsymbol{\theta}}', \breve{\boldsymbol{\lambda}}'\right)'$ is a consistent root of

$$\text{(3.17)} \quad \mathrm{Re}\int_{-\infty}^{\infty} \boldsymbol{\omega}(u)\left\{\varphi_T(u, \breve{\boldsymbol{\theta}}) - \varphi(u, \breve{\boldsymbol{\lambda}})\right\} du = \mathbf{0}, \qquad \boldsymbol{\omega}(u) = \overline{\boldsymbol{\tau}(u)}W(u).$$

Feuerverger and McDunnough (1981a, 1981b) studied in detail the asymptotic behavior of estimators satisfying the estimation equation (3.17) in a general framework. They found that, under general regularity conditions, the optimal weight function, allowing for an estimator virtually as efficient as the MLE, is of the form

$$\boldsymbol{\omega}(u) = \int_{-\infty}^{\infty} \frac{\partial \log f_{\boldsymbol{\vartheta}_0}(x)}{\partial \boldsymbol{\vartheta}} e^{-\mathrm{i}ux} dx.$$

In our framework, the regularity assumptions are not satisfied because, when $f_{\boldsymbol{\vartheta}_0}(x)$ is the SP density it is easy to see that the function $\frac{\partial \log f_{\boldsymbol{\vartheta}_0}(x)}{\partial \boldsymbol{\vartheta}} e^{-\mathrm{i}ux}$ is not integrable.

A solution to solve the problem consists in optimizing directly the asymptotic variance $\boldsymbol{S} = \boldsymbol{S}(W, \boldsymbol{\lambda}_0)$ defined in Theorem 3.4–Corollary 3.5 as a function of $W(\cdot)$, for a given value of $\boldsymbol{\lambda}_0$. Assume that the problem of interest is to minimize $\Psi(\boldsymbol{S})$ for some function $\Psi(\cdot)$ from the set of the 2×2 covariance matrices to $[0,\infty)$. The function $\Psi(\cdot)$ can be a norm or, if the focus is on the estimation of the tail index α_0, it can be $\Psi(\boldsymbol{S}) = \boldsymbol{S}(1,1)$. Let $\mathcal{W}$ be a set of functions $W(\cdot)$ satisfying **A3**(0) and **A3**(4), as requires in Theorems 3.3. Within this set, an optimal weight function is any measurable solution of

$$\text{(3.18)} \quad W^* = \arg\min_{W\in\mathcal{W}} \Psi\left\{\boldsymbol{S}(W, \boldsymbol{\lambda}_0)\right\}.$$

As an illustration, let us consider the Gamma-type weight functions defined by

(3.19) $\mathcal{W} = \{|u|^{\mathfrak{p}-1}e^{-\mathfrak{b}|u|} : \mathfrak{p} \geq 1, \mathfrak{b} > 0\}.$

Consider also the case $\boldsymbol{\lambda}_0 = (1.6, 0)$ with $\Psi(\boldsymbol{S}) = \boldsymbol{S}(1,1)$. The solution of (3.18) and the classical Gaussian weight function $W_\phi(u) := e^{-u^2}$ lead respectively to

$$\boldsymbol{S}(W^*, \boldsymbol{\lambda}_0) = \begin{pmatrix} 3.57 & 0.00 \\ 0.00 & 6.72 \end{pmatrix}, \quad \boldsymbol{S}(W_\phi, \boldsymbol{\lambda}_0) = \begin{pmatrix} 4.37 & 0.00 \\ 0.00 & 7.02 \end{pmatrix}.$$

One can see that the optimal weight function W^* leads asymptotically to a more accurate estimator of $\boldsymbol{\lambda}_0$ than the Gaussian weight function W_ϕ. Of course, one can expect some efficiency gain by optimizing over a larger set $\mathcal{W}$. Our numerical experiments lead us to think that the potential gain is modest, even at the price of a much more time-consuming optimization.

The optimal weight function W^* depends on the unknown parameter $\boldsymbol{\lambda}_0$. In the spirit of the optimal generalized method of moments (GMM) proposed by Hansen (1982), a standard solution consists in estimating $\boldsymbol{\lambda}_0$ in a first step by a suboptimal weight function, for instance W_ϕ, and replacing $\boldsymbol{\lambda}_0$ by the first-step estimate to solve (3.18) in a second step. Because we observed that the solution is not very sensitive to $\boldsymbol{\lambda}_0$, and because on the financial series that we have considered the estimated values of $\boldsymbol{\lambda}_0$ are often close to $(1.6, 0)$, we decided to keep the same weight function for all the forthcoming numerical illustrations. More precisely, we used the solution of (3.18)-(3.19), which turned out to be

$$W^*(u) = |u|^{\mathfrak{p}-1}e^{-\mathfrak{b}|u|} \qquad \text{with} \quad \mathfrak{p} = 1.69 \text{ and } \mathfrak{b} = 1.91.$$

4 Simulation results

The aim of our first simulation experiment is to illustrate that the ECF is consistent, but has a non standard asymptotic distribution concentrated on a line (see Corollary 3.5). We thus consider the following very simple version of the SP–PGARCH model

(4.1) $y_t = c_t\varepsilon_t, \quad c_t^2 = \mu_0 + b_0 y_{t-1}^2$

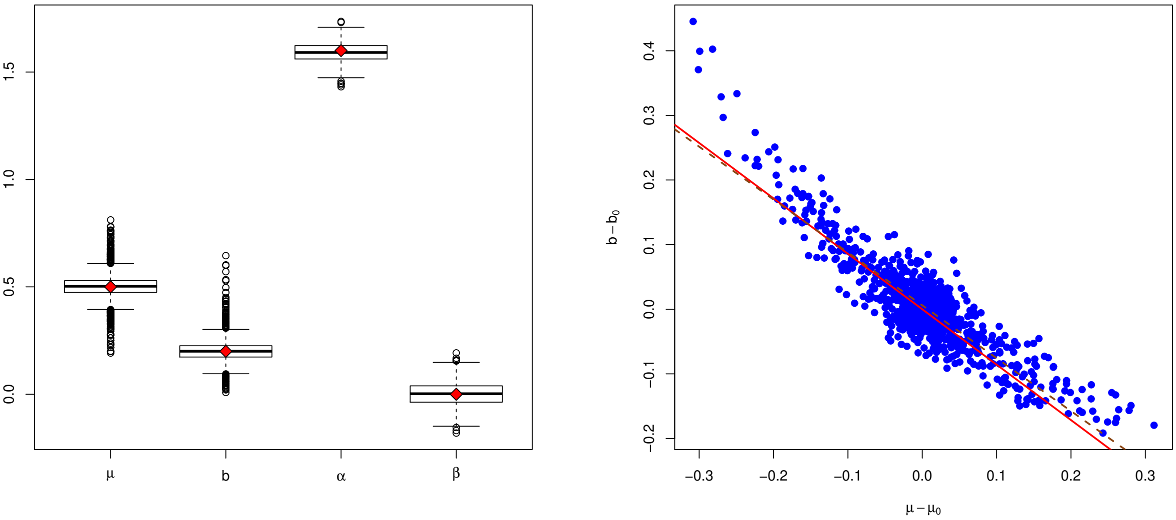

Figure 1: Empirical distribution of the ECF estimator over $1,000$ independent simulations of length $T = 2,000$ of the SP-PARCH(1) model (4.1). The red line of the scatter plot corresponds to the direction of the vector $Ec^{-1}\partial c_t(\boldsymbol{\theta}_0)/\partial\boldsymbol{\theta}$.

where $\boldsymbol{\vartheta}_0 = (0.5, 0.2, 1.6, 0)$, i.e., we have a SP-PARCH(1) model with symmetric SP innovations and tail index equal to 1.6. Moreover we assume that the value of ρ is known to be equal to 2. The left panel of Figure 1 displays the box-plots of the ECF estimates of the four parameters over $N = 1,000$ independent simulations of length $T = 2,000$ of the process. As expected, for each parameter, the median of the estimated values is very close to the true value (represented by a diamond symbol). The right panel of Figure 1 displays the scatter plot of the $1,000$ values of $(\hat{\mu} - \mu_0, \hat{b} - b_0)$. In accordance with Corollary 3.5, the points are concentrated along the red line, carried by the vector $Ec^{-1}\partial c_t(\boldsymbol{\theta}_0)/\partial\boldsymbol{\theta}$.[1] For comparison, we plotted the linear regression of $\hat{b} - b_0$ on $\hat{\mu} - \mu_0$ as a dotted line. This line is almost confused with the full red line. Figure 2 corresponds to simulations of length $T = 20,000$. Of course, the estimates are more accurate, and the points are more concentrated along the red line.

We now compare the ECF and ML estimators on the SP–PGARCH(1,1) model

$$(4.2) \qquad y_t = c_t\varepsilon_t, \quad c_t^2 = \mu_0 + b_0 y_{t-1}^2 + \gamma_0 c_{t-1}^2$$

[1]Because there exists no explicit form for this vector, it has been evaluated on the basis of a simulation of length $50,000$.

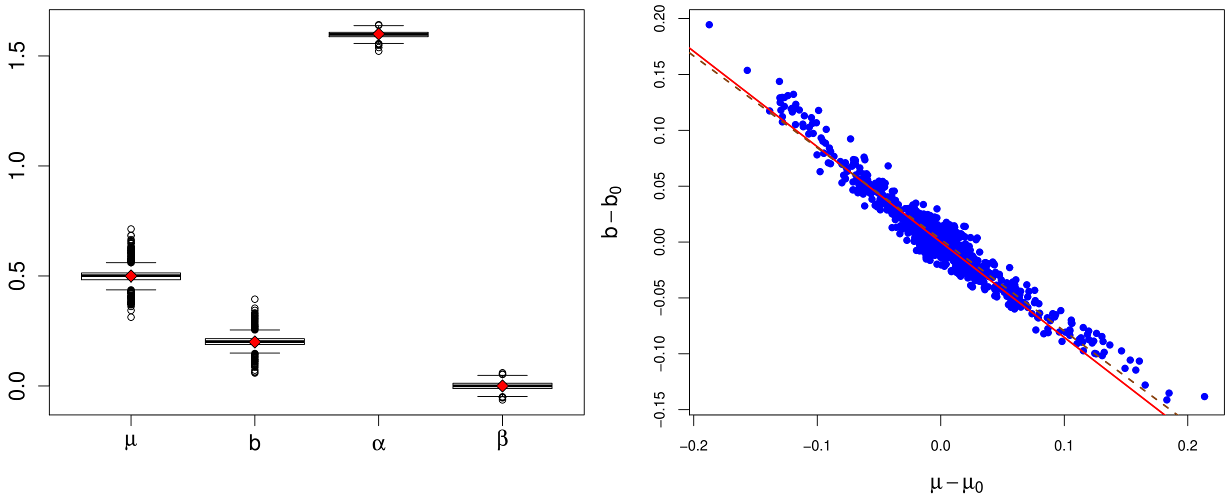

Figure 2: As Figure 1, but for the sample size $T = 20,000$.

where $\boldsymbol{\vartheta}_0 = (0.5, 0.05, 0.7, 1.6, 0)$. Table 1 shows that, for estimating the SP parameters α and β, the performance of ECF estimator seems similar to that of the MLE. For the estimation of the GARCH parameters, the MLE outperforms the ECF estimator, as expected from the asymptotic theory. Table 2 shows that, as expected, the ECF estimator is however much more advantageous than the MLE in terms of amount of computation time.

5 Multivariate extension

Consider now the problem of estimation of a multivariate SP–PGARCH model. We first specify that we will consider *i.i.d.* replications $\boldsymbol{\varepsilon}_t$, $t = 1, 2, ...,$ of the innovation vector $\boldsymbol{\varepsilon}$ of dimension $m \geq 1$, following a multivariate symmetric SP distribution with characteristic function

$$(5.1) \quad \varphi_{\boldsymbol{\varepsilon}}(\boldsymbol{u}) = e^{-\left|\frac{1}{2}\boldsymbol{u}'\boldsymbol{u}\right|^{\alpha/2}}, \ \boldsymbol{u} \in \mathbb{R}^m.$$

Such vectors have the convenient stochastic representation

$$(5.2) \quad \boldsymbol{\varepsilon} = A^{1/2}\boldsymbol{X},$$

with $\boldsymbol{X}$ distributed as multivariate standard normal, and A a totally skewed to the right SP random variable with shape parameter equal to $\alpha/2$, and independent of $\boldsymbol{X}$

Table 1: Comparison of the ECF and the MLE on 10 simulations of the SP–PGARCH(1,1) model (4.2).

	ECF estimator					MLE estimator				
Iter	$\widehat{\mu}$	$\widehat{b}$	$\widehat{\gamma}$	$\widehat{\alpha}$	$\widehat{\beta}$	$\widehat{\mu}$	$\widehat{b}$	$\widehat{\gamma}$	$\widehat{\alpha}$	$\widehat{\beta}$
1	0.39	0.04	0.75	1.67 (0.04)	0.08 (0.06)	0.42 (0.03)	0.04 (0.01)	0.73 (0.02)	1.64 (0.03)	0.04 (0.06)
2	0.45	0.05	0.72	1.57 (0.04)	-0.01 (0.06)	0.51 (0.07)	0.06 (0.01)	0.67 (0.03)	1.57 (0.03)	-0.01 (0.05)
3	0.68	0.04	0.68	1.65 (0.04)	0.07 (0.06)	0.56 (0.06)	0.06 (0.01)	0.67 (0.02)	1.63 (0.03)	0.06 (0.06)
4	0.26	0.02	0.83	1.47 (0.04)	0.01 (0.06)	0.45 (0.07)	0.05 (0.01)	0.70 (0.03)	1.52 (0.03)	0.01 (0.04)
5	0.61	0.04	0.72	1.62 (0.04)	0.06 (0.06)	0.64 (0.07)	0.05 (0.01)	0.70 (0.03)	1.63 (0.04)	0.03 (0.06)
6	0.82	0.10	0.52	1.65 (0.04)	0.09 (0.06)	0.58 (0.06)	0.06 (0.01)	0.67 (0.03)	1.67 (0.03)	0.11 (0.07)
7	0.54	0.04	0.72	1.56 (0.04)	0.06 (0.06)	0.59 (0.07)	0.04 (0.00)	0.71 (0.02)	1.57 (0.03)	0.08 (0.05)
8	0.37	0.03	0.77	1.56 (0.04)	-0.01 (0.06)	0.41 (0.04)	0.04 (0.00)	0.75 (0.02)	1.52 (0.02)	0.01 (0.04)
9	0.55	0.06	0.67	1.61 (0.04)	0.00 (0.06)	0.49 (0.06)	0.06 (0.00)	0.68 (0.02)	1.57 (0.03)	-0.03 (0.05)
10	0.60	0.03	0.72	1.55 (0.04)	0.03 (0.06)	0.47 (0.05)	0.04 (0.00)	0.72 (0.02)	1.59 (0.03)	0.00 (0.05)

Table 2: Computation time of the ECF and ML estimators for 10 simulations of length $T = 200$ and $T = 2,000$ of the SP–PGARCH(1,1) model (4.2) (the empirical standard deviations are given into brackets).

	$T = 200$	$T = 2,000$
ECF	7.9 (2.2)	9.4(4.9)
ML	123.7 (20.6)	1132.9 (126.2)

(Samorodnitsky and Taqqu, 1994, §2.5).

In view of the above we define the multivariate SP–PGARCH with observation vector $\boldsymbol{y}_t = (y_{1t}, ..., y_{mt})'$ as

$$\boldsymbol{y}_t = \mathbf{C}_t^{1/2}\boldsymbol{\varepsilon}_t, \tag{5.3}$$

where $\mathbf{C}_t^{1/2}$ is a $(m \times m)$ scale matrix which is assumed to be symmetric and positive definite, while the vectors $\boldsymbol{\varepsilon}_t$, $(t = 1, ..., T)$, have characteristic function given by (5.1). Following the lines of Section 2 we suggest to estimate the parameters of the multivariate SP–PGARCH model by minimizing the criterion

$$\widetilde{\Delta}_T(\boldsymbol{\vartheta}) = \int_{\mathbb{R}^m} \left|\widetilde{\varphi}_T(\boldsymbol{u}) - \varphi_\varepsilon(\boldsymbol{u})\right|^2 W(\boldsymbol{u})d\boldsymbol{u}, \tag{5.4}$$

where $\varphi_\varepsilon(\cdot)$ is given by (5.1), $\widetilde{\varphi}_T(\boldsymbol{u}) := \varphi_T(\boldsymbol{u}; \widetilde{\boldsymbol{\varepsilon}}_1, ..., \widetilde{\boldsymbol{\varepsilon}}_T)$ is the multivariate ECF defined by

$$(5.5) \quad \varphi_T(\boldsymbol{u}; \boldsymbol{x}_1, ..., \boldsymbol{x}_T) = \frac{1}{T}\sum_{t=1}^{T} e^{\mathrm{i}\boldsymbol{u}'\boldsymbol{x}_t},$$

and computed from the residuals $\widetilde{\boldsymbol{\varepsilon}}_t = \widetilde{\mathbf{C}}_t^{-1/2}\boldsymbol{y}_t$, with $\widetilde{\mathbf{C}}_t$ being a scale matrix depending on $\boldsymbol{\vartheta}$ and on past observations in a way that it will be specified below.

One specific instance of a multivariate GARCH model is the so–called constant conditional correlation (CCC)–GARCH specification whereby

$$(5.6) \quad \mathbf{C}_t = \mathbf{D}_t\mathbf{R}\mathbf{D}_t,$$

where $\mathbf{D}_t$ and $\mathbf{R}$ are $(m \times m)$ matrices with $\mathbf{R}$ being a correlation matrix, while $\mathbf{D}_t$ is related to the volatility vector $\boldsymbol{c}_t = (c_{1t}, ..., c_{mt})'$ by the equation $\mathbf{D}_t^2 = \mathrm{diag}(c_{1t}, ..., c_{mt})$. The aforementioned specification is adapted to the multivariate SP–PGARCH context by advocating a power GARCH volatility specification as

$$(5.7) \quad \boldsymbol{c}_t^{(\rho)} = \boldsymbol{b}_0 + \sum_{j=1}^{p}\mathbf{B}_j\boldsymbol{y}_{t-j}^{(\rho)} + \sum_{j=1}^{q}\boldsymbol{\Gamma}_j\boldsymbol{c}_{t-j}^{(\rho)},$$

with

$$\boldsymbol{c}_t^{(\rho)} = (c_{1t}^{\rho}, ..., c_{mt}^{\rho})', \qquad \boldsymbol{y}_t^{(\rho)} = (y_{1t}^{\rho}, ..., y_{mt}^{\rho})',$$

where the vector $\boldsymbol{b}_0$ is of dimension m and has positive elements, while the $(m \times m)$ matrices $\{\mathbf{B}_j\}_{j=1}^{p}$, and $\{\boldsymbol{\Gamma}_j\}_{j=1}^{q}$, are with non–negative elements. In this case the matrix $\widetilde{\mathbf{C}}_t$ involved in (5.4) is computed recursively based on the equations (5.6) and (5.7) and on initial values $(\boldsymbol{y}_0, ..., \boldsymbol{y}_{1-p})$ and $(\widetilde{\boldsymbol{c}}_0, ..., \widetilde{\boldsymbol{c}}_{1-q})$. We shall call this model CCC– SP–PGARCH model

6 Applications to exchange rates

We now consider daily returns of 19 exchange rates with respect to the Euro. The currencies that we have considered are the American Dollar (USD), the Japanese Yen (JPY), the Czech Koruna (CZK), the Danish Krone (DKK), the British Pound (BGP),

Table 3: Log-GARCH(1,1) models fitted by ECF estimator on daily returns of exchange rates. The estimated standard deviation are displayed into brackets.

	$\widehat{\mu}$	$\widehat{b}$	$\widehat{\gamma}$	$\widehat{\alpha}$	$\widehat{\beta}$		$\widehat{\mu}$	$\widehat{b}$	$\widehat{\gamma}$	$\widehat{\alpha}$	$\widehat{\beta}$
USD	0.00	0.02	0.96	1.88 (0.03)	-0.10 (0.04)	CHF	0.00	0.06	0.86	1.66 (0.03)	-0.16 (0.04)
JPY	0.02	0.01	0.89	1.72 (0.03)	-0.23 (0.04)	NOK	0.01	0.13	0.53	1.65 (0.03)	0.19 (0.04)
CZK	0.00	0.03	0.93	1.69 (0.03)	0.02 (0.04)	AUD	0.00	0.01	0.98	1.75 (0.03)	0.25 (0.04)
DKK	0.00	0.07	0.81	1.55 (0.03)	-0.02 (0.04)	CAD	0.00	0.02	0.96	1.90 (0.03)	0.00 (0.04)
GBP	0.00	0.02	0.95	1.88 (0.03)	0.29 (0.04)	HKD	0.00	0.02	0.96	1.86 (0.03)	-0.10 (0.04)
HUF	0.00	0.04	0.89	1.66 (0.03)	0.18 (0.04)	KRW	0.00	0.02	0.94	1.81 (0.03)	0.09 (0.04)
LTL	0.00	0.00	0.91	1.24 (0.03)	-0.10 (0.04)	NZD	0.00	0.02	0.94	1.76 (0.03)	0.33 (0.04)
LVL	0.00	0.06	0.81	1.47 (0.03)	0.07 (0.04)	SGD	0.00	0.01	0.98	1.79 (0.03)	0.00 (0.04)
PLN	0.00	0.02	0.96	1.75 (0.03)	0.32 (0.04)	ZAR	0.04	0.04	0.79	1.77 (0.03)	0.35 (0.04)
SEK	0.00	0.02	0.95	1.90 (0.03)	-0.01 (0.04)						

the Hungarian Forint (HUF), the Lithuanian Litas (LTL), the Latvian Lats (LVL), the Polish Zloty (PLN), the Swedish Krona (SEK), the Swiss Franc (CHF), the Norwegian Krone (NOK), the Australian Dollar (AUD), the Canadian Dollar (CAD), the Hong Kong Dollar (HKD), the South Korean Won (KRW), the New Zealand Dollar (NZD), the Singapore Dollar (SGD) and the South African Rand (ZAR). The observations cover the period from January 5, 1999 to August 10, 2012, which corresponds to 3488 observations.[2] Table 3 displays the estimated SP–PGARCH(1,1) models for each series.

We finally fitted a CCC–SP–PGARCH(1,1) model on the bivariate series $\boldsymbol{y}_t = (\text{USA}_t, \text{JPY}_t)'$ of the USA and JPY exchange rate returns. Using the ECF, the estimated model is

$$\boldsymbol{\varepsilon}_t = \mathbf{C}_t^{-1/2}\boldsymbol{y}_t, \quad \mathbf{C}_t = \mathbf{D}_t\mathbf{R}\mathbf{D}_t, \quad \mathbf{D}_t^2 = \text{diag}(\boldsymbol{c}_t)$$

where

$$\begin{aligned} \boldsymbol{c}_t^{(2)} &= \begin{pmatrix} 1.0\,10^{-6} \\ 2.3\,10^{-2} \end{pmatrix} + \begin{pmatrix} 0.032 & 0.003 \\ 0.000 & 0.027 \end{pmatrix} \boldsymbol{y}_{t-1}^{(2)} + \begin{pmatrix} 0.957 & 0.000 \\ 0.000 & 0.894 \end{pmatrix} \boldsymbol{c}_{t-1}^{(2)}, \\ \mathbf{R} &= \begin{pmatrix} 1 & 0.413 \\ 0.413 & 1 \end{pmatrix} \end{aligned}$$

[2]Data source: `http://www.ecb.int/stats/exchange/eurofxref/html/index.en.html`

and $\boldsymbol{\varepsilon}_t$ follows the SP distribution defined by the CF (5.1) with tail index $\alpha = 1.496$.

7 Conclusion

We propose an estimation method for the so–called power GARCH model with stable Paretian (SP) innovations. The method is based on the integrated weighted squared distance between the characteristic function of the SP distribution and an empirical counterpart computed from the GARCH residuals. Under fairly standard conditions the estimator was shown to be consistent. Its asymptotic distribution however proved non–standard and in fact splits into two parts: One regular Gaussian distribution corresponding to the parameters of the SP law, while the other part of the distribution corresponding to the GARCH parameters is singular and in particular it is concentrated on a hyperplane. For the regular Gaussian part it was possible to even optimize the choice of the weight function so that the estimators of the SP parameters attain minimum variance.

Although the simulations results show that the characteristic function–based estimator behaves reasonably and that it is by far less time–consuming than the MLE (which might suggest its use at least as an initial value), more work is needed in order to reveal the finite–sample properties of the proposed estimator. In this connection the proposed method may be viewed as a general method, and given the fact that it was shown to readily extend to multivariate GARCH, it could be considered for other GARCH models for which, like in the present SP–PGARCH model, the innovation distribution is more conveniently parametrized by the characteristic function, rather than by the corresponding density.

Acknowledgement This research was partially supported by Grant No. 11699 of the Special Account for Research Grants of the National and Kapodistrian University of Athens. Simos Meintanis also wishes to express his sincere thanks for the hospitality and financial support of the University of Lille 3. This work was also partially supported by the Agence Nationale de la Recherche (ANR) and the Economic and Social Research Council (ESRC) through the ORA Program (Project PRAM ANR-10-ORAR-008-01)

A Technical proofs

The first two results show that the choice of the unknown initial values is asymptotically unimportant for the objective function and its derivatives.

Proof of (3.2). Using the elementary relation $|\cos x - \cos y| \leq |x - y|$, we have

$$\text{(A.1)} \quad |\mathrm{Re}\,\{\widetilde{\varphi}_T(u) - \varphi_T(u)\}| \leq \frac{1}{T}\sum_{t=1}^{T}\left|uy_t\left(\frac{1}{\widetilde{c}_t} - \frac{1}{c_t}\right)\right|.$$

By the arguments used to show (5.3) in HZ, it is easily shown that

$$\sup_{\boldsymbol{\theta}\in\Theta}|c_t^\rho - \widetilde{c}_t^\rho| \quad \leq \quad K\varrho^t, \quad \forall t.$$

The mean-value theorem and the fact that $\inf_{\boldsymbol{\theta}\in\Theta}\min(c_t, \widetilde{c}_t) \geq \underline{\mu}^{1/\rho} > 0$, then imply that for c_t^{**} between c_t^ρ and $\widetilde{c}_t^\rho$,

$$\sup_{\boldsymbol{\theta}\in\Theta}|c_t - \widetilde{c}_t| \quad = \quad \sup_{\boldsymbol{\theta}\in\Theta}\left|(c_t^\rho - \widetilde{c}_t^\rho)\,\frac{1}{\rho}\,(c_t^{**})^{1/\rho-1}\right| \leq K\varrho^t\left(\max\{c_t, \widetilde{c}_t\}\right)^{1/\rho}.$$

Noting that Re can be replaced by Im in (A.1), we thus have

$$\text{(A.2)} \quad |\widetilde{\varphi}_T(u) - \varphi_T(u)| \leq \frac{K}{T}|u|\sum_{t=1}^{\infty}\varrho^t|y_t|\left(\max\{c_t, \widetilde{c}_t\}\right)^{1/\rho}.$$

The strict stationarity condition in **A2** entails that $E|y_t|^{2s} < \infty$ and $E|c_t|^{2s/\rho} < \infty$ for some small $s > 0$ (see Proposition A.1 in HZ). By the same arguments, we also have $E|\tilde{c}_t|^{2s/\rho} < \infty$. By the Cauchy-Schwarz inequality, the supremum over Θ of the sum of the right-hand side of the inequality (A.2) admits a moment of order s. Therefore this sum is almost surely finite, uniformly in Θ. It follows that

$$\begin{aligned}
& \left||\widetilde{\varphi}_T(u) - \varphi_\varepsilon(u)|^2 - |\varphi_T(u) - \varphi_\varepsilon(u)|^2\right| \\
= & \left|(\widetilde{\varphi}_T(u) - \varphi_T(u))\left(\overline{\widetilde{\varphi}_T}(u) - \overline{\varphi_\varepsilon}(u)\right) + \left(\overline{\widetilde{\varphi}_T}(u) - \overline{\varphi_T}(u)\right)(\varphi_T(u) - \varphi_\varepsilon(u))\right| \\
\leq & \frac{K}{T}|u|.
\end{aligned}$$

We then obtain (3.2) by **A3**(1). □

Proof of (3.12). In view of (3.6) and (3.7), we have

$$\left\|\frac{1}{T}\sum_{s=1}^{T}\frac{\partial g_s(u, \boldsymbol{\vartheta})}{\partial\boldsymbol{\vartheta}}\right\| \leq K + \frac{|u|}{T}\sum_{s=1}^{T}|\varepsilon_t(\boldsymbol{\theta}_0)|\frac{c_t(\boldsymbol{\theta}_0)}{c_t(\boldsymbol{\theta})}\left\|\frac{1}{c_t(\boldsymbol{\theta})}\frac{\partial c_t(\boldsymbol{\theta})}{\partial\boldsymbol{\vartheta}}\right\|.$$

By (3.8), (3.9) and $E\left|\varepsilon_t(\boldsymbol{\theta}_0)\right| < \infty$, the random variable of the right hand side of the last inequality is uniformly integrable in a neighborhood of $\boldsymbol{\vartheta}_0$. By the ergodic theorem, it follows that, when $\boldsymbol{\vartheta}$ is sufficiently close to $\boldsymbol{\vartheta}_0$, the right hand side is *a.s.* bounded by a constant or by u times a constant. Note also that $\left|T^{-1}\sum_{t=1}^{T}\overline{g_t(u,\boldsymbol{\vartheta})}\right| \le 2$. The dominated convergence theorem and **A3**(0)-**A3**(1) thus show that one can take the derivative under the integral symbol to obtain

$$\text{(A.3)} \quad \frac{\partial \Delta_T(\boldsymbol{\vartheta})}{\partial \boldsymbol{\vartheta}} = 2\mathrm{Re}\int_{-\infty}^{\infty} \frac{1}{T}\sum_{t=1}^{T}\overline{g_t(u,\boldsymbol{\vartheta})}\frac{1}{T}\sum_{s=1}^{T}\frac{\partial g_s(u,\boldsymbol{\vartheta})}{\partial \boldsymbol{\vartheta}}W(u)du,$$

at least when $\boldsymbol{\vartheta}$ is sufficiently close to $\boldsymbol{\vartheta}_0$. A similar expression holds for $\partial\widetilde{\Delta}_T(\boldsymbol{\vartheta})/\partial\boldsymbol{\vartheta}$. It follows that

$$\sqrt{T}\sup_{\boldsymbol{\vartheta}\in V(\boldsymbol{\vartheta}_0)}\left\|\frac{\partial \Delta_T(\boldsymbol{\vartheta})}{\partial \boldsymbol{\vartheta}} - \frac{\partial \widetilde{\Delta}_T(\boldsymbol{\vartheta})}{\partial \boldsymbol{\vartheta}}\right\| \le a_T + b_T$$

where, with the obvious notation $\widetilde{g}_t(u,\boldsymbol{\vartheta}) = e^{\mathrm{i}u\widetilde{\varepsilon}_t} - \varphi(u,\boldsymbol{\lambda})$,

$$a_T = \int_{-\infty}^{\infty}\frac{1}{\sqrt{T}}\sum_{t=1}^{\infty}\sup_{\boldsymbol{\vartheta}\in V(\boldsymbol{\vartheta}_0)}\left|g_t(u,\boldsymbol{\vartheta}) - \widetilde{g}_t(u,\boldsymbol{\vartheta})\right|\frac{1}{T}\sum_{s=1}^{T}\sup_{\boldsymbol{\vartheta}\in V(\boldsymbol{\vartheta}_0)}\left\|\frac{\partial g_s(u,\boldsymbol{\vartheta})}{\partial \boldsymbol{\vartheta}}\right\|W(u)du,$$

$$b_T = \int_{-\infty}^{\infty}\frac{1}{T}\sum_{t=1}^{T}\sup_{\boldsymbol{\vartheta}\in V(\boldsymbol{\vartheta}_0)}\left|g_t(u,\boldsymbol{\vartheta})\right|\frac{1}{\sqrt{T}}\sum_{s=1}^{\infty}\sup_{\boldsymbol{\vartheta}\in V(\boldsymbol{\vartheta}_0)}\left\|\frac{\partial g_s(u,\boldsymbol{\vartheta})}{\partial \boldsymbol{\vartheta}} - \frac{\partial \widetilde{g}_s(u,\boldsymbol{\vartheta})}{\partial \boldsymbol{\vartheta}}\right\|W(u)du.$$

By the argument used to show that the series in (A.2) is bounded, we obtain

$$\sum_{t=1}^{\infty}\sup_{\boldsymbol{\vartheta}\in V(\boldsymbol{\vartheta}_0)}\left|g_t(u,\boldsymbol{\vartheta}) - \widetilde{g}_t(u,\boldsymbol{\vartheta})\right| \le K|u| \quad a.s.$$

for some neighborhood $V(\boldsymbol{\vartheta}_0)$. By already used arguments, we also have

$$\frac{1}{T}\sum_{s=1}^{T}\sup_{\boldsymbol{\vartheta}\in V(\boldsymbol{\vartheta}_0)}\left\|\frac{\partial g_s(u,\boldsymbol{\vartheta})}{\partial \boldsymbol{\vartheta}}\right\| \le K + K|u| \quad a.s.$$

It follows that the integrand in a_T is almost surely bounded $KT^{-1/2}(|u|+u^2)W(u)$, when $V(\boldsymbol{\vartheta}_0)$ is sufficiently small. By the dominated convergence theorem and **A3**(1), **A3**(2), almost surely $a_T \to 0$ as $T \to \infty$. Similar arguments show that $b_T \to 0$, and (3.12) follows. □

Lemma A.1 *Let* $\varphi_{\boldsymbol{\lambda}}(u) = e^{-|u|^\alpha\{1-\mathrm{i}\beta\mathrm{sgn}(u)\tan(\pi\alpha/2)\}}$ *where* $\boldsymbol{\lambda} = (\alpha, \beta) \in (1, 2) \times [-1, 1]$. *If for some probability measure* ν *on* $[0, +\infty)$ *admitting a moment of order 10*

$$\int_0^\infty \varphi_{\boldsymbol{\lambda}}(xu)\nu(dx) = \varphi_{\boldsymbol{\lambda}_0}(u) \quad \forall u \in \mathbb{R},$$

then $\boldsymbol{\lambda} = \boldsymbol{\lambda}_0$ *and* ν *is the Dirac measure at the point 1.*

Proof. First order Taylor expansions lead to the following inequalities (see Equation (26.4) in Billingsley, 1995)

$$\left|e^{-|x|} - 1 + |x|\right| \leq \frac{x^2}{2}, \qquad \left|e^{\mathrm{i}x} - 1 - \mathrm{i}x\right| \leq \frac{x^2}{2}.$$

It follows that

$$\varphi_{\boldsymbol{\lambda}_0}(u) = 1 - |u|^{\alpha_0}\{1 - \mathrm{i}\beta_0\mathrm{sgn}(u)\tan(\pi\alpha_0/2)\} + R_{\boldsymbol{\lambda}_0}(u), \tag{A.4}$$

where $|R_{\boldsymbol{\lambda}}(u)| \leq K(|u|^{2\alpha} + |u|^{3\alpha})$ for some constant K. Moreover, as $u \to 0$ we have

$$\int_0^\infty \varphi_{\boldsymbol{\lambda}}(xu)\nu(dx) = 1 - |u|^\alpha \{1 - \mathrm{i}\beta\mathrm{sgn}(u)\tan(\pi\alpha/2)\} \int_0^\infty |x|^\alpha \nu(dx) + O(|u|^{2\alpha}). \tag{A.5}$$

Identifying the right-hand sides of (A.4) and (A.5) as $u \to 0$, we obtain $\alpha = \alpha_0$ and

$$\{1 - \mathrm{i}\beta\mathrm{sgn}(u)\tan(\pi\alpha_0/2)\} \int_0^\infty |x|^{\alpha_0}\nu(dx) = 1 - \mathrm{i}\beta_0\mathrm{sgn}(u)\tan(\pi\alpha_0/2).$$

The real and imaginary parts of both sides being equal, it follows that

$$\int_0^{+\infty} |x|^{\alpha_0}\nu(dx) = 1 \quad \text{and} \quad \beta\int_0^{+\infty} |x|^{\alpha_0}\nu(dx) = \beta_0,$$

from which we deduce that $\beta = \beta_0$. This is not sufficient to conclude concerning the measure ν. Doing Taylor expansions of higher orders, we have

$$\begin{aligned}\varphi_{\boldsymbol{\lambda}_0}(u) &= 1 - |u|^{\alpha_0}\left\{1 - \mathrm{i}\beta_0\mathrm{sgn}(u)\tan\left(\frac{\pi\alpha_0}{2}\right)\right\} \\ &\quad + \frac{|u|^{2\alpha_0}}{2}\left\{1 - \beta_0^2\tan^2\left(\frac{\pi\alpha_0}{2}\right) - 2\mathrm{i}\beta_0\mathrm{sgn}(u)\left(\frac{\pi\alpha_0}{2}\right)\right\} + R^*_{\boldsymbol{\lambda}_0}(u).\end{aligned}$$

where $|R^*_{\boldsymbol{\lambda}}(u)| \leq K(|u|^{3\alpha}+|u|^{4\alpha}+|u|^{5\alpha})$ for some constant K. Note that $\int R^*_{\boldsymbol{\lambda}}(xu)\nu(dx) = o(|u|^{2\alpha})$ as $u \to 0$ because $\int |x|^{5\alpha}\nu(dx) < \infty$. Identifying the approximations of $\varphi_{\boldsymbol{\lambda}_0}(u)$ and $\int_0^\infty \varphi_{\boldsymbol{\lambda}_0}(xu)\nu(dx)$ as $u \to 0$, we obtain $\int_0^{+\infty} |x|^{2\alpha_0}\nu(dx) = \int_0^{+\infty} |x|^{\alpha_0}\nu(dx) = 1$. It follows that ν is the Dirac measure at 1, which completes the proof. $\square$

Remark A.2 The need of moment assumptions on the probability measure ν is already evident from representation (5.2) which suggests that a symmetric SP random variable can be obtained as a scale mixture of normal distributions, with mixing distribution a SP distribution concentrated on $[0, \infty)$. A more general result involving non–normal mixtures of SP distributions is proved by Samorodnitsky and Taqqu (1994, §1.3). By way of example we consider the random variable $W = X^{1/2}Z$, where Z is standard normal with CF $\varphi_{\boldsymbol{\lambda}}(u) = e^{-(1/2)u^2}$ and X follows the Lévy distribution (see Section 1), which is a totally skewed to the right SP distribution with tail index $\alpha = 1/2$ and density $\nu(dx) = 1/(\sqrt{2\pi})x^{-3/2}e^{-1/(2x)}dx,\ x > 0$. Denote by $\varphi_W(u)$ the CF of W. Then it readily follows that

$$\varphi_W(u) = \int_0^\infty \varphi_{\boldsymbol{\lambda}}(x^{1/2}u)\nu(dx) = \frac{1}{\sqrt{2\pi}}\int_0^\infty \frac{1}{x^{3/2}}e^{-\frac{1}{2}\left(xu^2+\frac{1}{x}\right)}dx = e^{-|u|},$$

which shows that W follows the Cauchy distribution, and consequently that this distribution has a stochastic representation as a mixture of normal distributions with variance following the Lévy distribution.

The following lemma is similar to Lemma 1 in Tauchen (1985) and Lemma 2.4 in Newey and McFadden (1994), except that the assumption of iid observations is relaxed.

Lemma A.3 *Let (z_t) be a stationary and ergodic process. Assume that Θ is compact, that $\theta \mapsto a(z, \theta)$ is continuous on Θ for all $z \in \Omega_1$ such that $P(z_1 \in \Omega_1) = 1$, and that there exists $d(z)$ such that $\|a(z, \theta)\| \leq d(z)$ for all $\theta \in \Theta$ and $Ed(z_1) < \infty$. Then $\theta \mapsto Ea(z_1, \theta)$ is continuous and*

$$\sup_{\theta\in\Theta}\left\|\frac{1}{T}\sum_{t=1}^{T} a(z_t, \theta) - Ea(z_1, \theta)\right\| \to 0 \quad a.s. \quad as \quad T \to \infty.$$

Proof. Let $V_m(\theta)$ be the open ball of center θ and radius $1/m$. The dominated convergence theorem entails that for all $\theta_k \in \Theta$ and all $\epsilon > 0$ there exists m such that the neighborhood $V(\theta_k) = V_m(\theta_k)$ satisfies

$$\text{(A.6)} \quad E \sup_{\theta\in V(\theta_k)\cap\Theta} \|a(z_1, \theta) - a(z_1, \theta_k)\| \leq \epsilon.$$

By a compactness argument, there exist $\theta_1, \ldots, \theta_K$ such that $\cup_{k=1}^{K} V(\theta_k) \subseteq \Theta$ where

$V(\theta_k)$ satisfies (A.6). Now note that

$$\sup_{\theta \in V(\theta_k)\cap\Theta} \left\| \frac{1}{T}\sum_{t=1}^{T} a(z_t,\theta) - Ea(z_1,\theta) \right\| \leq \frac{1}{T}\sum_{t=1}^{T} \sup_{\theta \in V(\theta_k)\cap\Theta} \| a(z_t,\theta) - a(z_t,\theta_k) \|$$
$$+ \left\| \frac{1}{T}\sum_{t=1}^{T} a(z_t,\theta_k) - Ea(z_1,\theta_k) \right\| + \sup_{\theta \in V(\theta_k)\cap\Theta} \| Ea(z_1,\theta_k) - Ea(z_1,\theta) \| .$$

The ergodic theorem and (A.6) entail that, as $T \to \infty$, the almost sure limit of the first term of the right-hand side of the inequality is bounded by ϵ. The ergodic theorem also shows that the limit of the second term is zero. By (A.6), the last term is bounded by ϵ. Since ϵ is arbitrarily small, the conclusion follows. □

Lemma A.4 *Let ε be a random variable with the SP distribution of parameter $\boldsymbol{\lambda} = (\alpha, \beta) \in (1,2) \times [-1,1]$. For all $\nu \in (-3, \alpha - 1)$, there exists a constant K such that for all $c > 0$*

$$\left| E|\varepsilon|^{2+\nu} e^{\mathrm{i}c\varepsilon} \right| \leq K + \frac{K}{c}.$$

Proof. The density $f_{\boldsymbol{\lambda}}(x)$ of ε is bounded and satisfies

$$f_{\boldsymbol{\lambda}}(x) \sim \frac{K}{x^{\alpha+1}} \quad \text{as} \quad |x| \to \infty,$$

for some constant $K = K_{\boldsymbol{\lambda}}$ (see *e.g.* Theorem 1.12 in Nolan, 2012). To show the existence of $E|\varepsilon|^{2+\nu} e^{\mathrm{i}c\varepsilon}$, it is thus sufficient to show the existence of

$$\int_1^\infty \cos(cx) \frac{x^{2+\nu}}{x^{\alpha+1}} dx \quad \text{and} \quad \int_1^\infty \sin(cx) \frac{x^{2+\nu}}{x^{\alpha+1}} dx.$$

An integration by parts shows that the first integral is equal to

$$-\frac{\sin c}{c} + \frac{\alpha - \nu - 1}{c} \int_1^\infty \frac{\sin(cx)}{x^{\alpha-\nu}} dx.$$

Similarly, is can be seen that the second integral is also bounded by K/c. The conclusion follows. □

Proof of (3.14). First note that, similarly to (3.10), we have

$$\text{(A.7)} \quad E \sup_{\theta \in \Theta} \left\| \frac{1}{c_t} \frac{\partial^2 c_t(\boldsymbol{\theta})}{\partial \boldsymbol{\theta} \partial \boldsymbol{\theta}'} \right\|^r < \infty \quad \text{for all } r > 0.$$

We now consider the second order derivatives of $g_s(u,\boldsymbol{\vartheta})$. Note that $\partial^2 g_s(u,\boldsymbol{\vartheta})/\partial\boldsymbol{\theta}\partial\boldsymbol{\lambda}' = 0$, that $\partial^2 g_s(u,\boldsymbol{\vartheta})/\partial\boldsymbol{\lambda}\partial\boldsymbol{\lambda}'$ is a non random bounded function of u uniformly in $\boldsymbol{\vartheta}$, and

$$\frac{\partial^2 g_s(u,\boldsymbol{\vartheta})}{\partial\boldsymbol{\theta}\partial\boldsymbol{\theta}'} = ue^{\mathrm{i}u\varepsilon_t}\varepsilon_t\left\{-u\varepsilon_t\frac{1}{c_t^2}\frac{\partial c_t}{\partial\boldsymbol{\theta}}\frac{\partial c_t}{\partial\boldsymbol{\theta}'} + 2\mathrm{i}\frac{1}{c_t^2}\frac{\partial c_t}{\partial\boldsymbol{\theta}}\frac{\partial c_t}{\partial\boldsymbol{\theta}'} - \mathrm{i}\frac{1}{c_t}\frac{\partial^2 c_t}{\partial\boldsymbol{\theta}\partial\boldsymbol{\theta}'}\right\}.$$

Conditioning by the past, using Lemma A.4 with $c = |u|c_t(\boldsymbol{\theta}_0)/c_t(\boldsymbol{\theta})$ and $\nu = 0, -1, -2$, and using (3.8), (3.9) and (A.7), it can be shown that there exist a neighborhood $V(\boldsymbol{\vartheta}_0)$ of $\boldsymbol{\vartheta}_0$ and a constant K independent of u such that

$$E\sup_{\boldsymbol{\vartheta}\in V(\boldsymbol{\vartheta}_0)}\left\|\frac{\partial^2 g_s(u,\boldsymbol{\vartheta})}{\partial\boldsymbol{\vartheta}\partial\boldsymbol{\vartheta}'}\right\| \leq K(1+|u|+u^2).$$

Using this result, a Taylor expansion and the ergodic theorem, we obtain

$$\left\|\frac{1}{T}\sum_{s=1}^{T}\frac{\partial g_s(u,\widehat{\boldsymbol{\vartheta}}_T)}{\partial\boldsymbol{\vartheta}} - \frac{1}{T}\sum_{s=1}^{T}\frac{\partial g_s(u,\boldsymbol{\vartheta}_0)}{\partial\boldsymbol{\vartheta}}\right\| \leq K(1+|u|+u^2)\left\|\widehat{\boldsymbol{\vartheta}}_T - \boldsymbol{\vartheta}_0\right\|.$$

Now, note that

$$\frac{1}{T}\sum_{s=1}^{T}\frac{\partial g_s(u,\boldsymbol{\vartheta}_0)}{\partial\boldsymbol{\theta}} - E\frac{\partial g_1(u,\boldsymbol{\vartheta}_0)}{\partial\boldsymbol{\theta}} = \frac{1}{T}\sum_{t=1}^{T}(\boldsymbol{X}_t - E\boldsymbol{X}_1),$$

where, in view of Lemma A.3 and (3.9),

$$\boldsymbol{X}_t = -\mathrm{i}ue^{\mathrm{i}u\varepsilon_t(\boldsymbol{\theta}_0)}\varepsilon_t(\boldsymbol{\theta}_0)\frac{1}{c_t}\frac{\partial c_t(\boldsymbol{\theta}_0)}{\partial\boldsymbol{\theta}}$$

is such that $E\|\boldsymbol{X}_t\| \leq K(1+|u|)$. Note also that

$$\frac{1}{T}\sum_{s=1}^{T}\frac{\partial g_s(u,\boldsymbol{\vartheta}_0)}{\partial\boldsymbol{\lambda}} - E\frac{\partial g_1(u,\boldsymbol{\vartheta}_0)}{\partial\boldsymbol{\lambda}} = 0.$$

We thus have shown that

$$\text{(A.8)}\qquad \left\|\frac{1}{T}\sum_{s=1}^{T}\frac{\partial g_s(u,\widehat{\boldsymbol{\vartheta}}_T)}{\partial\boldsymbol{\vartheta}} - E\frac{\partial g_1(u,\boldsymbol{\vartheta}_0)}{\partial\boldsymbol{\vartheta}}\right\| \leq K(1+|u|+u^2)\left\{\left\|\widehat{\boldsymbol{\vartheta}}_T - \boldsymbol{\vartheta}_0\right\| + o_P(1)\right\}.$$

A Taylor expansion shows that

$$\frac{1}{\sqrt{T}}\sum_{t=1}^{T}\mathrm{Re}\ g_t(u,\widehat{\boldsymbol{\vartheta}}_T) = \frac{1}{\sqrt{T}}\sum_{t=1}^{T}\mathrm{Re}\ g_t(u,\boldsymbol{\vartheta}_0) + \frac{1}{T}\sum_{t=1}^{T}\frac{\partial\mathrm{Re}\ g_t(u,\boldsymbol{\vartheta}_T^*)}{\partial\boldsymbol{\vartheta}'}\sqrt{T}(\widehat{\boldsymbol{\vartheta}}_T - \boldsymbol{\vartheta}_0)$$

for some $\boldsymbol{\vartheta}_T^*$ between $\widehat{\boldsymbol{\vartheta}}_T$ and $\boldsymbol{\vartheta}_0$. A similar expansion holds for the imaginary part. We thus have

$$\frac{1}{\sqrt{T}}\sum_{t=1}^{T}\overline{g_t(u,\widehat{\boldsymbol{\vartheta}}_T)} = \frac{1}{\sqrt{T}}\sum_{t=1}^{T}\overline{g_t(u,\boldsymbol{\vartheta}_0)} + \left\{E\frac{\partial\overline{g_1(u,\boldsymbol{\vartheta}_0)}}{\partial\boldsymbol{\vartheta}} + R_T(u)\right\}\sqrt{T}(\widehat{\boldsymbol{\vartheta}}_T - \boldsymbol{\vartheta}_0)$$

where, in view of (A.8), $\|R_T(u)\| \leq K(1+|u|+u^2)\left\{\left\|\widehat{\boldsymbol{\vartheta}}_T - \boldsymbol{\vartheta}_0\right\| + o_P(1)\right\}$. By the previous result, (A.8), **A3**(0) and **A3**(4), Equation (3.13) yields (3.14). □

References

Adler, R.J., Feldman, R.E., and Taqqu, M.S. (Eds.) *A Practical Guide to Heavy Tails. Statistical Techniques and Applications*. Birkhäuser, Boston, 1998.

Akgül, I., and Sayyan, H. (2008). Modelling and forecasting long memory in exchange rate volatility vs. stable and integrated GARCH models. *Appl. Financ. Econom.*, *18*, 463–482.

Billingsley, P. *Probability and Measure*. Wiley and Sons, New York, 1995.

Bryant, J.L. and Paulson, A.S. (1979). Some comments on characteristic function based estimators. *Sankhy*ā, *41*, 109–116.

Bonato, M. (2009). Modeling fat tails in stock returns: a multivariate stable–GARCH approach. Electronic copy available at http://ssrn.com/abstract=1015477.

Bougerol, P., and Picard, N. (1992). Stationarity of GARCH processes and of some nonnegative time series. *J. Econometr.*, *52*, 115–127.

Carrasco, M., Chernov, M., Florens, J.P., and Ghysels, E. (2007). Efficient estimation of general dynamic models with a continuum of moment conditions. *J. Econometr.*, *140*, 529–573.

Curto, J.D., Pinto, J.C., and Tavares, G.N. (2009). Modelling stock markets' volatility using GARCH models with Normal, Student's t and stable Paretian distribution. *Statist. Papers*, *50*, 311–321.

DuMouchel, W.H. (1983). Estimating the stable index α in order to measure tail thickness: A critique. *Ann. Statist.*, *11*, 1019–1031.

Fama, E. (1965). The behavior of stock market prices. *J. Business*, *38*, 34–105.

Feuerverger, A. (1990). An efficiency result for the empirical characteristic function in stationery time–series models. *Canad. J. Statist.*, *18*, 155–161.

Feuerverger, A. and McDunnough, P. (1981a). On the efficiency of empirical characteristic function procedures. *J. Roy. Statist. Soc. Ser. B* , *38*, 20–27.

Feuerverger, A. and McDunnough, P. (1981b). On some Fourier methods for inference. *J. Amer. Statist. Assoc.*, *76*, 379–387.

Hamadeh, T., and Zakoïan, J.M. (2011). Asymptotic properties of LS and QML estimators for a class of nonlinear GARCH processes. *J. Statist. Plann. Inference*, *141*, 488–507.

Hansen, L.P. (1982). Large sample properties of generalized method of moments estimators. *Econometrica*, *50*, 1029–1054.

Heathcote, C.R. (1977). The integrated squared error estimation of parameters. *Biometrika*, *64*, 255–264.

Kotchoni, R. (2012). Applications of the characteristic function–based continuum GMM in finance. *Comput. Statist. Dat. Anal.* , *56*, 3599–3622.

Koutrouvelis, I.A., and Meintanis, S.G. (1999). Testing for stability based on the empirical characteristic function with applications to financial data. *J. Statist. Comput. Simul.*, *64*, 275–300.

Liu, S.M., and Brorsen, B.W. (1995a). GARCH–stable as a model for future price movements. *Rev. Quantitat. Financ. Account.*, *5*, 155–167.

Liu, S.M., and Brorsen, B.W. (1995b). Maximum likelihood estimation of a GARCH–stable model. *J. Appl. Econometr.*, *10*, 273–285.

Mandelbrot, B. (1963). The variation of certain speculative prices. *J. Business*, *36*, 394–419.

Mittnik, S., and Rachev, S.T. (1993). Modelling asset returns with alternative stable models. *Econometr. Rev.*, *12*, 261–330.

Mittnik, S., Rachev, S.T., Doganoglu, T, and Chenyao, D. (1999). Maximum likelihood estimation of stable Paretian models. *Math. Comput. Model.*, *29*, 275–293.

Newey, W.K., and McFadden, D. (1994). Large sample estimation and hypothesis testing. In: Engle and McFadden (Eds.) *Handbook of Econometrics, Vol. IV*, 2111–2245, Elsevier.

Nolan, J.P. *Stable Distributions - Models for Heavy Tailed Data*. Birkhauser, Boston, 2012.

Paolella, M.S. (2001). Testing the stable Paretian assumption. *Math. Comput. Model.* , *34*, 1095–1112.

Rachev, S. (Ed.) *Handbook of Heavy Tailed Distributions in Finance*. Elsevier/North–Holland, Amsterdam, 2003.

Rachev, S., and Mittnik, S. *Stable Paretian Models in Finance*. Wiley, New York, 2000.

Samorodnitsky, G., and Taqqu, M.S. *Stable Non–Gaussian Random Processes: Stochastic Models with Infinite Variance*. Chapman and Hall, New York, 1994.

Tauchen, G.E. (1985). Diagnostic testing and evaluation of maximum likelihood models . *J. Econometr.*, *30*, 415–443.

Tavares, A.B., Curto, J.D., and Tavares, G.N. (2008). Modelling heavy tails and asymmetry using ARCH–type models with stable Paretian distribution. *Nonlinear Dyn.*, *51*, 231–243.

Thornton, J.C. and Paulson, A.S. (1977). Asymptotic distribution of characteristic function–based estimators for the stable laws. *Sankhyā*, *39*, 341–354.

Tsionas, E. (2002). Likelihood–based comparison of stable Paretian and competing models: Evidence from daily exchange rates. *J. Statist. Comput. Simul.* , *72*, 341–353.

Xu, W., Wu, C., Dong, Y., and Xiao, W. (2011). Modeling Chinese stock returns with stable distribution. *Math. Comput. Model.*, *54*, 610–617.