Estimating the marginal law of a time series with applications to heavy tailed distributions

Christian Francq* Jean-Michel Zakoïan†

Abstract: This paper concerns estimating parametric marginal densities of stationary time series in absence of precise information on the dynamics of the underlying process. We propose to use an estimator obtained by maximization of the "quasi marginal" likelihood, which is a likelihood written as if the observations were independent. We study the effect of the (neglected) dynamics on the asymptotic behavior of this estimator. The consistency and asymptotic normality of the estimator are established under mild assumptions on the dependence structure. Applications of the asymptotic results to the estimation of stable, generalized extreme value and generalized Pareto distributions are proposed. The theoretical results are illustrated on financial index returns. All the supplemental materials used by this paper are available online.

Keywords: Alpha-stable distribution; Composite likelihood; GEV distribution; GPD; Pseudo likelihood; Quasi marginal maximum likelihood; Stock returns distributions.

*CREST and Université Lille 3 (EQUIPPE), BP 60 149, 59653 Villeneuve d'Ascq cedex, France. E-mail: christian.francq@univ-lille3.fr, tel: 33.3.20.41.64.87

†Corresponding author. CREST, 15 Bd G. Péri, 92245 Malakoff Cedex, France and Université Lille 3 (EQUIPPE). E-mail: zakoian@ensae.fr, tel: 33.1.41.17.77.25

1 Introduction

The marginal distribution of a stationary time series contains interesting information. If one is interested in prediction or value-at-risk evaluation over long horizons, this is the marginal distribution that matters (see *e.g.* Cotter, 2007). Another area where marginal densities play an important role is the estimation of copula-based stationary models: for instance Chen and Fan (2006) proposed a copula approach whose advantage is to "separate out the temporal dependence (such as tail dependence) from the marginal behavior (such as fat tailedness) of a time series." On the other hand, numerous statistical procedures require conditions on the marginal law, such as the existence of moments. Moreover, statistical inference on the marginal distribution can help validate or invalidate time series models. For example, a linear model with alpha-stable innovations entails the same type of distribution for the observations (see Remark 1 of Proposition 13.3.1 in Brockwell and Davis, 1991).

In principle, the marginal distribution is specified by the time series model. However, the closed parametric form of the marginal density is known only in special cases. Examples include the linear autoregressive processes with Gaussian or stable innovations, and some threshold autoregressive processes with very specific error distributions (see Anděl and Bartoň (1986), Anděl, Netuka and Zvára (1984), and more recently Loges (2004)). Moreover, in real situations, the dynamics of the series and the errors distribution are generally unknown.

Our aim in this paper is to estimate the parameterized marginal distribution of a stationary times series (X_t) without specifying its dependence structure. The focus is on the parameter of the marginal distribution, and the unknown dependence structure can be considered as a nuisance parameter in our framework.

To deal with situations where the computation of the exact likelihood is not feasible, Lindsay (1988) proposed the composite likelihood as a pseudo-likelihood for inference. A composite likelihood consists of a combination of likelihoods of small subsets of data. For example, the composite likelihood can be the product of the bivariate likelihood of pairs of observations (see Davis and Yau (2011) for the asymptotic properties of pairwise likelihood estimation procedures for linear time series models). Here the bivariate likelihood is not available, only the univariate likelihood is assumed to be known. Applying the composite likelihood principle to our framework, we thus write the likelihood corresponding to independent observations, neglecting the dependence structure. As will be seen, neglecting the dependence may however have important effects on the accuracy of the estimators. The corresponding estimator will be called Quasi-Marginal MLE (QMMLE). This estimator is actually widely employed with the name of MLE, but this estimator is not the MLE in the presence of time dependance. In the present paper, the asymptotic distribution of this estimator is studied by taking into account the temporal dependence, but without specifying a particular model. Our only assumption concerning the dependence structure is a classical mixing assumption, which is known to hold for an immense collection of time series models.

Our results apply, in particular, to heavy tailed time series, which have attracted a great deal of attention in recent years. Number of fields, in particular Environment, Insurance and Finance, use data sets which seem compatible with the assumption of heavy-tailed marginal distributions. For instance it has been long known that asset returns are not normally distributed. Mandelbrot (1963) and Fama (1965) pioneered the use of heavy-tailed random variables, with $P(X > x) \sim Cx^{-\alpha}$, for financial returns. Mandelbrot advocated the use of infinite-variance stable (Pareto-Lévy) distributions. See Rachev and Mittnik (2000) for a detailed analysis of stable distributions. The use of other heavy tailed distributions, for instance the Generalized Pareto Distribution (GPD) and the Generalized

Extreme Value distribution (GEV), was advocated by many authors. See Rachev (2003) for an account of the many applications of heavy-tailed distributions in finance. The GPD and GEV play a central role in extreme value theory (EVT) (see *e.g.* Beirlant et al. 2005).

Asymptotic theory of estimation for stable distributions has been established by DuMouchel (1973). He showed that, whenever $\alpha < 2$, the Maximum Likelihood Estimator (MLE) of the coefficient α has an asymptotic normal distribution. Asymptotic properties of the MLE of GPD and GEV parameters were obtained by Smith (1984, 1985). However, a limitation of those results is that their validity require independent and identically distributed (iid) observations. The independence assumption is clearly unsatisfied for most of the series to which these distributions are usually adjusted. This is in particular the case for financial returns. Autocorrelations of squares and volatility clustering, for instance, have been extensively documented for such series.

The paper is organized as follows. Section 2 defines the QMMLE and gives general regularity conditions for its consistency and asymptotic normality. The next section shows that the regularity conditions of Section 2 are satisfied for three important classes of heavy-tailed distributions. The alpha-stable, the generalized Pareto and the generalized extreme value distributions are considered respectively in Section 3.1, Section 3.2 and Section 3.3. Applications to the marginal distribution of financial returns are proposed in Section 4. Section 5 concludes. Proofs are relegated to an appendix.

2 The Quasi-Marginal MLE

In this section we consider the general problem of estimating the marginal distribution of a stationary time series $X_1, \dots, X_n$ defined on a probability space $(\Omega, \mathcal{A}, P)$

and taking its values in a non empty measurable space $(E, \mathcal{E})$. Assume that X_t admits a density f_{θ_0} with respect to some σ-finite measure μ on $(E, \mathcal{E})$. We consider the unknown dependent structure as a nuisance parameter and we concentrate on the estimation of the parameter $\theta_0 \in \Theta \subset \mathbb{R}^q$. In similar situations, where dependencies constitute a nuisance, one can use an estimator obtained by maximizing a quasi-likelihood (also known as pseudo-likelihood or composite likelihood) which treats the data values as being independent (see Lindsay (1988)). This leads to define a QMMLE[1] as any measurable solution of

$$\hat{\theta}_n = \arg\min_{\theta \in \Theta} \ell_n(\theta), \quad \ell_n(\theta) = -\frac{1}{n} \sum_{t=1}^{n} \log f_\theta(X_t). \tag{2.1}$$

To guarantee the existence of a solution to this optimization problem, we assume

A1: the set $\{x \in E : f_\theta(x) > 0\}$ does not depend on θ, the function $\theta \to f_\theta(x)$ is continuous for all $x \in E$ and Θ is compact.

Ignoring the time series dependence, the estimator $\hat{\theta}_n$ is often called MLE. Note however that, in general, $\hat{\theta}_n$ does not coincide with the MLE when the observations are not iid. Standard estimation methods based on the likelihood, or the quasi-likelihood, cannot be implemented when the conditional distribution of X_t given its past, or at least when the conditional moments of X_t given its past, are unknown. The main interest of the QMMLE is to avoid specifying a particular dynamics.

[1] We emphasize the difference with the so-called Quasi MLE: in the latter case, the first two conditional moments are supposed to be correctly specified and the criterion is written as if the conditional distribution were Gaussian; in the present paper, the marginal distribution is supposed to be correctly specified but the criterion is written as if the observations were independent.

2.1 Consistency and asymptotic normality of the quasi marginal MLE

The QMMLE $\hat{\theta}_n$ is CAN (consistent and asymptotically normal) under regularity assumptions similar to those made for the CAN of the MLE (see *e.g.* Tjøstheim, 1986, Pötscher and Prucha, 1997, Berkes and Horváth, 2004, McAleer and Ling, 2010). More precisely, the following standard identifiability and moment assumptions are made:

A2: $f_\theta(X_1) = f_{\theta_0}(X_1)$ almost surely (a.s.) implies $\theta = \theta_0$.

A3: $E\left|\log f_\theta(X_1)\right| < \infty$ for all $\theta \in \Theta$.

For the asymptotic normality, we need additional regularity assumptions.

A4: θ_0 belongs to the interior $\overset{\circ}{\Theta}$ of Θ, the function $\theta = (\theta_1, \dots, \theta_q)' \to f_\theta(x)$ admits third-order derivatives, for all $i, j, k \in \{1, \dots, q\}$ there exists a neighborhood $V(\theta_0)$ of θ_0 such that $E\sup_{\theta \in V(\theta_0)} \left|\frac{\partial^3 \log f_\theta(X_1)}{\partial\theta_i \partial\theta_j \partial\theta_k}\right| < \infty$, the matrices

$$I = \sum_{h=-\infty}^{\infty} E\frac{\partial \log f_{\theta_0}(X_1)}{\partial\theta}\frac{\partial \log f_{\theta_0}(X_{1+h})}{\partial\theta'} \quad \text{and} \quad J = -E\frac{\partial^2 \log f_{\theta_0}(X_1)}{\partial\theta\partial\theta'}$$

exist and J is nonsingular.

In the iid case, $J = I$ is the Fisher information matrix. In the general case, I is a so-called long-run variance (LRV) matrix. We also have to assume that the serial dependence is not too strong:

A5: $E\left\|\frac{\partial \log f_{\theta_0}(X_1)}{\partial\theta}\right\|^{2+\nu} < \infty$ and $\sum_{k=0}^{\infty}\{\alpha_X(k)\}^{\frac{\nu}{2+\nu}} < \infty$ for some $\nu > 0$,

where $\alpha_X(k)$, $k = 0, 1, \dots,$ denote the strong mixing coefficients of the process (X_t) (see *e.g.* Bradley, 2005, for a review on strong mixing conditions).

Theorem 2.1. *If (X_t) is a stationary and ergodic process with marginal density f_{θ_0}, and if* **A1-A3** *hold true, then $\hat{\theta}_n \to \theta_0$ a.s. Under the additional assumptions* **A4** *and* **A5**, *we have*

$$\sqrt{n}\left(\hat{\theta}_n - \theta_0\right) \xrightarrow{d} \mathcal{N}(0, \Sigma := J^{-1}IJ^{-1}) \text{ as } n \to \infty.$$

In the iid case, we have $I = J$. The following example shows that, for time series, Σ may be quite different from J^{-1}.

Example 2.1. Consider the simplistic example of an AR(1) of the form

$$X_t = a_0 X_{t-1} + \eta_t, \qquad \eta_t \text{ iid } \mathcal{N}(0, \sigma_0^2), \quad a_0 \in (-1, 1), \quad \sigma_0 > 0$$

and assume that the parameter of interest is $\theta_0 = \mathrm{Var} X_t = \sigma_0^2/(1 - a_0^2)$. We have

$$\frac{\partial \log f_{\theta_0}(x)}{\partial \theta} = \frac{x^2 - \theta_0}{2\theta_0^2}.$$

Therefore we have

$$J = \frac{1}{2\theta_0^2}, \quad I = \frac{1}{4\theta_0^4} \sum_{h=-\infty}^{\infty} \mathrm{Cov}\left(X_1^2, X_{1+h}^2\right) = \frac{1}{4\theta_0^4} \mathrm{Var}(X_1^2) \left(\frac{1 + a_0^2}{1 - a_0^2}\right)$$

with $\mathrm{Var}(X_1^2) = 2\theta_0^2$. The QMMLE is thus $\hat{\theta}_n = n^{-1} \sum_{t=1}^{n} X_t^2$ and it satisfies

$$\sqrt{n}\left(\hat{\theta}_n - \theta_0\right) \xrightarrow{d} \mathcal{N}\left(0, \Sigma = 2\theta_0^2 \frac{1 + a_0^2}{1 - a_0^2}\right) \text{ as } n \to \infty.$$

Figure 1 shows that the dynamics is crucial for the asymptotic distribution of the QMMLE, in the sense that Σ is much greater than J^{-1} when a_0 is far from 0.

It is well known that the MLE $\hat{\vartheta}_{MLE}$ of $\vartheta_0 = (a_0, \sigma_0^2)'$ satisfies

$$\sqrt{n}\left(\hat{\vartheta}_{MLE} - \vartheta_0\right) \xrightarrow{d} \mathcal{N}\left\{0, \begin{pmatrix} 1 - a_0^2 & 0 \\ 0 & 2\sigma_0^4 \end{pmatrix}\right\} \text{ as } n \to \infty.$$

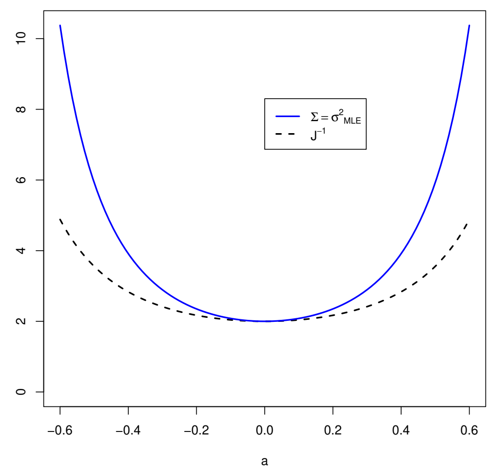

Figure 1: Asymptotic variances Σ of the QMMLE and J^{-1} of the iid MLE, for the AR(1) of Example 2.1 with $\sigma_0^2 = 1$.

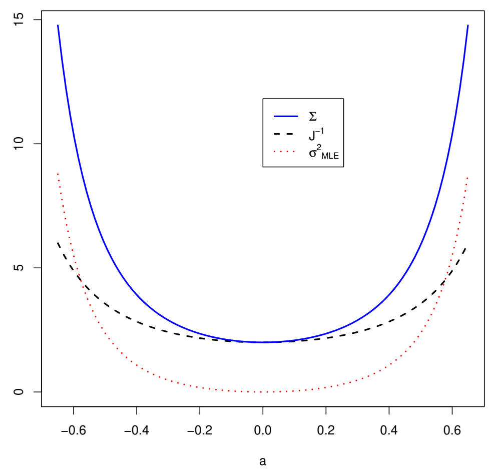

Figure 2: As in Figure 1 for the AR(1) of Example 2.2, with the asymptotic variance σ^2_{MLE} of the MLE.

By the delta method, the MLE $\hat{\theta}_{MLE}$ of θ_0 thus satisfies $\sqrt{n}\left(\hat{\theta}_{MLE} - \theta_0\right) \xrightarrow{d} \mathcal{N}\left(0, \sigma^2_{MLE}\right)$, with

$$\sigma^2_{MLE} = \begin{pmatrix} \frac{2a_0\sigma_0^2}{(1-a_0^2)^2} & \frac{1}{1-a_0^2} \end{pmatrix} \begin{pmatrix} 1 - a_0^2 & 0 \\ 0 & 2\sigma_0^4 \end{pmatrix} \begin{pmatrix} \frac{2a_0\sigma_0^2}{(1-a_0^2)^2} \\ \frac{1}{1-a_0^2} \end{pmatrix} = \frac{2\sigma_0^4(1+a_0^2)}{(1-a_0^2)^3}.$$

Note that $\Sigma = \sigma^2_{MLE}$. Thus, for this particular example, the QMMLE and the MLE have the same asymptotic distribution.

In the previous example, the QMMLE was as efficient as the MLE. The following example shows that, as expected, we may have an efficiency loss of the QMMLE with respect to the MLE, which can be considered as the price to pay for not having to specify the dynamics.

Example 2.2. Consider another example of an AR(1) of the form

$$X_t = a_0 X_{t-1} + \eta_t, \qquad a_0 \in (-1, 1), \quad \eta_t \text{ iid } \mathcal{N}(0, 1),$$

and assume that the parameter of interest is $\theta_0 = \mathrm{Var} X_t = (1-a_0^2)^{-1}$. Using the computation of the previous example, the QMMLE $\hat{\theta}_n = n^{-1}\sum_{t=1}^n X_t^2$ satisfies

$$\sqrt{n}\left(\hat{\theta}_n - \theta_0\right) \xrightarrow{d} \mathcal{N}\left\{0, \Sigma = 2\theta_0^2(2\theta_0 - 1)\right\} \text{ as } n \to \infty.$$

It is known that the MLE of a_0 satisfies

$$\sqrt{n}(\hat{a}_n - a_0) \xrightarrow{d} \mathcal{N}\left\{0, 1 - a_0^2\right\}.$$

Since $\theta'(a) = 2a/(1-a^2)^2$, the delta method shows that the MLE of θ_0 satisfies

$$\sqrt{n}\left(\hat{\theta}_{MLE} - \theta_0\right) \xrightarrow{d} \mathcal{N}\left\{0, \sigma_{MLE}^2 = 4(\theta_0 - 1)\theta_0^2\right\} \text{ as } n \to \infty.$$

Figure 2 shows that, for this very particular model, the MLE always clearly outperforms the QMMLE. Indeed, if we know that the observations are generated by an AR(1) with standard Gaussian innovations, than the marginal variance θ_0 is entirely defined by the AR coefficient. Thus it is not surprising that the estimator of θ_0 based on the MLE of a be more efficient than a simple empirical moment. Figure 2 also shows that J^{-1}, which is the asymptotical variance of the MLE of θ_0 in the iid case, is very far from the asymptotic variance of the MLE or of the QMMLE in the time series case.

Note that Theorem 2.1 does not allow to treat interesting cases where the support of the density depends on θ and/or cases where $\theta \to f_\theta(x)$ is not differentiable for all x. The GEV density is an example of such densities, that we would like to fit with QMMLE. To this purpose, consider the alternative assumptions.

A1*: for P_{θ_0} almost all x, the function $\theta \to f_\theta(x)$ is continuous and Θ is compact.

A3*: $E\left|\log f_{\theta_0}(X_1)\right| < \infty$ and $E \log^+ f_\theta(X_1) < \infty$ for all $\theta \in \Theta$.

A4*: there exists $\mathcal{X} \in \mathcal{E}$ such that $P(X_t \in \mathcal{X}) = 1$, for all $x \in \mathcal{X}$ the function $\theta \to f_\theta(x)$ admits third-order derivatives at θ_0, and all the other requirements of **A4** are satisfied.

Theorem 2.2. *If (X_t) is a stationary and ergodic process with marginal density f_{θ_0}, and if* **A1***, **A2** *and* **A3*** *hold true, then $\hat{\theta}_n \to \theta_0$ a.s. Under the additional assumptions* **A4*** *and* **A5**, *we have*

$$\sqrt{n}\left(\hat{\theta}_n - \theta_0\right) \xrightarrow{d} \mathcal{N}(0, \Sigma = J^{-1} I J^{-1}) \text{ as } n \to \infty.$$

As illustrated by Examples 2.1–2.2, it is essential to estimate consistently the standard Fisher information matrix J and the LRV matrix I. This problem is considered in the following section.

2.2 Estimation of the asymptotic variance

Since J is equal to the variance of the pseudo score $S_t := \partial \log f_{\theta_0}(X_t)/\partial\theta$, a natural estimator of that matrix is

$$\hat{J} = \frac{1}{n}\sum_{t=1}^{n} \hat{S}_t \hat{S}_t' \qquad \text{where} \quad \hat{S}_t = \frac{\partial \log f_{\hat{\theta}_n}(X_t)}{\partial \theta}.$$

Estimation of the LRV matrix I is more intricate. In the literature, two types of estimators are generally employed: Heteroskedasticity and Autocorrelation Consistent (HAC) estimators (see Newey and West, 1987, and Andrews, 1991, for general references; see Francq and Zakoian, 2007, for an application to testing strong linearity in weak ARMA models) and spectral density estimators (see e.g. de Haan and Levin, 1997, for a general reference and Francq, Roy and Zakoïan, 2005, for weak ARMA models). We will apply the second approach but HAC estimators could also be considered.

Note that, up to the factor 2π, the LRV matrix I is the spectral density at frequency zero of the process (S_t). For the numerical illustrations presented in this paper we used a VAR spectral estimator consisting in: i) fitting VAR(r) models for $r = 0, \ldots, r_{\max}$ to the series $\hat{S}_t$, $t = 1, \ldots, n$; ii) selecting the order r which minimizes an information criterion and estimating I by the matrix $\hat{I}$ defined as 2π times the spectral density at frequency zero of the estimated VAR(r) model. For the numerical illustrations presented in this paper, we used the AIC model selection criterion with $r_{\max} = 25$.

We now give a more precise description of the method and its asymptotic properties. The process (S_t) is both strictly and second-order stationary (by Assumption **A5**). If this process is purely deterministic (see e.g. Brockwell and Davis (1991) p. 189), it thus admits the Wold decomposition $S_t = u_t + \sum_{i=1}^{\infty} B_i u_{t-i}$, where (u_t) is a q-variate weak white noise (that is, a sequence of centered and uncorrelated random variables) with covariance matrix Σ_u. Assume that Σ_u is non singular, that $\sum_{i=1}^{\infty} \|B_i\| < \infty$, and that $\det\left(I_q + \sum_{i=1}^{\infty} B_i z^i\right) \neq 0$ when $|z| \leq 1$. Then (S_t) admits a VAR(∞) representation of the form

$$\mathcal{A}(B)S_t := S_t - \sum_{i=1}^{\infty} A_i S_{t-i} = u_t, \tag{2.2}$$

such that $\sum_{i=1}^{\infty} \|A_i\| < \infty$ and $\det\{\mathcal{A}(z)\} \neq 0$ for all $|z| \leq 1$, and we obtain

$$I = \mathcal{A}^{-1}(1)\Sigma_u \mathcal{A}'^{-1}(1). \tag{2.3}$$

Consider the regression of S_t on $S_{t-1}, \ldots, S_{t-r}$ defined by

$$S_t = \sum_{i=1}^{r} A_{r,i} S_{t-i} + u_{r,t}, \qquad u_{r,t} \perp \{S_{t-1} \cdots S_{t-r}\}. \tag{2.4}$$

The least squares estimators of $\underline{A}_r = (A_{r,1} \cdots A_{r,r})$ and $\Sigma_{u_r} = \mathrm{Var}(u_{r,t})$ are defined by

$$\hat{\underline{A}}_r = \hat{\Sigma}_{\hat{S},\hat{\underline{S}}_r} \hat{\Sigma}_{\hat{\underline{S}}_r}^{-1} \qquad \text{and} \qquad \hat{\Sigma}_{u_r} = \frac{1}{n} \sum_{t=1}^{n} \left(\hat{S}_t - \hat{\underline{A}}_r \hat{\underline{S}}_{r,t}\right) \left(\hat{S}_t - \hat{\underline{A}}_r \hat{\underline{S}}_{r,t}\right)'$$

where $\underline{\hat{S}}_{r,t} = (\hat{S}'_{t-1} \cdots \hat{S}'_{t-r})'$,

$$\hat{\Sigma}_{\hat{S},\underline{\hat{S}}_r} = \frac{1}{n}\sum_{t=1}^{n} \hat{S}_t \underline{\hat{S}}'_{r,t}, \qquad \hat{\Sigma}_{\underline{\hat{S}}_r} = \frac{1}{n}\sum_{t=1}^{n} \underline{\hat{S}}_{r,t}\underline{\hat{S}}'_{r,t},$$

with by convention $\hat{S}_t = 0$ when $t \leq 0$, and assuming $\hat{\Sigma}_{\underline{\hat{S}}_r}$ is non singular (which holds true asymptotically). We are now in a position to give conditions ensuring the consistency of $\hat{I}$ and $\hat{J}$. The proof, which is based on Berk (1974), is not given here, but is available from the authors.

Theorem 2.3. *Let the assumptions of Theorem 2.1 be satisfied. We have $\hat{J} \to J$ a.s. as $n \to \infty$. Assume in addition that the process (S_t) admits the VAR(∞) representation (2.2), where $\|A_i\| = o\left(i^{-2}\right)$ as $i \to \infty$, the roots of $\det(\mathcal{A}(z)) = 0$ are outside the unit disk, and Σ_u is non singular. We also need to complement Assumption* **A4** *by assuming that, with the same notation,*

A4': $\quad E \sup_{\theta \in V(\theta_0)} \left| \frac{\partial}{\partial \theta_i} \left\{ \frac{\partial \log f_\theta(X_1)}{\partial \theta_j} \frac{\partial \log f_\theta(X_1)}{\partial \theta_k} \right\} \right| < \infty,$

and to reinforce Assumption **A5** *by assuming that, for some $\nu > 0$,*

A5': $\quad E \left\| \frac{\partial \log f_{\theta_0}(X_1)}{\partial \theta} \right\|^{4+2\nu} < \infty$ *and* $\sum_{k=0}^{\infty} \{\alpha_X(k)\}^{\frac{\nu}{2+\nu}} < \infty.$

Then, when $r = r(n) \to \infty$ and $r^3/n \to 0$ as $n \to \infty$,

$$\hat{I} := \hat{\mathcal{A}}_r^{-1}(1)\hat{\Sigma}_{u_r}\hat{\mathcal{A}}_r'^{-1}(1) \to I \qquad \textit{in probability.}$$

Remark 2.1. In Theorems 2.2 and 2.3, we considered stationary processes with specified marginal distributions. Examples of dependent processes admitting a given cdf F can be constructed as follows. Take for instance a model of the form $Y_t = G_\theta(Y_{t-1}, \epsilon_t)$ with iid errors (ϵ_t). Under stationary assumptions, for any error distribution there exists a unique invariant marginal cdf F_Y for Y_t. For ease of presentation, let us assume that F and F_Y are invertible functions. Then, the process $X_t = F^{-1}\{F_Y(Y_t)\}$ is a strictly stationary solution of the model

$$X_t = F^{-1}\{F_Y(G_\theta[F_Y^{-1}\{F(X_{t-1})\}, \epsilon_t])\}$$

with marginal distribution F. This shows the existence of a non trivial stationary process with a specified marginal distribution. In the next section, we study a non iid and non linear stationary process admitting a marginal standard Gaussian distribution.

2.3 A GMM point of view

In general, our estimation problem could be reformulated in terms of Generalized Method of Moments (GMM) estimation, using the first-order condition

$$E\left[\frac{\partial \log f_{\theta_0}}{\partial \theta}(X_t)\right] = 0. \tag{2.5}$$

In fact, we do not use such first-order conditions for the consistency of our estimator. In Theorem 2.1, the consistency is established under **A1-A3**, that is, without any differentiability assumption on the density f_θ (continuity suffices). Under the assumptions of Theorem 2.1, however, the moment condition (2.5) holds and, in a GMM perspective, additional moments could be introduced to achieve efficiency gains. To illustrate this, consider estimating marginal Gaussian distributions. We follow the approach developed by Bontemps and Medahi (2005) for testing normality. Standard Gaussian distributions are characterized by the equalities

$$E[H_i(X)] = 0 \quad \text{for all } i > 0 \tag{2.6}$$

where the H_i are Hermite polynomials recursively defined by

$$H_0(x) = 1, \quad H_1(x) = x, \quad H_i(x) = \frac{1}{\sqrt{i}}\{xH_{i-1}(x) - \sqrt{i-1}H_{i-2}(x)\}.$$

We also have, using the Kronecker symbol δ_{ij},

$$E[H_i(X)H_j(X)] = \delta ij \quad \text{for all } i, j \geq 0. \tag{2.7}$$

Now assume that $X_1, \ldots, X_n$ are (possibly dependent) observations from the $\mathcal{N}(m_0, \sigma_0^2)$ distribution, with $\theta_0 = (m_0, \sigma_0^2)' \in \mathbb{R} \times \mathbb{R}^+$. Let for $p \geq 1$,

$$g_n(\theta) = \frac{1}{n} \sum_{t=1}^{n} H\left(\frac{X_t - m}{\sigma}\right), \quad \text{where } H(x) = (H_1(x), \ldots, H_p(x))'. \tag{2.8}$$

A GMM estimator of θ_0 based on the first p equalities in (2.6) is any measurable solution of

$$\hat{\theta}_n^W = \arg\min_{\theta \in \Theta} g_n(\theta)' W g_n(\theta) \tag{2.9}$$

for some positive definite weighting matrix W. For $p = 2$ we retrieve the QMMLE $\hat{\theta}_n$, whatever W, because $g_n(\hat{\theta}_n) = 0$. In the iid setting, an optimal weighting matrix is the identity matrix. Under appropriate assumptions (see Hansen (1982)), the GMM estimator is consistent and asymptotically normal:

$$\sqrt{n}(\hat{\theta}_n^W - \theta_0) \xrightarrow{d} \mathcal{N}(0, \Sigma(W)) \text{ as } n \to \infty,$$

where $\Sigma(W) = \{GWG'\}^{-1} GWVWG' \{GWG'\}^{-1}$ and

$$G = E\left(\frac{\partial H'\left(\frac{X_t - m_0}{\sigma_0}\right)}{\partial \theta}\right), \quad V = \sum_{h=-\infty}^{\infty} \text{Cov}\left\{H\left(\frac{X_t - m_0}{\sigma_0}\right), H\left(\frac{X_{t-h} - m_0}{\sigma_0}\right)\right\}.$$

The optimal GMM estimator is obtained for $W = V^{-1}$ and its asymptotic variance is $\Sigma^* = \{GV^{-1}G'\}^{-1}$. For $i \geq 1$ we have $\partial H_i(x)/\partial x = \sqrt{i} H_{i-1}(x)$. In view of (2.6) and (2.7), it follows that

$$G = -\begin{pmatrix} \frac{1}{\sigma_0} & 0 & 0 & \ldots & 0 \\ 0 & \frac{1}{\sqrt{2}\sigma_0^2} & 0 & \ldots & 0 \end{pmatrix}.$$

Thus, for the GMM estimator defined in (2.8)-(2.9), the optimal asymptotic covariance matrix takes the form

$$\Sigma^* = \sigma_0^2 \begin{pmatrix} V^{11} & \frac{1}{\sqrt{2}\sigma_0} V^{12} \\ \frac{1}{\sqrt{2}\sigma_0} V^{21} & \frac{1}{2\sigma_0^2} V^{22} \end{pmatrix}^{-1}, \quad \text{where } V^{-1} = (V^{ij}). \tag{2.10}$$

In particular, if $V = (V_{ij})$ is diagonal

$$\Sigma^* = \sigma_0^2 \begin{pmatrix} V_{11} & 0 \\ 0 & 2\sigma_0^2 V_{22} \end{pmatrix} = \begin{pmatrix} \sum_{h=0}^{\infty} \gamma(h) & 0 \\ 0 & \sum_{h=0}^{\infty} \gamma_2(h) \end{pmatrix}$$

where γ is the autocovariance function of (X_t) and γ_2 is the autocovariance function of $(X_t - m_0)^2$. Interestingly, when V is diagonal, Σ^* is independent of the number p of moments used for the GMM method. In other words, no asymptotic efficiency gains can be obtained from taking $p > 2$. But for $p = 2$, the GMM estimator coincides with the QMMLE. An example of process such that V is a diagonal matrix is the Gaussian AR(1) (see Bontemps and Meddahi, 2005, p.157). For the reader's convenience, we summarize these results in the next proposition.

Proposition 2.1. *Let (X_t) denote a stationary process with marginal Gaussian $\mathcal{N}(m_0, \sigma_0^2)$ distribution. Then, under conditions ensuring the asymptotic normality of GMM estimators (see Hansen, 1982), the asymptotic variance of the optimal GMM estimator is given by (2.10). For a Gaussian AR(1) process, more generally when V is diagonal, the QMMLE coincides with the optimal GMM for any number of moments p in (2.8).*

Simulations confirmed this proposition: for moderate and large sample size, no efficiency gains are reached from using $p > 2$. For nonlinear models, however, the matrix V is in general non diagonal and efficiency gains might be obtained. Let us consider the "absolute" autoregression

$$Y_t = \phi |Y_{t-1}| + \epsilon_t, \tag{2.11}$$

where (ϵ_t) is an iid sequence of $\mathcal{N}(0, 1)$ variables. Under the condition $|\phi| < 1$, there exists a strictly stationary solution (Y_t) which, for $\phi < 0$, admits the density

$$h_Y(x) = [2(1 - \phi^2)/\pi]^{1/2} \exp\left\{-\frac{1}{2}(1 - \phi^2)x^2\right\} \Phi(\phi x),$$

where Φ denotes the cdf of the standard Gaussian distribution (see Anděl, Netuka and Zvára (1984)). Let F_Y denote the marginal cdf of Y_t. It follows that, for any m and any $\sigma > 0$, the process $X_t = m + \sigma\Phi^{-1}\{F_Y(Y_t)\}$ is strictly stationary and has a marginal $\mathcal{N}(m, \sigma^2)$ distribution. To compare the performance of the QMMLE and the GMM estimators, we simulated $N = 1,\ 000$ independent trajectories of size $n = 1,\ 000$ of X_t, with $\phi = 0$ and 0.5, $m = 0$ and $\sigma = 1$. For the GMM estimators, we used $p = 4$ Hermite polynomials and two weighting matrices: GMM_I (resp. $\text{GMM}_{\hat{\Sigma}}$) denotes the GMM estimator with W equal to the identity matrix (resp. the estimated optimal weighting matrix). Results reported in Table 1 do not show much difference between the three estimators. For estimating the optimal weighting matrix we used the spectral estimator described in Section 2.2. We also tried several versions of the HAC estimators proposed by Andrews (1991), but the results remained qualitatively unchanged. Other sample size and parameter values lead to similar conclusions.

3 Application to heavy-tailed distributions

We now apply the general results of the previous section to three important classes of distributions.

3.1 Estimating stable marginal distributions

Assume that (X_t) has a univariate stable distribution $S(\theta)$, $\theta = (\alpha, \beta, \sigma, \mu)$, with tail exponent $\alpha \in (0, 2]$, parameter of symmetry (or skewness) $\beta \in [-1, 1]$, scale parameter $\sigma \in (0, \infty)$, and location parameter $\mu \in \mathbb{R}$. This class of density coincides with all the possible non degenerated limit distributions

Table 1: Sampling distribution, , over $n = 1,\ 000$ replications of 3 estimators of the $\mathcal{N}(0,1)$ marginal distribution of $X_t = \Phi^{-1}\{F_Y(Y_t)\}$, where Y_t is simulated from Model (2.11).

θ	ϕ	Method	bias	RMSE	min	Q_1	Q_2	Q_3	max
				$n = 1\,000$					
$m = 0$	0.0	QMMLE	-0.001	0.031	-0.134	-0.023	-0.002	0.018	0.104
		GMM_I	-0.001	0.031	-0.136	-0.024	-0.002	0.018	0.104
		$\text{GMM}_{\hat{\Sigma}}$	-0.001	0.031	-0.132	-0.023	-0.002	0.018	0.111
	0.5	QMMLE	-0.002	0.037	-0.137	-0.027	-0.001	0.024	0.116
		GMM_I	-0.002	0.038	-0.145	-0.027	-0.001	0.023	0.115
		$\text{GMM}_{\hat{\Sigma}}$	0.002	0.039	-0.146	-0.024	0.002	0.027	0.115
$\sigma^2 = 1$	0.0	QMMLE	0.001	0.046	0.843	0.969	1.001	1.031	1.143
		GMM_I	0.006	0.048	0.843	0.972	1.007	1.035	1.142
		$\text{GMM}_{\hat{\Sigma}}$	-0.009	0.048	0.840	0.958	0.991	1.024	1.138
	0.5	QMMLE	0.000	0.051	0.826	0.966	0.999	1.034	1.179
		GMM_I	0.006	0.053	0.825	0.970	1.004	1.041	1.179
		$\text{GMM}_{\hat{\Sigma}}$	-0.012	0.054	0.813	0.953	0.986	1.023	1.173

RMSE is the Root Mean Square Error, Q_i, $i = 1, 3$, denote the quartiles.

for standardized sums of iid random variables of the form $a_n^{-1}\sum_{i=1}^n Z_i - b_n$, where (a_n) and (b_n) are sequences of constants with $a_n > 0$. The location and scale parameters are such that $Y = \sigma X + \mu$, $\sigma > 0$, follows a stable distribution of parameter $(\alpha, \beta, \sigma, \mu)$ when X follows a stable distribution of parameter $(\alpha, \beta, 1, 0)$. In general, the density $f_\theta(x)$ of a stable distribution is not known explicitly, but the characteristic function $\phi(s) = \phi_{\alpha,\beta}(s)$ of a stable distribution of parameter $(\alpha, \beta, 1, 0)$ is defined by

$$\log \phi(s) = -|s|^\alpha \left\{1 + i\beta\,(\text{sign } s)\tan\left(\frac{\pi\alpha}{2}\right)\left(|s|^{1-\alpha} - 1\right)\right\}$$

if $\alpha \neq 1$ and

$$\log \phi(s) = -|s| \left\{1 + i\beta\,(\text{sign } s)\,\frac{2}{\pi}\log|s|\right\}$$

if $\alpha = 1$. There exist other parameterizations for the stable characteristic function, but this parameterization presents the advantage that

$$f_\theta(x) := (2\pi)^{-1} \int_{\mathbb{R}} \exp\{-is(x-\mu)\}\, \phi_{\alpha,\beta}(\sigma s) ds$$

is differentiable with respect to both $x \in \mathbb{R}$ and $\theta \in \Lambda := (0,2) \times (-1,1) \times (0,\infty) \times \mathbb{R}$ (see Nolan, 2003). Let $f_{\alpha,\beta}$ be the stable density of parameter $\theta = (\alpha, \beta, 1, 0)$. Because $f_{\alpha,\beta}(x)$ is real and $\phi(-s) = \overline{\phi(s)}$, we have

$$f_{\alpha,\beta}(x) = \frac{1}{\pi} \int_0^\infty e^{-s^\alpha} \cos\left\{ sx + \beta \tan\left(\frac{\pi\alpha}{2}\right)(s - s^\alpha) \right\} ds \tag{3.1}$$

for $\alpha \neq 1$, and

$$f_{\alpha,\beta}(x) = \frac{1}{\pi} \int_0^\infty e^{-s} \cos\left(sx + s\beta \frac{2}{\pi} \log s \right) ds \tag{3.2}$$

for $\alpha = 1$. From these expressions and the elementary series expansion $(1 - s^{\alpha-1}) \tan\left(\frac{\pi\alpha}{2}\right) = \frac{2}{\pi} \log s + o(\alpha - 1)$, the continuity at $\alpha = 1$ is clear.

Note that $f_\theta(x) = \sigma^{-1} f_{\alpha,\beta}\{\sigma^{-1}(x-\mu)\}$ can be numerically evaluated from (3.1)-(3.2), or alternatively using the function `dstable()` of the `R` package `fBasics`.

A stable distribution with exponent $\alpha = 2$ is a Gaussian distribution, a stable distribution with $\alpha < 2$ has infinite variance. The parameter α determines the tail of the distribution of $X \sim S(\theta)$ in the sense that, when $\alpha < 2$, $F_\theta(-x) := P(X < -x)$ and $1 - F_\theta(x)$ are equivalent to $C_\alpha(1-\beta)x^{-\alpha}$ and $C_\alpha(1+\beta)x^{-\alpha}$, respectively, as $x \to \infty$, with $C_\alpha > 0$. Moreover, still when $X \sim S(\theta)$ with $\alpha < 2$,

$$E|X|^p < \infty \quad \text{if and only if} \quad p < \alpha. \tag{3.3}$$

Theorem 3.1. *Assume that Θ is a compact subset of Λ and that $\theta_0 \in \Theta$. If (X_t) is a stationary and ergodic process whose marginal follows a stable distribution $S(\theta_0)$, then the QMMLE defined by (2.1) is such that $\hat{\theta}_n \to \theta_0$ a.s. If, in addition, $\theta_0 \in \overset{\circ}{\Theta}$ and there exists $\varepsilon \in (0,1)$ such that $\sum_{k=0}^{\infty} \{\alpha_X(k)\}^{1-\varepsilon} < \infty$, then*

$$\sqrt{n}\left(\hat{\theta}_n - \theta_0\right) \xrightarrow{d} \mathcal{N}(0, J^{-1}IJ^{-1}) \text{ as } n \to \infty,$$

where I and J are defined in **A4**.

We now show how to use the estimators $\hat{I}$ and $\hat{J}$ defined in Theorem 2.3 in the alpha-stable case. Since the alpha-stable densities and their derivatives are not explicit, we need to define a way to compute $\hat{S}_t$. By continuity, set $g_\alpha(s) = \tan(\pi\alpha/2)(s - s^\alpha)$ when $\alpha \neq 1$ and $g_\alpha(s) = (2s/\pi)\log s$ when $\alpha = 1$. Let $\psi_{\alpha,\beta}(x,s) = sx + \beta g_\alpha(s)$. By the arguments given in the proof of Theorem 3.1, differentiations of (3.1) under the integral sign yield

$$\begin{aligned}
\frac{\partial f_\theta(x)}{\partial \alpha} &= \frac{-1}{\sigma\pi}\int_0^\infty s^\alpha e^{-s^\alpha}\varphi_{\alpha,\beta}\left(\frac{x-\mu}{\sigma}\right)ds,\\
\frac{\partial f_\theta(x)}{\partial \beta} &= \frac{-1}{\sigma\pi}\int_0^\infty e^{-s^\alpha}\sin\psi_{\alpha,\beta}\left(\frac{x-\mu}{\sigma}, s\right)g_\alpha(s)ds,\\
\frac{\partial f_\theta(x)}{\partial \sigma} &= \frac{-1}{\sigma}f_\theta(x) + \frac{1}{\sigma^3\pi}\int_0^\infty s(x-\mu)e^{-s^\alpha}\sin\psi_{\alpha,\beta}\left(\frac{x-\mu}{\sigma}, s\right)ds,\\
\frac{\partial f_\theta(x)}{\partial \mu} &= \frac{1}{\sigma^2\pi}\int_0^\infty s e^{-s^\alpha}\sin\psi_{\alpha,\beta}\left(\frac{x-\mu}{\sigma}, s\right)ds,
\end{aligned}$$

with $\varphi_{\alpha,\beta}(x)$ is equal to

$$(\log s)\cos\psi_{\alpha,\beta}(x,s) - \beta\sin\psi_{\alpha,\beta}(x,s)\left\{(\log s)\tan\left(\frac{\pi\alpha}{2}\right) - \frac{\pi\left(s^{1-\alpha}-1\right)}{2\cos^2(\frac{\pi\alpha}{2})}\right\}$$

when $\alpha \neq 1$ and equal to $(\log s)\cos\psi_{1,\beta}(x,s) - (\beta/\pi)(\log s)^2\sin\psi_{1,\beta}(x,s)$ when $\alpha = 1$. These derivatives allow to compute the $\hat{S}_t$'s required for the estimators of I and J.

Proposition 3.1. *Under the assumptions of Theorem 3.1, Assumptions* **A4'** *and* **A5'** *are satisfied. Thus the consistency of $\hat{I}$ and $\hat{J}$ holds under the other assumptions of Theorem 2.3.*

3.2 Estimating generalized Pareto distributions

The GPD(γ_0, σ_0) with shape parameter $\gamma_0 \in \mathbb{R}$ and scale parameter $\sigma_0 > 0$, has the probability distribution function

$$F_{\gamma_0,\sigma_0}(x) = \begin{cases} 1 - \left(1 + \gamma_0 \frac{x}{\sigma_0}\right)^{-1/\gamma_0}, & \gamma_0 \neq 0, \\ 1 - \exp\left(-\frac{x}{\sigma_0}\right), & \gamma_0 = 0, \end{cases}$$

where for $\gamma_0 \geq 0$ the range is $x \geq 0$, while for $\gamma_0 < 0$ the range is $0 \leq x \leq -\sigma_0/\gamma_0$.

One attractive feature of the GPD is that it is stable with respect to "excess over threshold operations": if $X \sim GPD(\gamma_0, \sigma_0)$, then the distribution of $X - u$ conditional on $X > u$ is the GPD($\gamma_0, \sigma_0 + \gamma_0 u$). Moreover, when $\gamma_0 > 0$ the upper tail probability $P(X > x)$ of the GPD(γ_0, σ_0) behaves like $kx^{-\alpha}$ for large x, with $\alpha = 1/\gamma_0$, so that $1/\gamma_0$ is the tail index, comparable to α of the stable distribution. Note also that $E(X^s) < \infty$ for $s < 1/\gamma_0$. However, unlike the Pareto distribution, the GPD permits Paretian tail behavior with $\alpha \geq 2$. The GPD plays an important role in EVT. Indeed, it has been shown by Balkema and de Haan (1974) and Pickands (1975) that, for any random variable X whose distribution belongs to the maximum domain of attraction of an extreme value distribution, the law of the excess $X - u$ over a high threshold u, often called Peak Over Threshold (POT), is well approximated by a GPD($\gamma_0, \sigma_0(u)$) (see Theorem 3.4.13 in Embrechts, Klüppelberg and Mikosch, 1997).

Many approaches have been proposed to estimate the GPD (see the review by de Zea Bermudez and Kotz (2010)). Let $\theta_0 = (\gamma_0, \sigma_0)$ be the true parameter value of the GPD(γ_0, σ_0), where $\gamma_0, \sigma_0 > 0$. Let Θ denote a compact subset of $(0, \infty)^2$. The QMMLE is any measurable solution of (2.1) with, for $\theta = (\gamma, \sigma) \in \Theta$,

$$\ell_n(\theta) = \log \sigma^2 + \frac{1}{n}\left(\frac{1}{\gamma} + 1\right)\sum_{t=1}^{n} \log\left(\frac{\gamma X_t}{\sigma} + 1\right)^2.$$

Theorem 3.2. *If (X_t) is a stationary and ergodic process whose marginal follows a GPD(θ_0), then the QMMLE defined by (2.1) is such that $\hat{\theta}_n \to \theta_0$ a.s. If, in addition, $\theta_0 \in \overset{\circ}{\Theta}$ and there exists $\varepsilon \in (0,1)$ such that $\sum_{k=0}^{\infty} \{\alpha_X(k)\}^{1-\varepsilon} < \infty$, then*

$$\sqrt{n}\left(\hat{\theta}_n - \theta_0\right) \xrightarrow{d} \mathcal{N}(0, J^{-1} I J^{-1}) \text{ as } n \to \infty,$$

where I is defined in **A4** *and*

$$J^{-1} = \begin{pmatrix} (1+\gamma_0)^2 & -\sigma_0(1+\gamma_0) \\ -\sigma_0(1+\gamma_0) & 2\sigma_0^2(1+\gamma_0) \end{pmatrix}.$$

A drawback of the GPD, for instance in the aim of modeling log-returns distributions, is that its density is not positive over the real line. A simple extension of the GPD(γ_0, σ_0) is defined by the following density, which we can call double GPD$(\tau, \gamma_1, \sigma_1, \gamma_2, \sigma_2)$:

$$f_{\theta_0}(z) = \tau \frac{\sigma_1^{1/\gamma_1}}{(-\gamma_1 z + \sigma_1)^{1+1/\gamma_1}} \mathbb{1}_{z<0} + (1-\tau)\frac{\sigma_2^{1/\gamma_2}}{(\gamma_2 z + \sigma_2)^{1+1/\gamma_2}} \mathbb{1}_{z \geq 0} \qquad (3.4)$$

where $\theta_0 = (\tau, \gamma_1, \sigma_1, \gamma_2, \sigma_2)' \in \Theta$ where Θ denotes a compact subset of $[0,1] \times (0, \infty)^4$. A straightforward extension of Theorem 3.2, whose proof is omitted, is the following.

Theorem 3.3. *If (X_t) is a stationary and ergodic process whose marginal follows a double GPD(θ_0), then the QMMLE defined by (2.1) is such that $\hat{\theta}_n \to \theta_0$ a.s. If, in addition, $\theta_0 \in \overset{\circ}{\Theta}$ and there exists $\varepsilon \in (0,1)$ such that $\sum_{k=0}^{\infty} \{\alpha_X(k)\}^{1-\varepsilon} < \infty$, then*

$$\sqrt{n}\left(\hat{\theta}_n - \theta_0\right) \xrightarrow{d} \mathcal{N}(0, J^{-1} I J^{-1}) \text{ as } n \to \infty,$$

where I is defined in **A4** *and for $i = 1, 2$,*

$$J^{-1} = \begin{pmatrix} \tau(1-\tau) & 0 & 0 \\ 0 & \tau^{-1} J_1^{-1} & 0 \\ 0 & 0 & (1-\tau)^{-1} J_2^{-1} \end{pmatrix}, \quad J_i^{-1} = \begin{pmatrix} (1+\gamma_i)^2 & -\sigma_i(1+\gamma_i) \\ -\sigma_i(1+\gamma_i) & 2\sigma_i^2(1+\gamma_i) \end{pmatrix},$$

3.3 Estimating generalized extreme value distributions

We now consider another class of densities which is widely used in EVT. It is known (see *e.g.* Beirlant et al. 2005) that the possible limiting distributions for the maximum $X_{(n)}$ of a sample $X_1, \ldots, X_n$ are given by the class of the GEV whose densities are of the form

$$f_\theta(x) = \frac{1}{\sigma}\left\{1 + \gamma\left(\frac{x-\mu}{\sigma}\right)\right\}^{-1/\gamma - 1} e^{-\left\{1+\gamma\left(\frac{x-\mu}{\sigma}\right)\right\}^{-1/\gamma}} \mathbb{1}_{\{1+\gamma(x-\mu)/\sigma > 0\}},$$

with $\theta = (\mu, \sigma, \gamma) \in \mathbb{R} \times \mathbb{R}^+ \times \mathbb{R}$. Taking the limit, when $\gamma = 0$ the density is

$$f_\theta(x) = \sigma^{-1} e^{-(x-\mu)/\sigma} e^{-e^{-(x-\mu)/\sigma}}.$$

The density is called Weilbull, Gumbel or Fréchet when the shape parameter γ is respectively negative, null or positive. When the X_i's have Pareto tails of index $\alpha > 0$, the limiting distribution of $X_{(n)}$ as $n \to \infty$ is a Fréchet distribution with shape parameter $\gamma = 1/\alpha$. Let $\theta_0 = (\mu_0, \sigma_0, \gamma_0)$ be the true parameter value of the GEV(θ_0), where θ_0 belongs to a compact subset Θ of $\mathbb{R} \times \mathbb{R}^+ \times (\underline{\gamma}, \infty)$. We impose the constraint $\gamma_0 > \underline{\gamma}$ because, as shown by Smith (1985) in the iid case, the information matrix J does not exist when $\gamma_0 \leq -1/2$.

Theorem 3.4. *If (X_t) is a stationary and ergodic process whose marginal follows a GEV(θ_0), and if $\underline{\gamma} \geq -1$ then the QMMLE defined by (2.1) is such that $\hat{\theta}_n \to \theta_0$ a.s. If, in addition, $\theta_0 \in \overset{\circ}{\Theta}$, $\underline{\gamma} \geq -1/2$ and there exists $\varepsilon \in (0,1)$ such that $\sum_{k=0}^{\infty} \{\alpha_X(k)\}^{1-\varepsilon} < \infty$, then*

$$\sqrt{n}\left(\hat{\theta}_n - \theta_0\right) \xrightarrow{d} \mathcal{N}(0, J^{-1}IJ^{-1}) \text{ as } n \to \infty,$$

where I and J are defined in **A4**.

4 Modeling the unconditional distribution of daily returns

In this section, we consider an application to the marginal density of financial returns. We focus on two aspects of the shape of daily returns distributions, both widely discussed in the empirical finance literature, the asymmetry and the tail thickness.

Daily returns distribution are generally considered as approximately symmetric (see e.g. Taylor, 2007) but several studies documented the fact that they can be positively skewed (see e.g. Kon (1984)). Symmetry tests are generally based on the skewness coefficient, and the critical value is routinely obtained by assuming a sample from a normal distribution. In the symmetry test proposed by Premaratne and Bera (2005), the normality is replaced by a distribution that takes into account leptokurtosis explicitly, but the iid assumption is maintained. In the framework of this paper, we can test for asymmetry under general distributional assumptions, while allowing for serial dependence of observations.

By graphical methods, Mandelbrot (1963) showed that daily price changes in cotton have heavy tails with $\alpha \approx 1.7$, so that the mean exists but the variance is infinite. To mention only a few more recent studies, using the Hill estimator Jansen and de Vries (1991) found estimated values of α between 3 and 5 using the order statistics, for daily data of ten stocks from the S&P100 list and two indices. With the same estimator, Loretan and Phillips (1994) found estimated values of α between 2 and 4, for a daily and monthly returns from numerous stock indices and exchange rates, indicating that the variance of the price returns are finite but the fourth-order moments are not. The modified Hill estimator proposed by Huisman, Koedijk, Kool and Palm (2001) suggests higher α estimates. Using a MLE approach, McCulloch (1996) reestimated the coefficient α on the same data as Jansen and de Vries (1991) and Loretan and Phillips (1994), and found values between 1.5 and 2. By the same technique, using fast Fourier transforms to approximate the α-stable density, Rachev and Mittnik (2000) obtained values of α between 1 and 2, for a variety of stocks, stock indices and exchange rates.

The above-mentioned references show that the debate concerning the tail index α of the financial returns is not over. The estimated value of α seems to be very sensitive to the estimation method.[2]

In this paper, we participate in the debate on the typical value of α and the possible asymmetry of the marginal distribution of financial returns, by fitting alpha-stable, GPD and GEV distributions to daily returns of stock indices, using the QMMLE. We consider nine major world stock indices: CAC (Paris), DAX (Frankfurt), FTSE (London), Nikkei (Tokyo), NSE (Bombay), SMI (Switzerland), SP500 (New York), SPTSX (Toronto), and SSE (Shanghai). The observations cover the period from January, 2 1991 to August, 26 2011 (except for the NSE, SPTSX and SSE whose first observations are posterior to 1991). The period includes the recent sovereign-debt crises in Europe and US. We checked that the results presented below are not changed much by withdrawing this recent turbulent period.

4.1 Fitting alpha-stable distributions to the series

Table 2 shows that the tail index estimated when fitting alpha-stable distributions is always between 1.5 and 1.7, for all the series, which is comparable with the values found by Mandelbrot (1963), Leitch and Paulson (1975), McCulloch (1996) or Rachev and Mittnik (2000). It is interesting to note that all distributions are negatively skewed ($\beta < 0$). Table 3 shows that, for all but one returns the distribution is significantly asymmetric. Table 4 shows that the estimated value $\hat{\mu}$ of the position parameter is often significantly positive. It should be however underlined that these results are valid under

[2]Several methods based on EVT have been proposed for the sole estimation of the tail parameter α, mainly in the iid case (see Beirlant, Vynckier and Teugels (1996), Einmahl, Li and Liu (2009) and the references therein).

Table 2: Stable distributions fitted by QMMLE on daily stock market returns. The estimated standard deviation are displayed into brackets.

Index	$\hat{\alpha}$	$\hat{\beta}$	$\hat{\sigma}$	$\hat{\mu}$
CAC	1.72 (0.07)	-0.19 (0.05)	0.81 (0.03)	0.07 (0.02)
DAX	1.64 (0.07)	-0.17 (0.05)	0.79 (0.04)	0.09 (0.02)
FTSE	1.70 (0.06)	-0.19 (0.04)	0.62 (0.02)	0.07 (0.01)
Nikkei	1.65 (0.05)	-0.14 (0.03)	0.79 (0.03)	0.05 (0.02)
NSE	1.60 (0.09)	-0.21 (0.07)	0.87 (0.05)	0.17 (0.04)
SMI	1.66 (0.06)	-0.22 (0.05)	0.64 (0.02)	0.09 (0.02)
SP500	1.62 (0.05)	-0.10 (0.03)	0.50 (0.01)	0.05 (0.01)
SPTSX	1.55 (0.11)	-0.25 (0.05)	0.60 (0.03)	0.11 (0.02)
SSE	1.54 (0.06)	-0.12 (0.07)	0.83 (0.03)	0.09 (0.04)

Table 3: p-values for the t-test of $H_0 : \beta = 0$ against $\beta \neq 0$.

CAC	DAX	FTSE	Nikkei	NSE	SMI	SP500	SPTSX	SSE
0.000	0.000	0.000	0.000	0.002	0.000	0.001	0.000	0.080

the assumption that the marginal distribution belongs to the class of the alpha-stable distributions.

4.2 Fitting double GPD to double POT

It is worth studying the sensitivity of the results to a change of distribution. According to the EVT, the tail index of a series of returns r_t should also be well estimated by fitting a GPD to the POT's $\{r_t - u : r_t > u\}$. In order

Table 4: p-values for the t-test of $H_0 : \mu = 0$ against $\mu > 0$.

CAC	DAX	FTSE	Nikkei	NSE	SMI	SP500	SPTSX	SSE
0.001	0.000	0.000	0.003	0.000	0.000	0.000	0.000	0.012

to estimate indices for both the positive and negative tails, we fitted double GPD distributions to $\{r_t - u : r_t > u\} \cup \{r_t + u : r_t < -u\}$, for the different series r_t of returns considered in Table 2. The choice of the threshold u is crucial. If u is chosen too small, estimation biases may occur due to the inadequacy of the GPD distribution for the whole data set. If u is chosen too large, the variance of the estimates is likely to be too large because of the small number of tail observations.

In order to propose a practical choice for the threshold, we conducted the following experiment. Let k be a positive integer, and let (η_t) be an iid sequence of alpha-stable distribution $S(\theta_k)$. Assume $\theta_k = (\alpha, 0, k^{-1/\alpha}, 0)$, *i.e.* the location parameter is $\mu = 0$, the symmetry parameter is $\beta = 0$ and the scale parameter is $\sigma = k^{-1/\alpha}$. For any $k \geq 1$, the moving average process

$$X_t = \sum_{i=1}^{k} \eta_{t+1-i} \tag{4.1}$$

has the marginal distribution $S(\theta)$, with $\theta = (\alpha, 0, 1, 0)$. For the numerical illustrations we took $\alpha = 1.6$, which is a value close to the estimated values in Table 2. Even if the marginal distribution does not vary with k, the dynamics of the k-dependent process (X_t) strongly depends on k (Figure 3).

We simulated $1,000$ independent realizations of length $n = 4,000$ of Model (4.1). The sample size $n = 4,000$ is a typical sample size for the daily series considered in Table 2. On each series, we fitted a double GPD, whose density is displayed in (3.4), to the proportion π of the data with largest absolute values. Figure 4 shows, in function of π, the bias and root

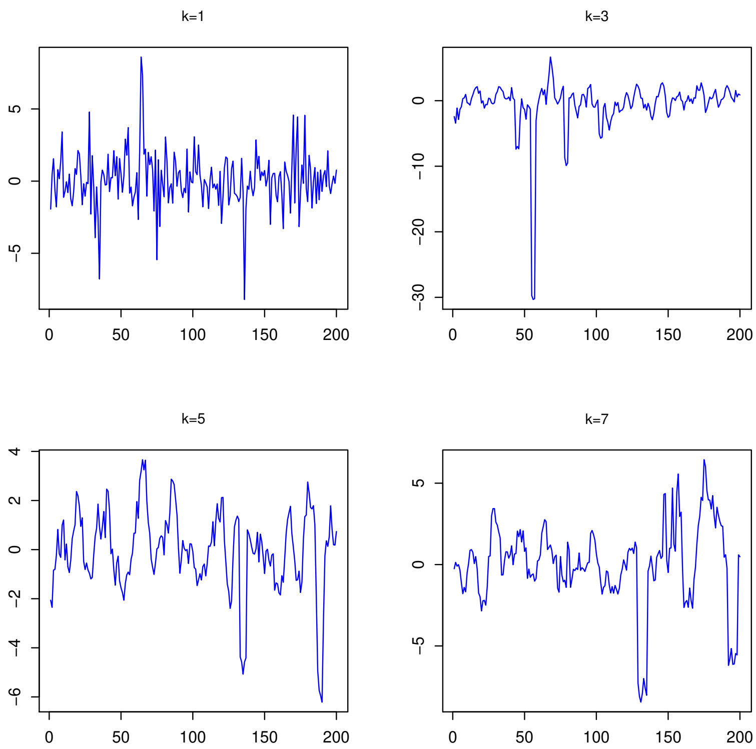

Figure 3: Trajectories of the moving average (4.1) of order k for different values of k.

mean squared error (RMSE) of estimation of the tail parameter $\alpha_2 := 1/\gamma_2$ of the positive tail. We do not present the graph of the RMSE of estimation of $\alpha_1 := 1/\gamma_1$, which is obviously very similar to that of Figure 4. For computing these RMSE's we used 5%- trimmed means, which eliminate few simulations for which the estimate of γ_2 is close to zero (and thus the estimate of α is clearly not compatible with that of a stable distribution). It can be seen that the bias and RMSE's tend to increase with the degree k of dependence. Interestingly, the shapes of the curves are however similar for the different values of k, with a minimum corresponding to π of about 12.5%. We thus decided to define the threshold u as being the quantile of order 87.5% of the absolute values of the returns. We then adjusted double GPD's on the subset of the returns with absolute value greater than u. Table 5 displays the values of the QMMLE for the nine series of returns. The most noticeable output is that the estimated standard deviations of $\hat{\alpha}_1$, and to a lesser extent $\hat{\alpha}_2$, are very high, ruling out any clear conclusion concerning the tail index parameters. We tried other values of the threshold, but even for much smaller values of u the estimated standard deviations remained very large.

The POT approach seems difficult to apply to get an accurate estimate of α for typical sample sizes of daily series of returns. A very small proportion of the most extreme observations is required to get a negligible bias, but the RMSE is then relatively large. The estimated values of the other parameters give more conclusive information. Note that if the marginal distribution of the returns was symmetric, one should have $\tau = 1/2$ and $\sigma_1 = \sigma_2$. Table 6 shows that this assumption is often rejected, confirming the outputs of Tables 3 and 4.

From this study, based two large classes of distributions for the daily returns, one can conclude that for general volatility models (i.e. GARCH, stochastic volatility ...) of the form $r_t = \sigma_t \eta_t$ with η_t iid, centered and independent of σ_t, an asymmetric distribution can be recommended for η_t. Indeed, a symmetric distribution for η_t would entail a symmetric distribution for r_t. The commonly used Gaussian, Student distributions, or GED (Generalized Error Distribution), should thus be avoided for η_t.

4.3 Fitting GEV to block maxima

Table 7 displays the estimated tail indices obtained by fitting a GEV on the maxima of blocks of m consecutive returns. The main result of that table

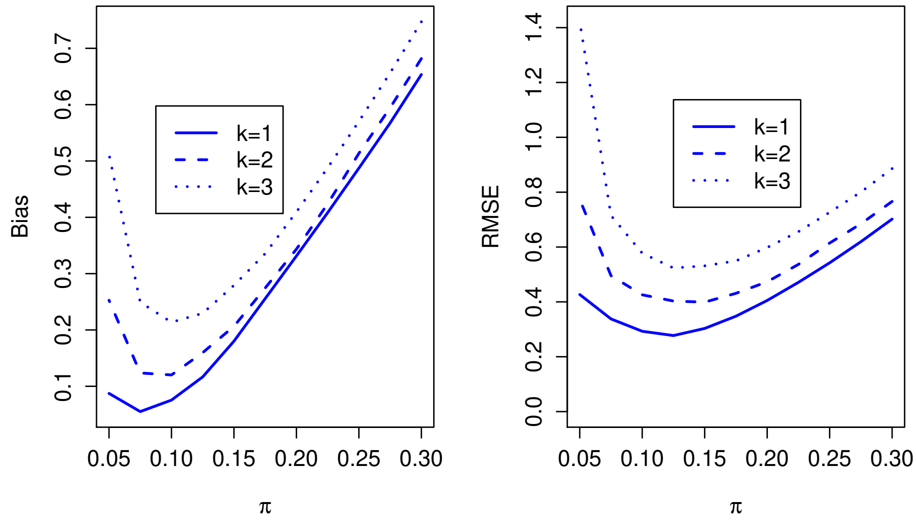

Figure 4: RMSE for the estimates for the positive tail index $\alpha = 1/\gamma_2$ of the process (4.1), when γ_2 is estimated by fitting a double GPD to the proportion π of the data with largest absolute values.

is that the estimated tail indices are around 3, which is much higher than what was obtained by fitting stable distributions. This is not very surprising since, under the Pareto-tail assumption, α is only a tail parameter of the *asymptotic* distribution of the maxima. Observe that when m increases, the estimation of α decreases for all assets and tends to be closer to what was obtained for the stable distribution (in particular for the SMI, 1.89 with the GEV against 1.66 with the stable law). Note also that the estimated standard deviation are large, but do not increase much when m increases (although the number of observations $[n/m]$ decreases). This is certainly due to the fact that, roughly speaking, the dependence of the observations decreases when the size m of the blocks increases.

5 Conclusion

It is often of interest to have information about the marginal distribution of a time series. A typical example is provided by financial series, for which recur-

Table 5: Generalized Pareto distributions fitted by QMMLE on 12.5% of the most extreme daily stock market returns. The estimated standard deviation are displayed into brackets. The estimate of the tail index is NA (not available) when the estimate of GPD parameter γ is not positive.

Index	$\hat{\tau}$	$\hat{\alpha}_1 = 1/\hat{\gamma}_1$	$\hat{\sigma}_1$	$\hat{\alpha}_2 = 1/\hat{\gamma}_2$	$\hat{\sigma}_2$
CAC	0.53 (0.02)	11.16 (13.65)	0.97 (0.13)	3.69 (1.13)	0.73 (0.1)
DAX	0.51 (0.02)	24.72 (51.39)	1.14 (0.12)	3.96 (1.34)	0.76 (0.08)
FTSE	0.52 (0.02)	4.72 (2.33)	0.72 (0.08)	5.5 (2.23)	0.68 (0.08)
Nikkei	0.54 (0.02)	4.57 (1.38)	0.83 (0.07)	6.29 (2.35)	0.91 (0.08)
NSE	0.54 (0.03)	6.68 (4.26)	1.21 (0.15)	5.65 (2.96)	1.1 (0.15)
SMI	0.52 (0.02)	22.09 (45.77)	0.98 (0.12)	3.8 (1.24)	0.66 (0.08)
SP500	0.5 (0.01)	3.81 (0.79)	0.57 (0.05)	5.11 (1.55)	0.59 (0.05)
SPTSX	0.56 (0.03)	5.21 (2.74)	0.93 (0.27)	7 (6.51)	0.87 (0.19)
SSE	0.52 (0.03)	184 (3301.67)	1.3 (0.13)	4.28 (2.22)	0.88 (0.12)

Table 6: p-value for the Wald test of $H_0 : \tau = 0.5$ and $\sigma_1 = \sigma_2$.

CAC	DAX	FTSE	Nikkei	NSE	SMI	SP500	SPTSX	SSE
0.008	0.007	0.163	0.011	0.395	0.016	0.916	0.01	0.049

rent debates concerning the shape of the distributions exist in the literature. In particular, a large literature has been devoted to testing for the presence of heavy tails, and the asymmetry of marginal distributions of stock returns. However, tests developed in the iid framework are abusively applied, without taking into account the dynamics. In this paper we proposed a method for estimating a parametric specification of the marginal distribution, without specifying the dynamics. We showed that the consistency holds under mild conditions. The dynamic plays an important role, however, in the asymptotic distribution of estimators.

In the present work, the marginal density is assumed to belong to a specific class of parametric densities. Goodness-of-fit tests based on non-

Table 7: Estimated tail index α when GEV distributions are fitted by QMMLE on maxima of m consecutive daily stock market returns.

Index	$m = 8$	$m = 16$	$m = 24$	$m = 32$	$m = 40$	$m = 48$
CAC	5.75 (1.39)	4.11 (1.23)	3.63 (1.02)	3.54 (1.22)	3.22 (1.06)	3.26 (1.31)
DAX	5.69 (1.60)	4.60 (1.52)	3.97 (1.38)	3.75 (1.50)	3.68 (1.67)	3.23 (1.53)
FTSE	5.73 (1.12)	3.84 (0.89)	3.65 (0.94)	3.04 (0.94)	2.81 (0.79)	3.30 (1.05)
Nikkei	6.11 (1.26)	4.73 (1.08)	4.67 (1.15)	5.12 (1.44)	5.08 (1.64)	5.10 (1.77)
NSE	6.52 (1.53)	3.09 (0.77)	3.03 (0.85)	2.32 (0.60)	1.76 (0.19)	2.71 (0.16)
SMI	6.81 (1.48)	2.96 (0.71)	3.11 (0.71)	2.89 (0.84)	2.94 (0.78)	1.89 (0.65)
SP500	5.61 (1.19)	4.94 (1.15)	4.10 (1.10)	4.84 (1.45)	5.40 (1.82)	4.88 (1.87)
SPTSX	4.88 (1.94)	3.13 (1.28)	2.57 (1.11)	3.17 (1.45)	3.17 (0.42)	2.78 (1.18)
SSE	9.28 (3.32)	4.73 (1.66)	3.96 (1.44)	3.06 (1.32)	3.41 (1.68)	3.57 (3.07)

parametric kernel density estimators can be used (see Liu and Wu, 2010) to assess a particular form for the marginal density. In future works, we intend to consider the related problem of testing whether the marginal density belongs to a given parametric class.

Appendix: Proofs

A Proof of Theorem 2.1

First note that

$$\hat{\theta}_n = \arg\min_{\theta\in\Theta} Q_n(\theta), \quad \text{with } Q_n(\theta) = \frac{1}{n}\sum_{t=1}^{n} D_t(\theta), \quad D_t(\theta) = \log\frac{f_{\theta_0}(X_t)}{f_\theta(X_t)}. \tag{A.1}$$

Let $V_k(\theta)$ be the open sphere with center θ and radius $1/k$ and let $\tilde{\theta} \in \Theta$, $\tilde{\theta} \neq \theta_0$. Applying Assumption **A3** and the ergodic theorem to the stationary ergodic process $\left\{\inf_{\theta\in V_k(\tilde{\theta})\cap\Theta} D_t(\theta)\right\}_t$ shows that

$$\liminf_{n\to\infty} \inf_{\theta\in V_k(\tilde{\theta})\cap\Theta} Q_n(\theta) \geq E \inf_{\theta\in V_k(\tilde{\theta})\cap\Theta} D_1(\theta). \tag{A.2}$$

By Beppo Levi's theorem, $E\inf_{\theta\in V_k(\tilde{\theta})\cap\Theta} D_1(\theta)$ increases to $ED_1(\tilde{\theta})$ as $k \to \infty$. Moreover, Jensen's inequality and **A2** entail

$$ED_1(\tilde{\theta}) \geq -\log E\frac{f_{\tilde{\theta}}(X_t)}{f_{\theta_0}(X_t)} = -\log \int_E f_{\tilde{\theta}}(x)d\mu(x) = 0$$

with equality iff $\tilde{\theta} = \theta_0$. It follows that for all $\tilde{\theta} \neq \theta_0$, there exists a neighborhood $V(\tilde{\theta})$ of $\tilde{\theta}$ such that

$$\liminf_{n\to\infty} \inf_{\theta\in V(\tilde{\theta})\cap\Theta} Q_n(\theta) > 0 \geq \limsup_{n\to\infty} \inf_{\theta\in V(\theta_0)\cap\Theta} Q_n(\theta), \tag{A.3}$$

where $V(\theta_0)$ is an arbitrary neighborhood of θ_0. The consistency then follows from a standard compactness argument.

The proof of the asymptotic normality rests on the Taylor expansion:

$$0 = \sqrt{n}\frac{\partial \ell_n(\theta_0)}{\partial\theta} + \frac{\partial^2 \ell_n(\theta_n^*)}{\partial\theta\partial\theta'}\sqrt{n}\left(\hat{\theta}_n - \theta_0\right), \text{ with } \|\theta_n^* - \theta_0\| \leq \|\hat{\theta}_n - \theta_0\|. \tag{A.4}$$

The central limit theorem of Herrndorf (1984) and **A5** entail

$$\sqrt{n}\frac{\partial \ell_n(\theta_0)}{\partial\theta} \xrightarrow{d} \mathcal{N}(0, I) \text{ as } n \to \infty.$$

A new Taylor expansion, Assumption **A4**, the consistency of $\hat{\theta}_n$ and the ergodic theorem show that $\frac{\partial^2 \ell_n(\theta_n^*)}{\partial\theta\partial\theta'} \to -J$ a.s.

B Proof of Theorem 2.2

We maintain the notation introduced in the proof of Theorem 2.1. Note that, with probability one, $f_{\theta_0}(X_t) > 0$ for all t. Using **A1*** and the standard convention $D_t(\theta) = +\infty$ when $f_\theta(X_t) = 0$, almost surely the criterion $Q_n(\theta)$ is a continuous function valued in $(-\infty, \infty]$, taking a finite value at θ_0. Therefore $\arg\min_{\theta\in\Theta} Q_n(\theta)$ exists (but is not necessarily unique) with probability one. Thus $\hat{\theta}_n$ is still defined as a measurable solution of (A.1). Now **A3*** entails that $ED_t(\theta_0) = 0$ and $ED_t(\theta) \in (-\infty, \infty]$ for all $\theta \in \Theta$. Applying the ergodic theorem to the stationary ergodic process $\left\{\inf_{\theta\in V_k(\tilde{\theta})\cap\Theta} D_t(\theta)\right\}_t$ whose expectation is defined in $(-\infty, \infty]$ (see Billingsley 1995, pages 284 and 495) we still have (A.2), where the expectation of the right-hand side can be equal to $+\infty$. Finally (A.3) continues to hold, and the consistency follows.

The asymptotic normality is shown as in the proof of Theorem 2.1, on the set of probability one $\cap_{t=1}^{\infty} (X_t \in \mathcal{X})$.

C Proof of Theorem 3.1 and Proposition 3.1

DuMouchel (1973) showed the CAN of the MLE for stable iid variables. Note that DuMouchel used a parametrization with a discontinuity at $\alpha = 1$. With the chosen parameterization, $f_\theta(x)$ is continuous with respect to $\theta \in \Lambda$ for all x and its support is $\mathbb{R}$ (see Nolan, 2003). Assumption **A1** is thus satisfied with $E = \mathbb{R}$. The identifiability assumption **A2** follows from the identifiability of the characteristic function (see Condition 5 in DuMouchel, 1973). Since

$$f_{\theta_0}(x) \sim c_{\theta_0}|x|^{-(\alpha_0+1)} \qquad \text{as} \quad |x| \to \infty \tag{C.1}$$

(see for example Feller, 1975), $|\log f_{\theta_0}(x)| f_{\theta_0}(x) \sim (\alpha_0+1)c_{\theta_0}|x|^{-(\alpha_0+1)} \log|x|$ as $|x| \to \infty$. It follows that $\int_{|x|>A} |\log f_{\theta_0}(x)| f_{\theta_0}(x) dx < \infty$ for A large enough. Moreover $f_{\theta_0}(x)$ is bounded and bounded away from zero on any compact: $0 < m \leq f_{\theta_0}(x) \leq M < \infty$ for all $x \in [-A, A]$. It follows that $\int_{|x|\leq A} |\log f_{\theta_0}(x)| f_{\theta_0}(x)dx < \infty$, and eventually **A3** holds true. The consistency then follows from Theorem 2.1.

From asymptotic expansions in DuMouchel (1973) (see also equations (2.5)-(2.10) in Andrews, Calder and Davis (2009)), there exists a neighborhood $V(\theta_0)$ of θ_0 such that

$$\sup_{\theta\in V(\theta_0)} \left| \frac{\partial^k \log f_\theta(x)}{\partial\theta_{i_1}\partial\theta_{i_k}} \right| = O\left([\log |x|]^k \right), \tag{C.2}$$

as $|x| \to \infty$, for $k \in \{1, 2, 3\}$ and $i_1, \dots, i_k \in \{1, \dots, 4\}$. From (3.1)-(3.2), it is clear that $f_\theta(x)$ admits derivatives of any order with respect to the components of θ, and that these derivatives can be obtained by differentiation under the integral sign. By continuity arguments and the compactness of Θ, the function $f_\theta(x)$, its derivatives and its inverse are bounded uniformly on $\theta \in \Theta$ and $x \in [-A, A]$ for all $A \in \mathbb{R}$. We thus have

$$\int_{-A}^{A} \sup_{\theta\in\Theta} \left| \frac{\partial \log f_\theta(x)}{\partial\theta_i} \right|^\tau f_{\theta_0}(x)dx < \infty$$

for all $\tau \geq 0$ and all $A \geq 0$. The same bound holds when the first-order derivative is replaced by higher-order derivatives. In view of (C.2) with $k = 1$ and (C.1), we also have

$$\int_{(-\infty,-A)\cup(A,\infty)} \sup_{\theta\in\Theta} \left| \frac{\partial \log f_\theta(x)}{\partial\theta_i} \right|^\tau f_\theta(x)dx < \infty$$

for all $\tau \geq 0$. By (C.2) with $k = 2, 3$ the same holds true with second and third order derivatives. It follows that the moments conditions of **A4** are satisfied, in particular the existence of J is established. The invertibility of J is proved by Condition 6 in DuMouchel (1973). By Davydov's inequality (1968), the existence of I is a consequence of the mixing condition and of the fact that $\|\partial \log f_{\theta_0}(X_1)/\partial\theta\|$ admits moment of any order τ. Assumptions **A4** and **A5** are thus satisfied, and the conclusion follows from Theorem 2.1.

Proposition 3.1 is established by the arguments used to show **A4**, in particular (C.1)-(C.2).

D Proof of Theorem 3.2

The theorem is a consequence of Theorem 2.1. Assumption **A1** is satisfied with $E = \mathbb{R}^+$. Assumptions **A2** and **A3** are clearly satisfied, with the density of the GPD(θ) given, for $\gamma, \sigma > 0$, by $f_\theta(z) = \frac{\sigma^{1/\gamma}}{(\gamma z+\sigma)^{1+1/\gamma}}, \quad z \geq 0$. From the second- and third-order derivatives, available as JBES on-line supplement, we have

$$\sup_{\theta\in V(\theta_0)} \left| \frac{\partial \log f_\theta(x)}{\partial\theta_i} \right| = O\left(\log|x|\right), \quad \sup_{\theta\in V(\theta_0)} \left| \frac{\partial^2 \log f_\theta(x)}{\partial\theta_i\partial\theta_j} \right| = O\left(\log|x|\right),$$

as $|x| \to \infty$, for all $i, j \in \{1, \dots, 4\}$. It can be seen that the third-order derivatives are of the same order, from which Assumption **A4** follows. Finally, $\|\partial \log f_\theta(X_1)/\partial\theta\|$ admits moment of any order, and Assumption **A5** is thus satisfied. The formula for J^{-1} is available as JBES on-line supplement.

E Proof of Theorem 3.4

Note that Theorem 2.1 does not apply here because the support of f_θ depends on θ. We will therefore apply Theorem 2.2. Note that $f_\theta(x) \sim \sigma^{-1}y^{-1/\gamma-1}$ when $y := 1 + \gamma\left(x-\mu\right)/\sigma \to 0^+$ and $\gamma < 0$. Because $\gamma > \underline{\gamma} \geq -1$, the continuity assumption **A1*** holds true. Moreover, when $\gamma > -1$ the function $f_\theta(\cdot)$ is bounded. The condition $E\log^+ f_\theta(X_1) < \infty$ of **A3*** is thus satisfied.

Now note that as $y \to +\infty$, we have $|\log f_\theta(x)|\, f_\theta(x) = O(y^{-1/\gamma-1}\log y)$ when $\gamma > 0$ and $|\log f_\theta(x)|\, f_\theta(x) = O\left(\exp(y^{-1/\gamma})\right)$ when $\gamma < 0$. Note also that as $y \to 0^+$, $|\log f_\theta(x)|\, f_\theta(x)$ tends to zero at an exponential rate when $\gamma > 0$ and tends to zero like a positive power of y when $-1 < \gamma < 0$. This shows that $E|\log f_{\theta_0}(X_1)| < \infty$ when $\gamma_0 \neq 0$. When $\gamma_0 = 0$, the function $x \to |\log f_{\theta_0}(x)|\, f_{\theta_0}(x)$ is bounded away from zero on any compact set and tends to zero at an exponential rate when $x \to \pm\infty$, which shows that $E|\log f_{\theta_0}(X_1)| < \infty$ also when $\gamma_0 = 0$. We thus have shown that Assumption **A3*** is satisfied, and the consistency follows from Theorem 2.2.

Now observe that $\hat{\theta}_n$ and θ_0 necessarily belong to

$$\Theta_n := \left\{\theta : 1 + \gamma\left(X_{(1)} - \mu\right)/\sigma > 0 \text{ and } 1 + \gamma\left(X_{(n)} - \mu\right)/\sigma > 0\right\}$$

where $X_{(1)}$ and $X_{(n)}$ denote the minimum and maximum of the observations. Indeed, $\ell_n(\theta) = +\infty$ when $\theta \notin \Theta_n$. Moreover, $n^{-1}\ell_n(\theta) \to -E\log f_\theta(X_1)$ which is finite at θ_0, and thus also finite in a neighborhood of θ_0 by **A1***. This entails that $\ell_n(\theta)$ is finite, and admits derivatives of any order, on this neighborhood for n large enough. The Taylor expansion (A.4) thus holds. The existence and invertibility of J does not depend on the dynamics, and has already been proven by Smith (1985) in the iid case under the condition $\gamma_0 > -1/2$. Explicit expressions for the derivatives of $\log f_\theta(x)$ can be found in Beirlant et al. (2005). From these expressions, it can be seen that, for $\gamma < 0$, $\|\partial f_\theta(x)/\partial\theta\|^2 f_\theta(x)$ tends to zero at the exponential rate when $y \to -\infty$ and is equivalent to a constant multiplied by $y^{-3-1/\gamma}$ when $y \to 0^+$. It follows that, when $\gamma_0 > -1/2$, we have $E\|\partial f_{\theta_0}(X_1)/\partial\theta\|^{2+\varepsilon} < \infty$ for some $\varepsilon > 0$. The existence of I then follows from the mixing condition, using Davydov's inequality (1968). The conclusion follows.

SUPPLEMENTARY MATERIALS

Online appendix: Appendix A derives the matrix J^{-1} for the GPD. Appendix B provides complementary numerical illustrations. In particular, we consider alpha stable distributions fitted on aggregated series. Finally, Appendix C provides a proof for Theorem 2.3.

ACKNOWLEDGMENTS

We are grateful to the Agence Nationale de la Recherche (ANR), which supported this work via the Project ECONOM&RISK (ANR 2010 blanc 1804 03). We are also grateful to an anonymous Associate Editor and anonymous reviewers for helpful comments.

References

Anděl, J., Netuka, I. and K. Zvára (1984) On threshold autoregressive processes. *Kybernetika* 20, 89–106.

Andrews, D.W.K., (1991) Heteroskedasticity and autocorrelation consistent covariance matrix estimation. *Econometrica* 59, 817–858.

Andrews, B., Calder, M. and R.A. Davis (2009) Maximum likelihood estimation for α-stable autoregressive processes. *Annals of Statistics* 37, 1946–1982.

Balkema, A. and L. de Haan (1974) Residual life time at great age. *Annals of Probability* 2, 792–804.

Beirlant, J., Vynckier, P. and Teugels, J.L. (1996) Tail index estimation, Pareto quantile plots, and regression diagnostics. *Journal of the American Statistical Association* 91, 1659–1667.

Beirlant, J., Goegebeur, Y., Teugels, J., Segers, J. (2005) *Statistics of Extremes: Theory and Applications*. John Wiley & Sons, New York.

Berk, K.N. (1974) Consistent Autoregressive Spectral Estimates. *Annals of Statistics* 2, 489–502.

Berkes, I. and L. Horváth (2004) The efficiency of the estimators of the parameters in GARCH processes. *The Annals of Statistics* 32, 633–655.

Bontemps, C. and N. Meddahi (2005) Testing normality: a GMM approach. *Journal of Econometrics* 124, 149–186.

Billingsley, P. (1995) *Probability and Measure.* John Wiley & Sons, New York.

Bradley, R.C. (2005) Basic properties of strong mixing conditions. A survey and some open questions. *Probability Surveys* 2, 107–144.

Brockwell, P.J. and R.A. Davis (1991) *Time Series: Theory and Methods*, Springer-Verlag, New York, 2nd edition.

Chen, X. and Y. Fan (2006) Estimation of copula-based semiparametric time series models. *Journal of Econometrics* 130, 307–335.

Cotter, J. (2007) Varying the VaR for unconditional and conditional environments. *Journal of International Money and Finance* 26, 1338–1354.

Davis, R.A. and C.Y. Yau (2011) Comments on pairwise likelihood in time series models. *Statistica Sinica* 21, 255–277.

Davydov, Y.A. (1968) Convergence of distributions generated by stationary stochastic processes. *Theory of Probability and Applications* 13, 691–696.

de Zea Bermudez, P. and S. Kotz (2010) Parameter estimation of the generalized Pareto distribution. *Journal of Statistical Planning and Inference,* 140, 1353–1388.

den Hann, W.J. and A. Levin (1997) A practitioner's guide to robust covariance matrix estimation. *In Handbook of Statistics* 15, Rao, C.R. and G.S. Maddala (eds), 291–341.

DuMouchel, W.H. (1973) On the asymptotic normality of the maximum-likelihood estimate when sampling from a stable distribution. *The Annals of Statistics* 1, 948-957.

Einmahl, J.H.J., Li, J. and R.Y. Liu (2009) Thresholding events of extreme in simultaneous monitoring of multiple risks. *Journal of the American Statistical Association* 104, 982–992.

Embrechts, P., Klüppelberg, C. and T. Mikosch (1997) *Modelling Extremal Events*, Springer, New York.

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

Feller, W. (1975) *An Introduction to Probability Theory and Its Applications, Vol. II, Second Edition.* New York: John Wiley and Sons.

Francq, C., Roy, R., and J.-M. Zakoïan (2005) Diagnostic checking in ARMA models with uncorrelated errors. *Journal of the American Statistical Association* 100, 532–544.

Francq, C. and J.-M. Zakoïan (2007) HAC estimation and strong linearity testing in weak ARMA models. *Journal of Multivariate Analysis* 98, 114–144.

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

Herrndorf, N. (1984) A functional central limit theorem for weakly dependent sequences of random variables. *Annals of Probability* 12, 141–153.

Huisman, R., Koedijk, K.G., Kool, C.J.M. and F. Palm (2001) Tail-index estimates in small samples. *Journal of Business & Economic Statistics* 19, 208–216.

Jansen, D.W. and C.G. de Vries (1991) On the Frequency of Large Stock Returns: Putting Booms and Busts into Perspective. *The Review of Economics and Statistics* 73, 18–24.

Kon, S.J. (1984) Models of stochastic returns - A comparison. *Journal of Finance* 39, 147–165.

Lindsay, B.G. (1988) Composite likelihood methods. In *Statistical Inference from Stochastic Processes,* Ed. N. U. Prabhu, pp. 221–39. Providence: American Mathematical Society.

Liu, W. and W.B. Wu (2010) Simultaneous nonparametric inference of time series. *The Annals of Statistics* 38, 2388–2421.

Leitch, R.A. and A.S. Paulson (1975). Estimation of stable law parameters: Stock price behaviour application. *Journal of the American Statistical Association* 70, 690–697.

Loges, W. (2004) The stationary marginal distribution of a threshold AR(1) process. *Journal of Time Series Analysis* 25, 103–125.

Loretan, M. and P.C.B. Phillips (1994) Testing the covariance stationarity of heavy-tailed time series: An overview of the theory with applications to several financial data sets. *Journal of Empirical Finance* 1, 211–248.

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

McAleer, M. and S. Ling (2010) A general asymptotic theory for time-series models. *Statistica Neerlandica* 64, 97–111

McCulloch, J.H. (1996) Financial applications of stable distributions, in G.S. Maddala, C.R. Rao (eds.), *Handbook of Statistics,* 14, Elsevier, 393–425.

Newey, W.K. and K.D. West (1987) A simple, positive semi-definite, heteroskedasticity and autocorrelation consistent covariance matrix. *Econometrica* 55, 703–708.

Nolan, J. (2003) Modeling Financial Data with Stable Distributions. In Rachev S.T. Editor, *Handbook of Heavily Tailed Distributions in Finance,* Elsevier/North Holland.

Pickands, J. (1975) Statistical Inference Using Extreme Order Statistics. *The Annals of Statistics* 3, 119–131.

Pötscher, B.M. and I.R. Prucha (1997) *Dynamic Nonlinear Econometric Models*, Springer, Berlin.

Premaratne, G. and A. Bera (2005) A test for symmetry with leptokurtic financial data. *Journal of Financial Econometrics* 3, 169–187.

Rachev, S.T., ed. (2003) *Handbook of Heavy-tailed Distributions in Finance,* Elsevier/North Holland.

Rachev, S.T. and S. Mittnik (2000) *Stable Paretian Models in Finance.* New-York: Wiley.

Smith, R.L. (1984) Threshold methods for sample extremes. In *Statistical Extremes and Applications, ed. J. Tiago de Oliveira*, 621–38, Dordrecht: Reidel.

Smith, R.L. (1985) Maximum likelihood estimation in a class of nonregular cases. *Biometrika* 72, 67–92.

Taylor, S.J. (2007) *Asset price dynamics, volatility and prediction*, Princeton: Princeton University Press.

Tjøstheim, D. (1986) Estimation in nonlinear time series models. *Stochastic Processes and Applications* 21, 251–273.

Online appendix
Estimating the marginal distribution of heavy tailed time series

A Matrix J^{-1} for the GPD

The first-order derivatives of $\log f_\theta$ with respect to (γ, σ) are

$$\begin{aligned}
\frac{\partial \log f_\theta(z)}{\partial \gamma} &= \frac{1}{\gamma^2}\log\left(1+\gamma\frac{z}{\sigma}\right) - (1+\gamma)\frac{z}{\gamma(\gamma z+\sigma)},\\
\frac{\partial \log f_\theta(z)}{\partial \sigma} &= \frac{z-\sigma}{\sigma(\gamma z+\sigma)},
\end{aligned}$$

and the second-order derivatives are

$$\begin{aligned}
\frac{\partial^2 \log f_\theta(z)}{\partial \gamma^2} &= \frac{-2}{\gamma^3}\log\left(1+\gamma\frac{z}{\sigma}\right) + \frac{2}{\gamma^2}\frac{z}{\gamma z+\sigma} + \left(1+\frac{1}{\gamma}\right)\frac{z^2}{(\gamma z+\sigma)^2},\\
\frac{\partial^2 \log f_\theta(z)}{\partial \gamma \partial \sigma} &= \frac{-(z-\sigma)z}{\sigma(\gamma z+\sigma)^2},\\
\frac{\partial^2 \log f_\theta(z)}{\partial \sigma^2} &= \frac{(z-\sigma)^2 - z^2(1+\gamma)}{\sigma^2(\gamma z+\sigma)^2}.
\end{aligned}$$

Now let

$$m_{k,j} = E\left\{\frac{Z^k}{(\gamma Z+\sigma)^j}\right\}, \qquad 0 \le k \le j + \frac{1}{\gamma}.$$

We have, by integration by part,

$$m_{k,j} = \frac{k}{1+\gamma j} m_{k-1,j-1}, \qquad 1 \le k \le j + \frac{1}{\gamma}.$$

By direct integration we have $m_{0,j} = \frac{1}{\sigma^j(1+j\gamma)}$. It follows that

$$m_{1,1} = \frac{1}{1+\gamma}, \quad m_{1,2} = \frac{1}{\sigma(1+\gamma)(1+2\gamma)}, \quad m_{2,2} = \frac{2}{(1+\gamma)(1+2\gamma)}.$$

We also have

$$E\left\{\log\left(1+\gamma\frac{Z}{\sigma}\right)\right\} = \gamma.$$

Table 8: Stable distributions fitted by QMMLE on daily stock market returns. The estimated standard deviation are displayed into brackets.

Index	$\hat{\alpha}$	$\hat{\beta}$	$\hat{\sigma}$	$\hat{\mu}$
CAC	1.70 (0.08)	-0.16 (0.05)	0.80 (0.04)	0.07 (0.02)
DAX	1.62 (0.08)	-0.16 (0.05)	0.78 (0.05)	0.09 (0.02)
FTSE	1.64 (0.08)	-0.10 (0.04)	0.62 (0.03)	0.05 (0.01)
Nikkei	1.74 (0.06)	-0.09 (0.06)	0.90 (0.03)	0.01 (0.02)
NSE	1.55 (0.08)	-0.24 (0.07)	0.90 (0.05)	0.22 (0.04)
SMI	1.66 (0.07)	-0.22 (0.05)	0.65 (0.03)	0.09 (0.02)
SP500	1.55 (0.10)	-0.11 (0.05)	0.58 (0.04)	0.06 (0.01)
SPTSX	1.52 (0.12)	-0.23 (0.06)	0.61 (0.04)	0.11 (0.02)
SSE	1.49 (0.06)	-0.16 (0.06)	0.81 (0.03)	0.08 (0.04)

It follows that

$$\begin{aligned} E\left\{-\frac{\partial^2 \log f_\theta(z)}{\partial\gamma^2}\right\} &= \frac{2}{(1+\gamma)(1+2\gamma)}, \\ E\left\{-\frac{\partial^2 \log f_\theta(z)}{\partial\gamma\partial\sigma}\right\} &= \frac{1}{\sigma(1+\gamma)(1+2\gamma)}, \\ E\left\{-\frac{\partial^2 \log f_\theta(z)}{\partial\sigma^2}\right\} &= \frac{1}{\sigma^2(1+2\gamma)}. \end{aligned}$$

The matrix J^{-1}, as given in Theorem 3.2, follows.

B Complementary numerical illustrations

We now replicate the numerical illustrations of Section 4 on a sub-period which does not include the recent crisis. More precisely, we consider the nine stock returns during the period from January, 2 1991 to July, 3 2009 (except, of course, for the series whose first observations are posterior to 1991). The results of Tables 8-12 are similar to those displayed in the paper, in Tables 2-6.

Figure 5 shows that the estimated stable distributions actually resemble the non parametric kernel density estimator of the marginal distributions.

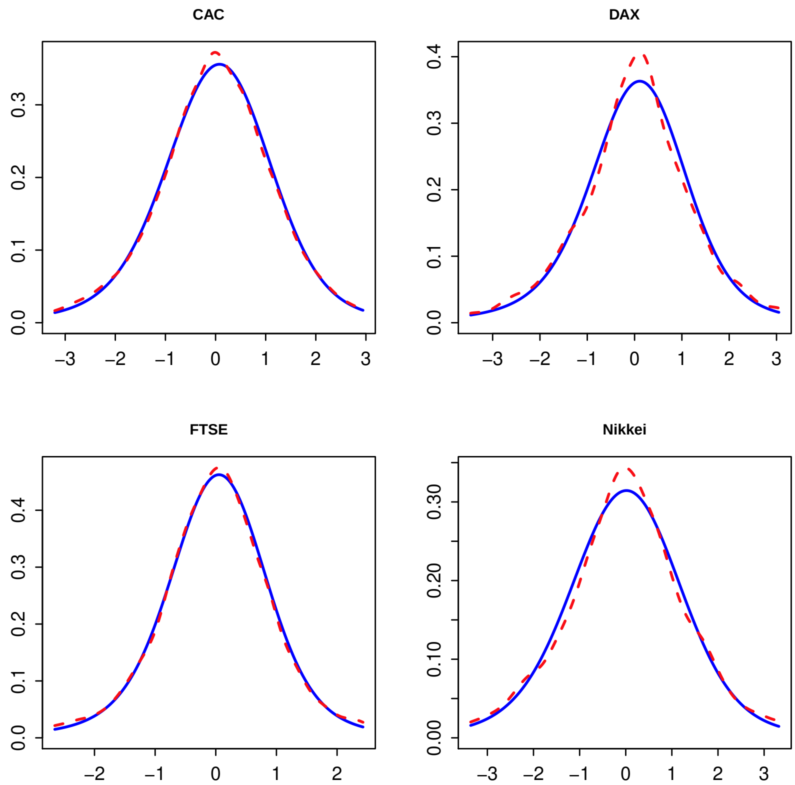

Figure 5: Comparison between the estimated stable density (full line) and the kernel density estimate (dashed line) of the marginal distribution of the returns of 4 stock market indices.

Table 9: p-values for the t-test of $H_0 : \beta = 0$ against $\beta \neq 0$.

CAC	DAX	FTSE	Nikkei	NSE	SMI	SP500	SPTSX	SSE
0.002	0.001	0.023	0.095	0.001	0.000	0.021	0.000	0.321

Table 10: p-values for the t-test of $H_0 : \mu = 0$ against $\mu > 0$.

CAC	DAX	FTSE	Nikkei	NSE	SMI	SP500	SPTSX	SSE
0.001	0.000	0.001	0.342	0.000	0.000	0.000	0.000	0.033

We now return to the most recent data sets. Table 13 displays the alpha stable distributions fitted on the aggregated series $X_t = \sum_{i=1}^{m} r_{5t+i}$ of each series of returns (r_t), for $m = 5$. Note that if the series r_t was iid, with a distribution which is not necessary stable but belongs to the domain of attraction of a stable distribution with tail index α, then, in view of the generalized CLT (see *e.g.* Feller, 1975), the distribution of X_t should be close to a stable distribution with tail index α for large m. To illustrate this point, let $\tilde{S}_t = S_t + N_t$, where (S_t) and (N_t) are two independent iid sequences, $S_t \sim S(\alpha, \beta, \sigma, \mu)$ and $N_t \sim \mathcal{N}(m, s)$. Figure 6 shows that, according to the asymptotic theory, the distribution of $\sum_{i=1}^{m} \tilde{S}_{5t+i}$ tends to the stable distribution of $\sum_{i=1}^{m} S_{5t+i}$ when m increases. For this figure, we took $\alpha = 0.8$, $\beta = \mu = m = 0$ and $\sigma = s = 1$. This simple illustration highlights that there exist obviously situations where a stable distribution is more plausible after temporal aggregation, and that the tail index is not changed by this transformation. Interestingly, Table 13 shows that the tail index estimated on the aggregated series is similar to that of the initial series of returns. Surprisingly

Table 11: Generalized Pareto distributions fitted by QMMLE on 12.5% of the most extreme daily stock market returns. The estimated standard deviation are displayed into brackets. The estimate of the tail index is NA (not available) when the estimate of GPD parameter γ is not positive.

Index	$\hat{\tau}$	$\hat{\alpha}_1 = 1/\hat{\gamma}_1$	$\hat{\sigma}_1$	$\hat{\alpha}_2 = 1/\hat{\gamma}_2$	$\hat{\sigma}_2$
CAC	0.53 (0.02)	10.69 (13.26)	0.99 (0.14)	4.51 (1.87)	0.81 (0.11)
DAX	0.50 (0.02)	89.15 (672.28)	1.22 (0.13)	3.74 (1.32)	0.77 (0.08)
FTSE	0.51 (0.02)	9.18 (9.30)	0.87 (0.12)	6.70 (5.90)	0.78 (0.14)
Nikkei	0.53 (0.01)	4.95 (3.64)	0.86 (0.11)	7.48 (4.82)	0.94 (0.11)
NSE	0.54 (0.03)	11.36 (12.28)	1.39 (0.20)	5.60 (3.37)	1.19 (0.18)
SMI	0.51 (0.02)	24.17 (58.24)	1.01 (0.13)	3.79 (1.27)	0.68 (0.08)
SP500	0.52 (0.02)	4.57 (2.28)	0.78 (0.14)	5.34 (2.96)	0.81 (0.14)
SPTSX	0.57 (0.03)	5.79 (4.36)	1.03 (0.33)	12.25 (17.05)	1.04 (0.22)
SSE	0.49 (0.03)	NA (NA)	1.38 (0.17)	3.72 (1.89)	0.88 (0.12)

Table 12: p-value for the Wald test of $H_0 : \tau = 0.5$ and $\sigma_1 = \sigma_2$.

CAC	DAX	FTSE	Nikkei	NSE	SMI	SP500	SPTSX	SSE
0.065	0.004	0.227	0.005	0.334	0.024	0.343	0.016	0.049

the estimated standard deviation of the estimator of α is not deteriorated by the aggregation (although the number of observations is obviously divided by $m = 5$). A possible explanation is that the temporal dependencies should decrease as $m \to \infty$, which could facilitate the estimation of that parameter. Another surprising output of Table 13 is that the asymmetry parameter β is much more negative for $m = 5$ than for $m = 1$. This is certainly due to the presence of clusters of negative returns. Table 14 display the estimated tail index α for different values of m. The main output of that table is that $\hat{\alpha}$ is always greater than 1.5 and less than 2, for all indices and any m, leading to the conclusion that the moments of order 1 should exist, whereas those of order 2 should not.

Table 13: Stable distributions fitted by QMMLE on rolling sums of $m = 5$ consecutive daily stock market returns.

Index	$\hat{\alpha}$	$\hat{\beta}$	$\hat{\sigma}$	$\hat{\mu}$
CAC	1.81 (0.06)	-0.48 (0.10)	1.95 (0.10)	0.28 (0.08)
DAX	1.74 (0.07)	-0.47 (0.11)	1.86 (0.17)	0.44 (0.12)
FTSE	1.74 (0.08)	-0.29 (0.11)	1.40 (0.06)	0.28 (0.07)
Nikkei	1.75 (0.05)	-0.44 (0.11)	1.82 (0.08)	0.24 (0.09)
NSE	1.61 (0.11)	-0.50 (0.20)	2.25 (0.17)	0.90 (0.20)
SMI	1.65 (0.08)	-0.45 (0.09)	1.47 (0.10)	0.44 (0.08)
SP500	1.77 (0.05)	-0.32 (0.09)	1.29 (0.05)	0.27 (0.04)
SPTSX	1.55 (0.15)	-0.47 (0.13)	1.33 (0.13)	0.42 (0.08)
SSE	1.78 (0.09)	-0.26 (0.32)	2.34 (0.16)	0.32 (0.32)

Table 14: Estimated tail index α when stable distributions are fitted by QMMLE on rolling sums of m consecutive daily stock market returns.

Index	$m = 1$	$m = 2$	$m = 4$	$m = 8$	$m = 16$	$m = 32$
CAC	1.72 (0.07)	1.73 (0.08)	1.83 (0.06)	1.86 (0.05)	1.82 (0.08)	1.73 (0.14)
DAX	1.64 (0.07)	1.66 (0.06)	1.75 (0.07)	1.71 (0.09)	1.65 (0.13)	1.63 (0.23)
FTSE	1.70 (0.06)	1.73 (0.06)	1.79 (0.06)	1.79 (0.07)	1.70 (0.11)	1.80 (0.19)
Nikkei	1.65 (0.05)	1.70 (0.06)	1.80 (0.06)	1.77 (0.06)	1.80 (0.13)	1.85 (0.18)
NSE	1.60 (0.09)	1.64 (0.08)	1.63 (0.11)	1.76 (0.10)	1.66 (0.14)	1.68 (0.14)
SMI	1.66 (0.06)	1.67 (0.07)	1.68 (0.07)	1.74 (0.07)	1.61 (0.11)	1.76 (0.12)
SP500	1.62 (0.05)	1.73 (0.05)	1.77 (0.04)	1.80 (0.05)	1.82 (0.05)	1.82 (0.11)
SPTSX	1.55 (0.11)	1.64 (0.12)	1.52 (0.09)	1.64 (0.09)	1.62 (0.10)	1.68 (0.20)
SSE	1.54 (0.06)	1.71 (0.05)	1.73 (0.07)	1.81 (0.07)	1.97 (0.03)	1.91 (0.06)

In order to further assess the previous assumptions on the marginal moments, we draw the empirical moments $M_{r,n} = n^{-1} \sum_{t=1}^{n} |r_t|^r$ as function of n, for $r = 1$ (Figure 7) and $r = 2$ (Figure 8). The ergodic theorem entails that, if the tail indices are correctly estimated, $M_{1,n}$ should converge and

$M_{2,n}$ should diverge. The main output of these figures is that the empirical moments $M_{r,n}$ of the returns do not resemble those of iid sequences with the stable distribution fitted on the returns by QMMLE. An obvious explanation for that is that the returns r_t are not independent. This is not the sole reason because if the marginal distribution where the estimated stable distribution, by the ergodic theorem $M_{r,n}$ should however converge to the corresponding moment, which does not seem to be the case. Indeed, the empirical moments $M_{r,n}$ computed on the real series r_t are always smaller that those computed on the simulations of stable distribution. We draw the conclusion that the marginal distribution of the returns are not be well approximated by a stable distribution. It is a much more difficult to infer if the sequence $M_{r,n}$ converge or not, and thus to assess if the estimated tail indices are plausible, by simple inspection of the graphs. By the previous arguments based on generalized CLT, the marginal distribution of rolling sums of m consecutive returns are expected to be closer to a stable distribution, at least for m large enough. Figures 9 and 10 confirm that the empirical moments are indeed closest to those of the estimated stable distributions, but these average are still smaller that expected. We thus have a serious doubt on the adequacy the class of the stable distributions for modeling the marginal distribution of the returns or even of aggregates of r returns, at least for moderate values of r.

Figures 11 and 12 indicate that the behavior of the empirical moments $M_{r,n}$ are in accordance with the assumption of a marginal GEV for the block maxima, but the size m of the blocks must be large.

To have an idea on how large should be the size m of the blocks, we made a last experiment. We fitted GEV to block maxima of $1,000$ independent realizations of length $n = 4,000$ of the moving average Model (4.1) whose marginal is the stable distribution of parameter $\alpha = 1.6$, $\beta = 0$, $\sigma = 1$ and $\mu = 0$. Table 15 gives the estimated value of the tail index α. The main output is that the size m needs to be dramatically large. Even for $m = 48$, the estimation of α is still largely positively biased. The numbers between the brackets are the observed standard deviations of the estimates over the $1,000$ replications. Surprisingly, these standard deviations do not systematically increase with m (although the number of observation $[1000/m]$ decreases). This is in accordance with the estimated standard deviations that we obtained in Table 7. This can be explained by the fact that the time dependence decreases when m increases. The effect of the time dependence is indeed clear, because the estimation results worsen when the order of the dependence parameter k increases.

Table 15: Estimated tail index α of the stable MA(k) (4.1) by fitting GEV distributions on block maxima of size m.

MA order k	$m = 8$	$m = 16$	$m = 24$
1	3.35 (0.55)	2.33 (0.38)	2.00 (0.35)
2	5.03 (1.42)	3.16 (0.77)	2.57 (0.60)
3	2011.57 (44698.62)	4.09 (1.59)	3.15 (1.02)
4	15012.88 (121611.7)	5.12 (2.95)	3.81 (1.75)

MA order k	$m = 32$	$m = 40$	$m = 48$
1	1.83 (0.36)	1.74 (0.37)	1.71 (0.40)
2	2.28 (0.56)	2.08 (0.52)	2.00 (0.59)
3	2.75 (0.92)	2.52 (1.25)	2.32 (0.85)
4	3.21 (1.25)	2.86 (1.02)	2.64 (0.98)

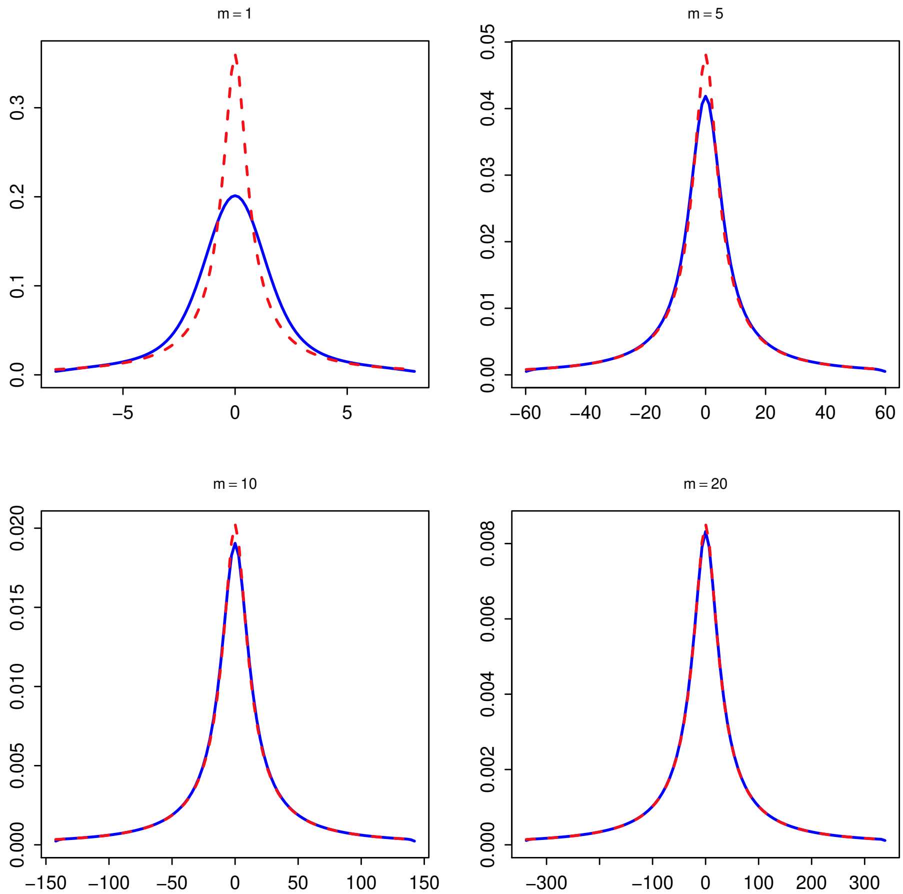

Figure 6: Convergence of the distribution of $\sum_{i=1}^{m} \tilde{S}_{5t+i}$ (full blue line) to that of a stable distribution (dashed red line) as $m \to \infty$, for an iid sequence $\tilde{S}_t$ which does not follow a stable distribution (see the text for details).

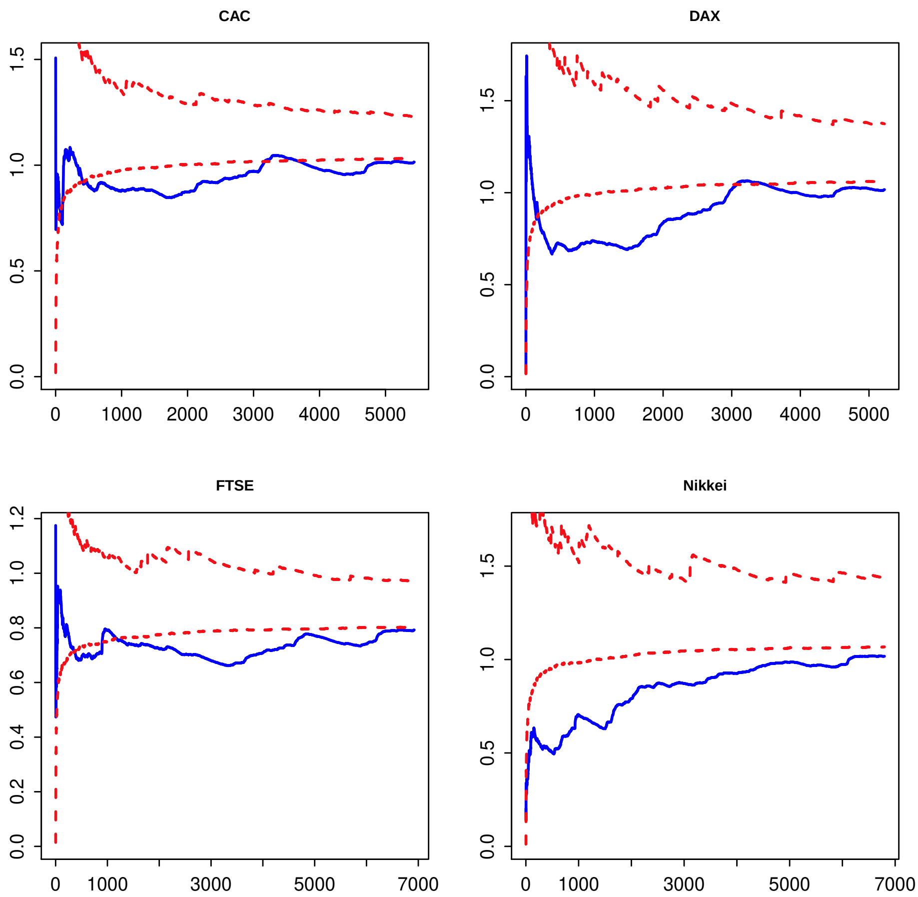

Figure 7: Empirical moment $M_{1,n} = n^{-1} \sum_{t=1} |r_t|$ (full line) as function of n, for the returns r_t of 4 stock market indices. The dotted lines are the 1% and 99% empirical quantiles of 1000 trajectories of $n^{-1} \sum_{t=1} |X_t|$ where X_t is an iid sequence of the stable distribution fitted by QMMLE.

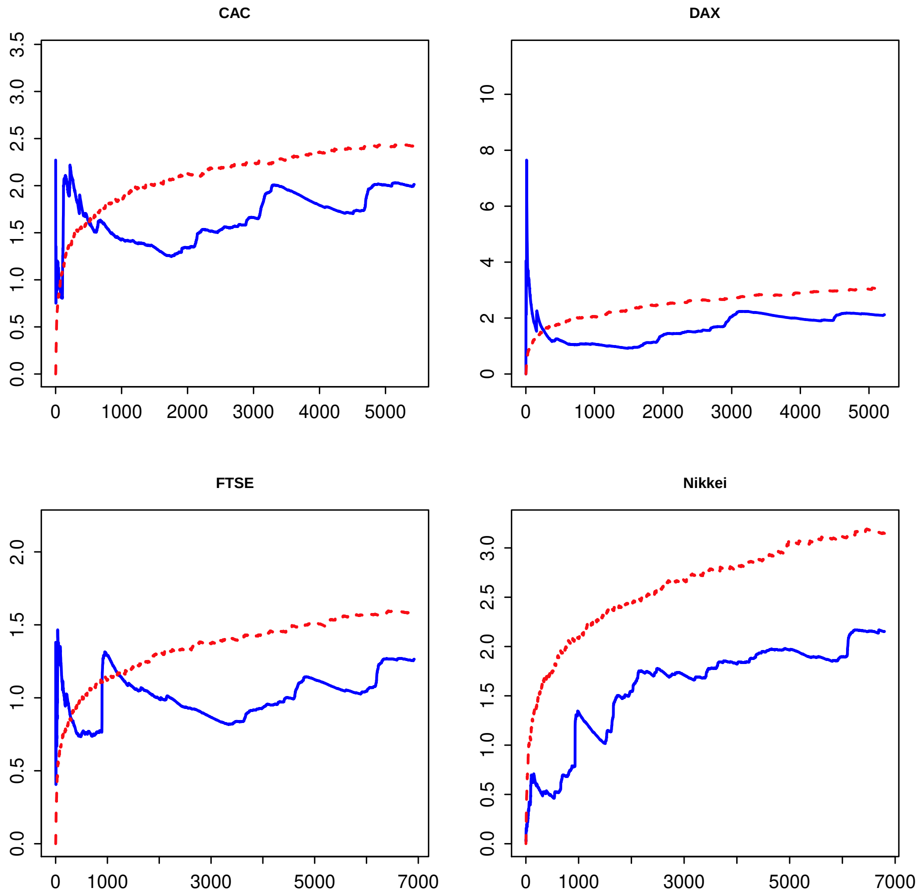

Figure 8: As Figure 8, but for the empirical moment $M_{2,n} = n^{-1} \sum_{t=1} r_t^2$ (the 99% upper bound is outside the frame).

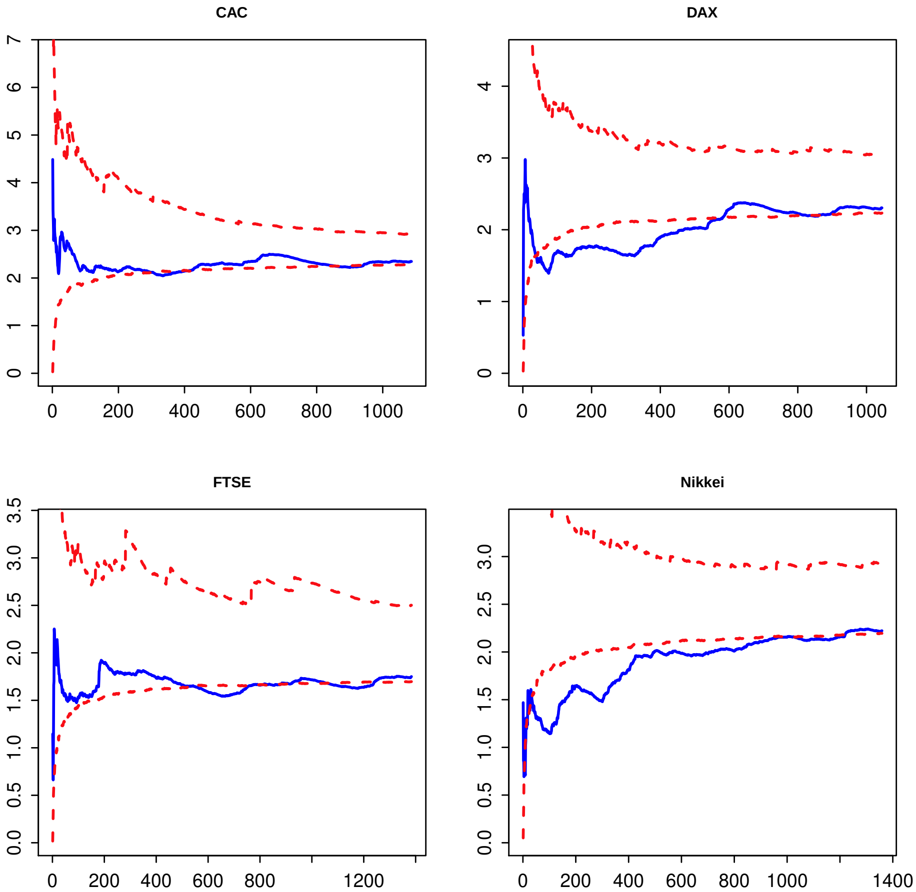

Figure 9: As Figure 7, but for rolling sums of 5 consecutive returns.

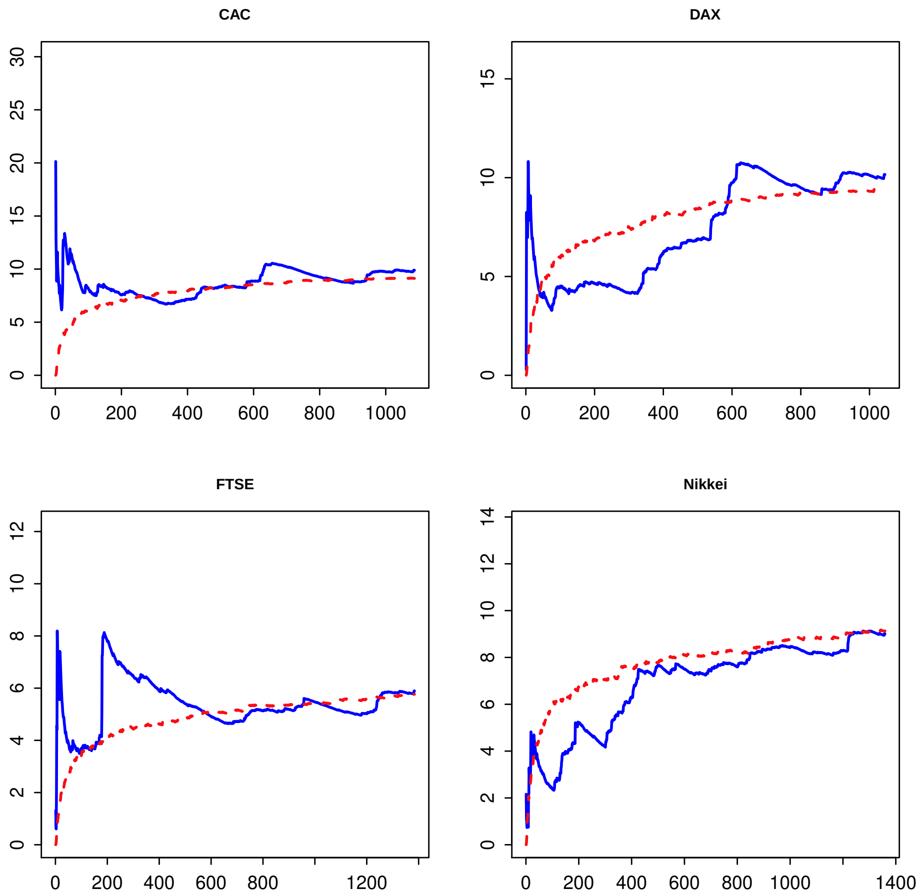

Figure 10: As Figure 8, but for rolling sums of 5 consecutive returns.

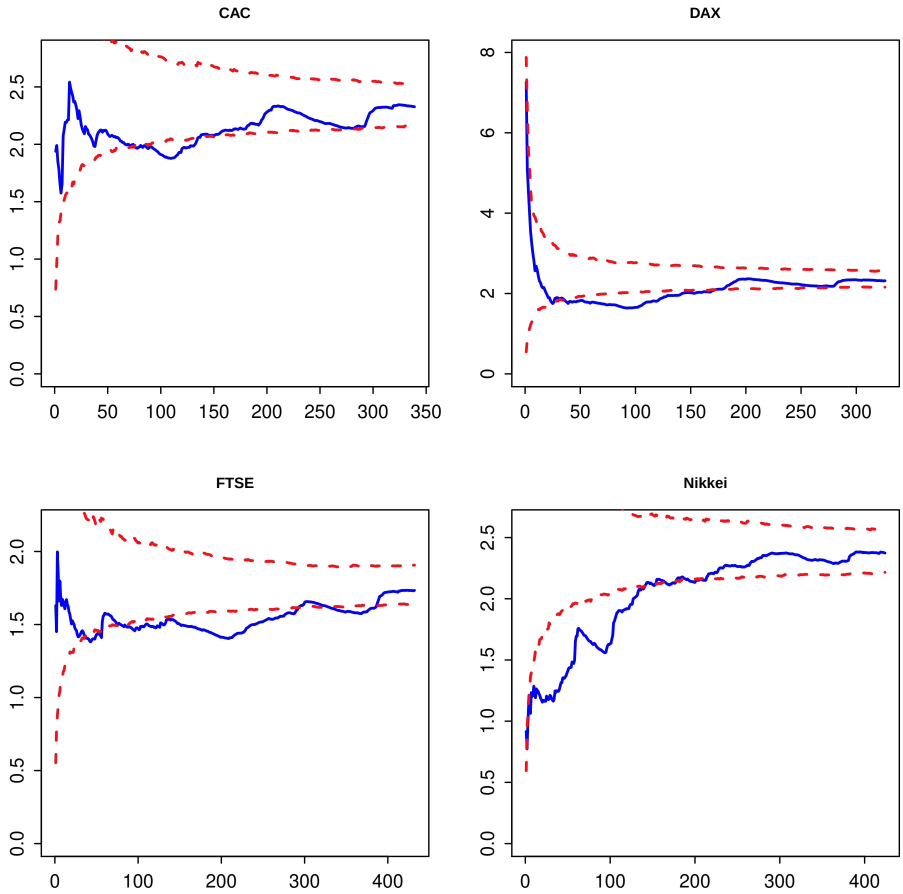

Figure 11: As Figure 7, but r_t is replaced by the maximum $max\{r_{mt+1},\ldots,r_{mt+m}\}$ of $m = 16$ consecutive returns, and the dotted lines are the 1% and 99% confidence bounds for $n^{-1}\sum_{t=1}|X_t|$ when X_t is iid with GEV distribution.

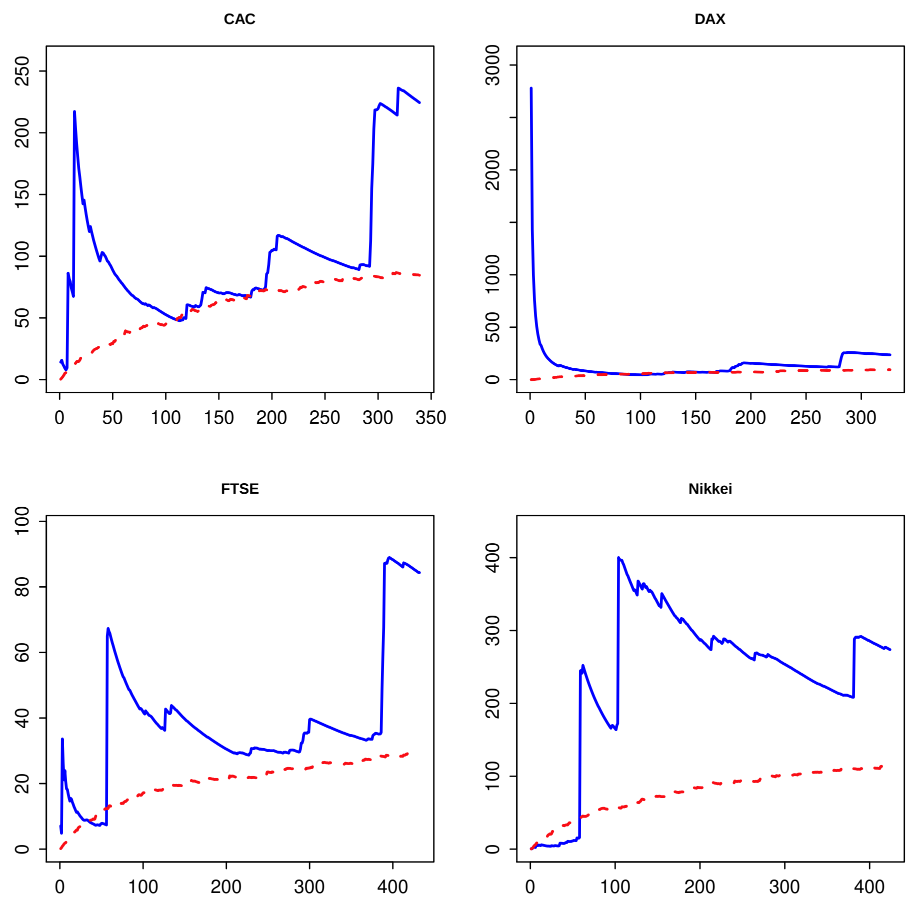

Figure 12: As Figure 11, but $M_{1,n}$ is replaced by $M_{4,n}$.

C Proof of Theorem 2.3

The proof is based on a series of lemmas. Similar proofs can be found in the supplementary files of Francq, Roy and Zakoïan (2005) and Boubacar, Carbon and Francq (2011). We begin by proving that $\hat{J}$ is a consistent estimator of J. It will be convenient to introduce the notation $\hat{\Sigma}_{\hat{S}} = \hat{J}$, $\hat{\Sigma}_S = n^{-1}\sum_{t=1}^{n} S_t S_t'$ and $\Sigma_S = J = ES_tS_t'$.

Lemma C.1. *Under the assumptions of Theorem 2.1, $\hat{\Sigma}_{\hat{S}} \to \Sigma_S$ a.s. when $n \to \infty$.*

Proof of Lemma C.1. A Taylor expansion yields

$$\hat{\Sigma}_{\hat{S}}(i,j) = \hat{\Sigma}_S(i,j) + (\hat{\theta}_n - \theta_0)' \frac{1}{n}\sum_{t=1}^{n} \frac{\partial}{\partial\theta}\left\{\frac{\partial \log f_\theta(X_t)}{\partial\theta_i}\frac{\partial \log f_\theta(X_t)}{\partial\theta_j}\right\}(\theta^*) \tag{C.1}$$

for some θ^* between $\hat{\theta}_n$ and θ_0. The consistency of $\hat{J}$ then follows from Assumption **A4'**, the consistency of $\hat{\theta}_n$ and the ergodic theorem. □

We use the multiplicative matrix norm $\|A\| = \sup_{\|x\|\leq 1}\|Ax\| = \varrho^{1/2}(A'A)$, where A is a $d_1 \times d_2$ matrix, $\|x\|$ is the Euclidean norm of the vector $x \in \mathbb{R}^{d_2}$, and $\varrho(\cdot)$ denotes the spectral radius. This choice of the norm is crucial for the following lemma to hold (with e.g. the Euclidean norm, this result is not valid). Let $\underline{S}_{r,t} = \left(S_{t-1}', \ldots, S_{t-r}'\right)'$ and

$$\Sigma_{S,\underline{S}_r} = ES_t\underline{S}_{r,t}', \quad \Sigma_{\underline{S}_r} = E\underline{S}_{r,t}\underline{S}_{r,t}'.$$

In the sequel, K and ρ denote generic constant such as $K > 0$ and $\rho \in (0,1)$, whose exact values are unimportant.

Lemma C.2. *Under the assumptions of Theorem 2.3,*

$$\sup_{r\geq 1} \max \left\{ \left\| \Sigma_{S,\underline{S}_r} \right\|, \left\| \Sigma_{\underline{S}_r} \right\|, \left\| \Sigma_{\underline{S}_r}^{-1} \right\| \right\} \leq \infty.$$

Proof. We readily have

$$\|\Sigma_{\underline{S}_r} x\| \leq \|\Sigma_{\underline{S}_{r+1}}(x', 0_q')'\| \quad \text{and} \quad \|\Sigma_{S,\underline{S}_r} x\| \leq \|\Sigma_{\underline{S}_{r+1}}(0_q', x')'\|$$

for any $x \in \mathbb{R}^{qr}$. Therefore

$$0 < \|\text{Var}\,(S_t)\| = \left\| \Sigma_{\underline{S}_1} \right\| \leq \left\| \Sigma_{\underline{S}_2} \right\| \leq \cdots$$

and

$$\left\| \Sigma_{S,\underline{S}_r} \right\| \leq \left\| \Sigma_{\underline{S}_{r+1}} \right\|.$$

Let $f(\lambda)$ be the spectral density of S_t. Because the autocovariance function of S_t is absolutely summable, $\|f(\lambda)\|$ is bounded by a finite constant K, say. Denoting by $\delta = (\delta_1', \dots, \delta_r')'$ an eigenvector of $\Sigma_{\underline{S}_r}$ associated with its largest eigenvalue, such that $\|\delta\| = 1$ and $\delta_i \in \mathbb{R}^q$ for $i = 1, \dots, r$, we have

$$\begin{aligned} \left\| \Sigma_{\underline{S}_r} \right\| &= \varrho^{1/2}(\Sigma_{\underline{S}_r}^2) = \varrho(\Sigma_{\underline{S}_r}) = \delta' \Sigma_{\underline{S}_r} \delta \\ &= \sum_{j,k=1}^{r} \delta_j' \int_{-\pi}^{\pi} e^{i(k-j)\lambda} f(\lambda) d(\lambda) \delta_k \leq 2\pi K. \end{aligned}$$

By similar arguments, the smallest eigenvalue of $\Sigma_{\underline{S}_r}$ is greater than a positive constant independent of r. Using the fact that $\|\Sigma_{\underline{S}_r}^{-1}\|$ is equal to the inverse of the smallest eigenvalue of $\Sigma_{\underline{S}_r}$, the proof is completed. □

Denote by $S_t(i)$ the i-th element of S_t.

Lemma C.3. *Under* **A5'**, *there exits a finite constant K_1 such that for $m_1, m_2 = 1, \dots, q$*

$$\sup_{s\in\mathbb{Z}} \sum_{h=-\infty}^{\infty} |Cov\{S_1(m_1)S_{1+s}(m_2), S_{1+h}(m_1)S_{1+s+h}(m_2)\}| < K_1.$$

Proof. See for instance Corollary A.3 in Francq and Zakoïan (2010). □

Let $\hat{\Sigma}_{\underline{S}_r}$, $\hat{\Sigma}_S$ and $\hat{\Sigma}_{S,\underline{S}_r}$ be the matrices obtained by replacing $\hat{S}_t$ by S_t in $\hat{\Sigma}_{\underline{\hat{S}}_r}$, $\hat{\Sigma}_{\hat{S}}$ and $\hat{\Sigma}_{\hat{S},\underline{\hat{S}}_r}$.

Lemma C.4. *Under the assumptions of Theorem 2.3,* $\sqrt{r}\|\hat{\Sigma}_{\underline{S}_r} - \Sigma_{\underline{S}_r}\|$, $\sqrt{r}\|\hat{\Sigma}_S - \Sigma_S\|$, *and* $\sqrt{r}\|\hat{\Sigma}_{S,\underline{S}_r} - \Sigma_{S,\underline{S}_r}\|$ *tend to zero in probability as* $n \to \infty$ *when* $r = o(n^{1/3})$.

Proof. For $1 \leq m_1, m_2 \leq q$ and $1 \leq r_1, r_2 \leq r$, the element of the $\{(r_1-1)q + m_1\}$-th row and $\{(r_2-1)q + m_2\}$-th column of $\hat{\Sigma}_{\underline{S}_r}$ is of the form $n^{-1}\sum_{t=1}^n Z_t$ where $Z_t = S_{t-r_1}(m_1)S_{t-r_2}(m_2)$. By stationarity of (Z_t), we have

$$\mathrm{Var}\left(\frac{1}{n}\sum_{t=1}^{n} Z_t\right) = \frac{1}{n^2}\sum_{h=-n+1}^{n-1}(n-|h|)\,\mathrm{Cov}\left(Z_t, Z_{t-h}\right) \leq \frac{K_1}{n}, \tag{C.2}$$

where, by Lemma C.3, K_1 is a constant independent of r_1, r_2, m_1, m_2 and r, n. Note that the sup-norm satisfies

$$\|A\|^2 \leq \sum_{i,j} a_{i,j}^2 \tag{C.3}$$

with obvious notations.

In view of (C.3) and (C.2), using arguments of the proof of Lemma C.2, we have

$$\begin{aligned} E\left\{r\|\hat{\Sigma}_S - \Sigma_S\|^2\right\} &\leq E\left\{r\|\hat{\Sigma}_{S,\underline{S}_r} - \Sigma_{S,\underline{S}_r}\|^2\right\} \\ &\leq E\left\{r\|\hat{\Sigma}_{\underline{S}_r} - \Sigma_{\underline{S}_r}\|^2\right\} \leq \frac{K_1 q^2 r^3}{n} = o(1) \end{aligned}$$

as $n \to \infty$ when $r = o(n^{1/3})$. The result follows. □

We now show that the previous lemma applies when S_t is replaced by $\hat{S}_t$.

Lemma C.5. *Under the assumptions of Theorem 2.3,* $\sqrt{r}\|\hat{\Sigma}_{\underline{\hat{S}}_r} - \Sigma_{\underline{S}_r}\|$, $\sqrt{r}\|\hat{\Sigma}_{\hat{S}} - \Sigma_S\|$, *and* $\sqrt{r}\|\hat{\Sigma}_{\hat{S},\underline{\hat{S}}_r} - \Sigma_{S,\underline{S}_r}\|$ *tend to zero in probability as* $n \to \infty$ *when* $r = o(n^{1/3})$.

Proof. Similarly to (C.1), for $1 \leq m_1, m_2 \leq q$ and $1 \leq r_1, r_2 \leq r$, the element of the $\{(r_1-1)q+m_1\}$-th row and $\{(r_2-1)q+m_2\}$-th column of $\hat{\Sigma}_{\underline{\hat{S}}_r} - \hat{\Sigma}_{\underline{S}_r}$ is of the form

$$(\hat{\theta}_n - \theta_0)' \frac{1}{n} \sum_{t=1}^{n} \frac{\partial}{\partial \theta} \left\{ \frac{\partial \log f_\theta(X_{t-r_1})}{\partial \theta_{m_1}} \frac{\partial \log f_\theta(X_{t-r_2})}{\partial \theta_{m_2}} \right\} (\theta^*)$$

for some θ^* between $\hat{\theta}_n$ and θ_0. By Assumption **A4'**, the expectation of the absolute value of the latter empirical mean is bounded by a constant K independent of n, r_1, r_2, m_1 and m_2. Thus, using again (C.3),

$$\|\hat{\Sigma}_{\underline{\hat{S}}_r} - \hat{\Sigma}_{\underline{S}_r}\|^2 \leq r^2 \left\| \hat{\theta}_n - \theta_0 \right\|^2 O_P(1).$$

Since $\left\| \hat{\theta}_n - \theta_0 \right\| = O_P\left(n^{-1/2}\right)$, we obtain for $r = o(n^{1/3})$

$$\sqrt{r}\|\hat{\Sigma}_{\underline{\hat{S}}_r} - \hat{\Sigma}_{\underline{S}_r}\| = o_P(1). \tag{C.4}$$

By Lemma C.4 , (C.4) shows that $\sqrt{r}\|\hat{\Sigma}_{\underline{\hat{S}}_r} - \Sigma_{\underline{S}_r}\| = o_P(1)$. The other results are obtained similarly. □

Write $\underline{A}_r^* = (A_1 \cdots A_r)$ where the A_i's are defined by (2.2).

Lemma C.6. *Under the assumptions of Theorem 2.3,*

$$\sqrt{r}\,\|\underline{A}_r^* - \underline{A}_r\| \to 0,$$

as $r \to \infty$.

Proof. Recall that by (2.2) and (2.4)

$$S_t = \underline{A}_r \underline{S}_{r,t} + u_{r,t} = \underline{A}_r^* \underline{S}_{r,t} + \sum_{i=r+1}^{\infty} A_i S_{t-i} + u_t := \underline{A}_r^* \underline{S}_{r,t} + u_{r,t}^*.$$

Hence, using the orthogonality conditions in (2.2) and (2.4)

$$\underline{A}_r^* - \underline{A}_r = -\Sigma_{u_r^*, \underline{S}_r} \Sigma_{\underline{S}_r}^{-1} \tag{C.5}$$

where $\Sigma_{u_r^*, \underline{S}_r} = E u_{r,t}^* \underline{S}_{r,t}'$. By Assumption **A4**, there exists a constant K_2 independent of s and m_1, m_2 such that

$$E\left|S_1(m_1) S_{1+s}(m_2)\right| \leq K_2.$$

By (C.3), we then have

$$\left\|\mathrm{Cov}\left(S_{t-r-h}, \underline{S}_{r,t}\right)\right\| \leq K_2 r^{1/2} q.$$

Thus,

$$\begin{aligned} \|\Sigma_{u_r^*, \underline{S}_r}\| &= \|\sum_{i=r+1}^{\infty} A_i E S_{t-i} \underline{S}_{r,t}'\| \leq \sum_{h=1}^{\infty} \|A_{r+h}\| \left\|\mathrm{Cov}\left(S_{t-r-h}, \underline{S}_{r,t}\right)\right\| \\ &= O(1) r^{1/2} \sum_{h=1}^{\infty} \|A_{r+h}\|. \end{aligned} \tag{C.6}$$

Note that the assumption $\|A_i\| = o\left(i^{-2}\right)$ entails $r \sum_{h=1}^{\infty} \|A_{r+h}\| = o(1)$ as $r \to \infty$. The lemma therefore follows from (C.5), (C.6) and Lemma C.2. □

The following lemma is similar to Lemma 3 in Berk (1974).

Lemma C.7. *Under the assumptions of Theorem 2.3,*

$$\sqrt{r} \|\hat{\Sigma}_{\hat{\underline{S}}_r}^{-1} - \Sigma_{\underline{S}_r}^{-1}\| = o_P(1)$$

as $n \to \infty$ when $r = o(n^{1/3})$ and $r \to \infty$.

Proof. We have

$$\begin{aligned}\left\|\hat{\Sigma}_{\hat{\underline{S}}_r}^{-1} - \Sigma_{\underline{S}_r}^{-1}\right\| &= \left\|\left\{\hat{\Sigma}_{\hat{\underline{S}}_r}^{-1} - \Sigma_{\underline{S}_r}^{-1} + \Sigma_{\underline{S}_r}^{-1}\right\}\left\{\Sigma_{\underline{S}_r} - \hat{\Sigma}_{\hat{\underline{S}}_r}\right\}\Sigma_{\underline{S}_r}^{-1}\right\| \\ &\leq \left(\left\|\hat{\Sigma}_{\hat{\underline{S}}_r}^{-1} - \Sigma_{\underline{S}_r}^{-1}\right\| + \left\|\Sigma_{\underline{S}_r}^{-1}\right\|\right)\left\|\hat{\Sigma}_{\hat{\underline{S}}_r} - \Sigma_{\underline{S}_r}\right\|\left\|\Sigma_{\underline{S}_r}^{-1}\right\|.\end{aligned}$$

Iterating this inequality, we obtain

$$\left\|\hat{\Sigma}_{\hat{\underline{S}}_r}^{-1} - \Sigma_{\underline{S}_r}^{-1}\right\| \leq \left\|\Sigma_{\underline{S}_r}^{-1}\right\| \sum_{i=1}^{\infty} \left\|\hat{\Sigma}_{\hat{\underline{S}}_r} - \Sigma_{\underline{S}_r}\right\|^i \left\|\Sigma_{\underline{S}_r}^{-1}\right\|^i.$$

Thus, for every $\varepsilon > 0$,

$$\begin{aligned}&P\left(\sqrt{r}\left\|\hat{\Sigma}_{\hat{\underline{S}}_r}^{-1} - \Sigma_{\underline{S}_r}^{-1}\right\| > \varepsilon\right) \\ \leq\ & P\left(\sqrt{r}\frac{\left\|\Sigma_{\underline{S}_r}^{-1}\right\|^2\left\|\hat{\Sigma}_{\hat{\underline{S}}_r} - \Sigma_{\underline{S}_r}\right\|}{1 - \left\|\hat{\Sigma}_{\hat{\underline{S}}_r} - \Sigma_{\underline{S}_r}\right\|\left\|\Sigma_{\underline{S}_r}^{-1}\right\|} > \varepsilon \text{ and } \left\|\hat{\Sigma}_{\hat{\underline{S}}_r} - \Sigma_{\underline{S}_r}\right\|\left\|\Sigma_{\underline{S}_r}^{-1}\right\| < 1\right) \\ &+P\left(\sqrt{r}\left\|\hat{\Sigma}_{\hat{\underline{S}}_r} - \Sigma_{\underline{S}_r}\right\|\left\|\Sigma_{\underline{S}_r}^{-1}\right\| \geq 1\right) \\ \leq\ & P\left(\sqrt{r}\left\|\hat{\Sigma}_{\hat{\underline{S}}_r} - \Sigma_{\underline{S}_r}\right\| > \frac{\varepsilon}{\left\|\Sigma_{\underline{S}_r}^{-1}\right\|^2 + \varepsilon r^{-1/2}\left\|\Sigma_{\underline{S}_r}^{-1}\right\|}\right) \\ &+P\left(\sqrt{r}\left\|\hat{\Sigma}_{\hat{\underline{S}}_r} - \Sigma_{\underline{S}_r}\right\| \geq \left\|\Sigma_{\underline{S}_r}^{-1}\right\|^{-1}\right) = o(1)\end{aligned}$$

by Lemmas C.4 and C.2. This establishes Lemma C.7. □

Lemma C.8. *Under the assumptions of Theorem 2.3,*

$$\sqrt{r}\left\|\underline{\hat{A}}_r - \underline{A}_r\right\| = o_P(1)$$

as $r \to \infty$ *and* $r = o(n^{1/3})$.

Proof. By the triangle inequality and Lemmas C.2 and C.7, we have

$$\left\|\hat{\Sigma}_{\hat{\underline{S}}_r}^{-1}\right\| \leq \left\|\hat{\Sigma}_{\hat{\underline{S}}_r}^{-1} - \Sigma_{\underline{S}_r}^{-1}\right\| + \left\|\Sigma_{\underline{S}_r}^{-1}\right\| = O_P(1). \qquad \text{(C.7)}$$

Note that the orthogonality conditions in (2.4) entail that $\underline{A}_r = \Sigma_{S,\underline{S}_r}\Sigma_{\underline{S}_r}^{-1}$. By Lemmas C.2, C.4, C.7, and (C.7), we then have

$$\begin{aligned}
&\sqrt{r}\left\|\underline{\hat{A}}_r - \underline{A}_r\right\| = \sqrt{r}\left\|\hat{\Sigma}_{\hat{S},\underline{\hat{S}}_r}\hat{\Sigma}_{\underline{\hat{S}}_r}^{-1} - \Sigma_{S,\underline{S}_r}\Sigma_{\underline{S}_r}^{-1}\right\| \\
= \;& \sqrt{r}\left\|\left(\hat{\Sigma}_{\hat{S},\underline{\hat{S}}_r} - \Sigma_{S,\underline{S}_r}\right)\hat{\Sigma}_{\underline{\hat{S}}_r}^{-1} + \Sigma_{S,\underline{S}_r}\left(\hat{\Sigma}_{\underline{\hat{S}}_r}^{-1} - \Sigma_{\underline{S}_r}^{-1}\right)\right\| = o_P(1).
\end{aligned}$$

□

Proof of Theorem 2.3. In view of (2.3), it suffices to show that $\hat{\mathcal{A}}_r(1) \to \mathcal{A}(1)$ and $\hat{\Sigma}_{u_r} \to \Sigma_u$ in probability. Let the $r \times 1$ vector $\mathbf{1}_r = (1, \ldots, 1)'$ and the $rq \times q$ matrix $\mathbf{E}_r = \mathbb{I}_q \otimes \mathbf{1}_r$, where $\otimes$ denotes the matrix Kronecker product and $\mathbb{I}_d$ the $d \times d$ identity matrix. Using (C.3), and Lemmas C.6 and C.8, we obtain

$$\begin{aligned}
\left\|\hat{\mathcal{A}}_r(1) - \mathcal{A}(1)\right\| &\leq \left\|\sum_{i=1}^{r}\hat{A}_{r,i} - A_{r,i}\right\| + \left\|\sum_{i=1}^{r}A_{r,i} - A_i\right\| + \left\|\sum_{i=r+1}^{\infty}A_i\right\| \\
&= \left\|\left(\underline{\hat{A}}_r - \underline{A}_r\right)\mathbf{E}_r\right\| + \left\|\left(\underline{A}_r^* - \underline{A}_r\right)\mathbf{E}_r\right\| + \left\|\sum_{i=r+1}^{\infty}A_i\right\| \\
&\leq \sqrt{qr}\left\{\left\|\underline{\hat{A}}_r - \underline{A}_r\right\| + \left\|\underline{A}_r^* - \underline{A}_r\right\|\right\} + \left\|\sum_{i=r+1}^{\infty}A_i\right\| \\
&= o_P(1).
\end{aligned}$$

Now note that

$$\hat{\Sigma}_{u_r} = \hat{\Sigma}_{\hat{S}} - \underline{\hat{A}}_r\hat{\Sigma}'_{\hat{S},\underline{\hat{S}}_r}$$

and, by (2.2)

$$\begin{aligned}
\Sigma_u &= Eu_tu_t' = Eu_tS_t' = E\left\{\left(S_t - \sum_{i=1}^{\infty}A_iS_{t-i}\right)S_t'\right\} \\
&= \Sigma_S - \sum_{i=1}^{\infty}A_iES_{t-i}S_t' = \Sigma_S - \underline{A}_r^*\Sigma'_{S,\underline{S}_r} - \sum_{i=r+1}^{\infty}A_iES_{t-i}S_t'.
\end{aligned}$$

Thus,

$$\begin{aligned}\left\|\hat{\Sigma}_{u_r} - \Sigma_u\right\| &= \Bigg\|\hat{\Sigma}_{\hat{S}} - \Sigma_S - \left(\underline{\hat{A}}_r - \underline{A}_r^*\right)\hat{\Sigma}'_{\hat{S},\underline{\hat{S}}_r} \\ &\quad -\underline{A}_r^*\left(\hat{\Sigma}'_{\hat{S},\underline{\hat{S}}_r} - \Sigma'_{S,\underline{S}_r}\right) + \sum_{i=r+1}^{\infty} A_i E S_{t-i} S'_t\Bigg\| \\ &\leq \left\|\hat{\Sigma}_{\hat{S}} - \Sigma_S\right\| + \left\|\left(\underline{\hat{A}}_r - \underline{A}_r^*\right)\left(\hat{\Sigma}'_{\hat{S},\underline{\hat{S}}_r} - \Sigma'_{S,\underline{S}_r}\right)\right\| \\ &\quad + \left\|\left(\underline{\hat{A}}_r - \underline{A}_r^*\right)\Sigma'_{S,\underline{S}_r}\right\| + \left\|\underline{A}_r^*\left(\hat{\Sigma}'_{\hat{S},\underline{\hat{S}}_r} - \Sigma'_{S,\underline{S}_r}\right)\right\| \\ &\quad + \left\|\sum_{i=r+1}^{\infty} A_i E S_{t-i} S'_t\right\|. \end{aligned} \tag{C.8}$$

In the right-hand side of this inequality, the first norm is $o_P(1)$ by Lemma C.4. By Lemmas C.6 and C.8, we have $\|\underline{\hat{A}}_r - \underline{A}_r^*\| = o_p(r^{-1/2}) = o_p(1)$, and by Lemma C.4, $\|\hat{\Sigma}'_{\hat{S},\underline{\hat{S}}_r} - \Sigma'_{S,\underline{S}_r}\| = o_p(r^{-1/2}) = o_p(1)$. Therefore the second norm in the right-hand side of (C.8) tends to zero in probability. The third norm tends to zero in probability because $\|\underline{\hat{A}}_r - \underline{A}_r^*\| = o_p(1)$ and, by Lemma C.2, $\|\Sigma'_{S,\underline{S}_r}\| = O(1)$. The fourth norm tends to zero in probability because, in view of Lemma C.4, $\|\hat{\Sigma}'_{\hat{S},\underline{\hat{S}}_r} - \Sigma'_{S,\underline{S}_r}\| = o_p(1)$, and, in view of (C.3), $\|\underline{A}_r^*\|^2 \leq \sum_{i=1}^{\infty} \mathrm{Tr}(A_i A'_i) < \infty$. Clearly, the last norm tends to zero, which completes the proof. □

Additional references

Boubacar Mainassara, Y., Carbon, M. and Francq, C. (2011) Computing and estimating information matrices of weak ARMA models. *Computational Statistics and Data Analysis* DOI: 10.1016/j.csda.2011.07.006

Francq, C., Roy, R. and J-M. Zakoïan (2005) Diagnostic checkig in ARMA Models with Uncorrelated Errors. *Journal of the American Statistical Association* 100, 532–544.

Francq, C. and Zakoïan, J-M. (2010) *GARCH Models: Structure, Statistical Inference and Financial Applications.* Wiley, Chichester.