Exact confidence sets and goodness-of-fit methods for stable distributions*

Marie-Claude Beaulieu † Université Laval

Jean-Marie Dufour ‡ McGill University

Lynda Khalaf § Carleton University

October 30, 2013

* The authors thank two anonymous referees and the Editor Jeong Kurz-Kim for several useful comments. This work was supported by the William Dow Chair in Political Economy (McGill University), the Bank of Canada (Research Fellowship), the Toulouse School of Economics (Pierre-de-Fermat Chair of excellence), the Universitad Carlos III de Madrid (Banco Santander de Madrid Chair of excellence), the Chair on the Economics of Electric Energy (Université Laval), the Social Sciences and Humanities Research Council of Canada, the Natural Sciences and Engineering Research Council of Canada, the Fonds de recherche sur la société et la culture (Québec), the Institut de Finance mathématique de Montréal (IFM2), a Guggenheim Fellowship, a Konrad-Adenauer Fellowship (Alexander-von-Humboldt Foundation, Germany), and the Canadian Network of Centres of Excellence [program on *Mathematics of Information Technology and Complex Systems* (MITACS)].

† Chaire RBC en innovations fiancières, Centre Interuniversitaire de Recherche sur les Politiques Économiques et de l'Emploi (CIRPÉE), Département de finance, assurance et immobilier, Pavillon Palasis-Prince, local 3632, Université Laval, Québec, Canada G1K 7P4. TEL: (418) 656-2131-2926, FAX: (418) 656-2624. email: Marie-Claude.Beaulieu@fsa.ulaval.ca

‡ William Dow Professor of Economics (Department of Economics, McGill University), Centre interuniversitaire de recherche en économie quantitative (CIREQ), Centre interuniversitaire de recherche en analyse des organisations (CIRANO). Mailing address: Leacock Building, Rm 519, 855 Sherbrooke Street West, Montréal, Québec, Canada H3A 2T7. TEL: (514) 398 8879; FAX: (514) 398 4938; e-mail: jean-marie.dufour@mcgill.ca.

§ Groupe de recherche en économie de l'énergie, de l'environnement et des ressources naturelles (GREEN) Université Laval, Centre interuniversitaire de recherche en économie quantitative (CIREQ), and Economics Department, Carleton University. Mailing address: Economics Department, Carleton University, Loeb Building 1125 Colonel By Drive, Ottawa, Ontario, K1S 5B6 Canada. Tel (613) 520-2600-8697; FAX: (613)-520-3906. email: Lynda_Khalaf@carleton.ca.

ABSTRACT

Usual inference methods for stable distributions are typically based on limit distributions. But asymptotic approximations can easily be unreliable in such cases, for standard regularity conditions may not apply or may hold only weakly. This paper proposes finite-sample tests and confidence sets for tail thickness and asymmetry parameters (α and β) of stable distributions. The confidence sets are built by inverting exact goodness-of-fit tests for hypotheses which assign specific values to these parameters. We propose extensions of the Kolmogorov-Smirnov, Shapiro-Wilk and Filliben criteria, as well as the quantile-based statistics proposed by McCulloch (1986) in order to better capture tail behavior. The suggested criteria compare empirical goodness-of-fit or quantile-based measures with their hypothesized values. Since the distributions involved are quite complex and non-standard, the relevant hypothetical measures are approximated by simulation, and p-values are obtained using Monte Carlo (MC) test techniques. The properties of the proposed procedures are investigated by simulation. In contrast with conventional wisdom, we find reliable results with samples sizes as small as 25. The proposed methodology is applied to daily electricity price data in the U.S. over the period 2001-2006. The results show clearly that heavy kurtosis and asymmetry are prevalent in these series.

Key words: stable distribution; skewness; asymmetry; exact test; Monte Carlo test; specification test; goodness-of-fit; tail parameter; electricity price.

1. Introduction

Drawing inference on the parameters of stable distributions is an enduring statistical problem. Such distributions appear in general central limit theorems, and thus provide an attractive alternative to the Gaussian distribution. So they are commonly considered in various fields of statistics, econometrics and finance; see Mandelbrot (1963), Ibragimov and Linnik (1975), Zolotarev (1986), Samorodnitsky and Taqqu (1994), Embrechts, Klüppelberg and Mikosch (1997), Rachev, Kim and Mittnik (1999*a*, 1999*b*), Rachev and Mittnik (2000), and Dufour, Kurz-Kim and Palm (2010).

In finance, stable distributions are often considered to account for heavy tails and asymmetries typically observed in financial returns and speculative price data. Furthermore, the family of stable distributions is consistent with optimization arguments underlying fundamental financial models such as the Capital Asset Pricing Model (CAPM); see, for example, Mandelbrot (1963), Samuelson (1967), and our own reviews and applications in Dufour, Khalaf and Beaulieu (2003) and Beaulieu, Dufour and Khalaf (2005).

In the latter papers, we consider asset pricing models based on multivariate linear regressions with stable error distributions, and we derive tests for the efficiency of the market portfolio (zero intercepts), allowing for stable error distributions with unknown tail thickness and asymmetry parameters (α and β). To estimate these parameters, we "inverted" goodness-of-fit (**GF**) tests based on multivariate kurtosis and skewness coefficients computed from model residuals. By test "inversion", we mean the operation of finding the set of parameter values which are not rejected by the test. We found that abnormal returns are less prevalent when skewness is allowed, so allowing for skewness has crucial implications for testing asset pricing models. These results also indicate that inference on the asymmetry parameter tends to be much less precise than inference on the tail parameter. Indeed, the distribution is fundamentally determined by the vector (α, β), and there is generally no reason why the values of its components could be separately determined in a precise way. This suggests that inference should focus on the pair (α, β), so a joint approach may be more informative.

In this paper, we reconsider the problem of building *joint* confidence sets for the tail and skewness parameters of a stable distribution, with the view of improving inference on the skewness parameter. Almost invariably, tests and confidence sets which have been proposed for inference on models with stable distributions are based on asymptotic approximations. The latter can easily be unreliable, since standard regularity conditions and asymptotic distributional theory may easily not apply [or apply only weakly] to such distributions. Consequently, it is important from an inference viewpoint that we approach this problem from a finite-sample perspective.

Stable distributions, despite their analytical complexities, can be easily simulated; see Chambers, Mallows and Stuck (1976), and Weron (1996). To get reliable inference, we thus use the technique of Monte Carlo (MC) tests. This method [originally proposed by Dwass (1957) and Barnard (1963)] is an exact simulation-based test procedure related to the parametric bootstrap in the sense that the distribution of the test statistic is simulated under the null hypothesis.[1] While

[1] See, for example, Dufour (2006), Dufour and Khalaf (2001, 2002*a*, 2002*b*, 2003), Dufour and Kiviet (1996, 1998), Kiviet and Dufour (1997), Dufour, Farhat, Gardiol and Khalaf (1998), Dufour, Khalaf, Bernard and Genest (2004), Dufour et al. (2003).

typical bootstrap methods are justified only asymptotically, the level of a MC test can be controlled in finite samples as soon as the distribution of the test statistic under the null hypothesis can be simulated once parameter values are set. Test statistics with very complicated distributions may thus be considered; the existence of a limiting distribution is not even required, which is particularly relevant for stable distributions.

As illustrated in Dufour et al. (1998), Dufour and Khalaf (2001) and Dufour et al. (2003), MC test methods are especially suited for testing goodness-of-fit. In this paper, we exploit our earlier research into such problems for inference on the parameters of stable distributions. To be more specific, we derive exact joint confidence sets for the tail and asymmetry parameters by inverting exact GF tests. The tests we propose to invert are new and provide useful model diagnostics.

MC test methods are used in two ways in our analysis. Typically, GF test criteria compare sample measures, *e.g.* moments, order statistics or the empirical distribution function (**EDF**), with hypothesized values, and discrepancies between the *observed* and *hypothesized* (or *expected*) measures suggest that the null hypothesis should be rejected. Two difficulties must be addressed in this process. *First*, computing the hypothesized measure may not be straightforward. In the case of stable laws, a simple closed-form expression is not even available for the density or distribution function. *Second*, GF test statistics often have complex null distributions. In many cases, even limiting null distributions are not pivotal. As a matter of fact, both difficulties remain present even in the Gaussian case; see Thode (2002), Dufour et al. (1998) and the references therein.

Here we approach both problems via a *two-stage* MC test procedure. In the first stage, we obtain *simulation-based estimates* of the hypothesized (or expected) measures considered; in the *second stage*, we obtain test p-values by the MC test technique. The parameter pairs for which the p-values are greater than the level α_* constitute a confidence set with level $1-\alpha_*$.

Our methodology considerably expands the class of statistics which can be used for building confidence sets. We use extensions of the Kolmogorov-Smirnov [Kolmogorov (1933), Smirnov (1939)], Shapiro-Wilk [Shapiro and Wilk (1965)] and Filliben (1975) criteria, as well as the quantile-based statistics proposed by McCulloch (1986), which we consider to capture tail behavior. Our results provide further avenues for development in general GF testing problems.

The properties of the proposed procedures are investigated in a large-scale Monte Carlo study. Since the size of the tests we invert is controlled by construction, our simulation study allows us to precisely assess their effective power advantages. This experiment reveals notable power differences in testing the skewness parameter.

We also apply the proposed methodology to electricity price data. This empirical analysis illustrates the usefulness of our joint estimation approach. We study the on-peak (daily) electricity spot price initially provided by ICAP US, over the period from January 3, 2001 to May 15, 2006. We assess the fit of a stable distribution to this series and derive joint confidence regions for both skewness and tail index parameters. Results may differ depending on the tests considered. Overall, however, our confidence sets reveal heavy kurtosis and asymmetries in the series analyzed.

The paper is organized as follows. Section 2 specifies the statistical framework under consideration. In section 3, we present the proposed inference methods. In section 4, we report the results of an illustrative MC study. The empirical application is discussed in section 5. Section 6 concludes.

2. Framework

If a random variable Y follows a stable distribution $S(\mu, \sigma, \alpha, \beta)$, where μ, σ, α and β represent location, scale, tail and skewness parameters, then its characteristic function $\phi(t)$ takes the form:

$$\phi(t) = \mathsf{E}\big[\exp(itY)\big] = \begin{cases} \exp\left\{-\sigma^{\alpha}|t|^{\alpha}\left[1 - i\beta\,\mathrm{sgn}(t)\tan(\pi\alpha/2)\right] + i\mu t\right\}, & \text{for } \alpha \neq 1, \\ \exp\left\{-\sigma|t|\left[1 + i\beta(2/\pi)\,\mathrm{sgn}(t)\ln|t|\right] + i\mu t\right\}, & \text{for } \alpha = 1, \end{cases}$$

where $0 < \alpha \leq 2$ and $-1 \leq \beta \leq 1$, and $\mathrm{sgn}(t)$ is the sign function, *i.e.*

$$\mathrm{sgn}(t) = \begin{cases} 1, & \text{if } t > 0 \\ 0, & \text{if } t = 0 \\ -1, & \text{if } t < 0 \end{cases} .$$

For inference on α and β, location and scale parameterizations raise difficult issues. For example, the above parametrization for the location parameter in $\phi(t)$ implies a discontinuity in the distribution around $\alpha = 1$ when $\beta \neq 0$. The scale parameter in the expression for $\phi(t)$ also involves an irregularity at $\alpha = 1$, which calls for caution in interpreting usual standardizations; see McCulloch (1996). We thus consider the following location-scale representation:

$$Y_i = \mu + \sigma y_i, \tag{2.1}$$

$$y_i \overset{i.i.d.}{\sim} \mathscr{S}(0, 1, \alpha, \beta), \quad i = 1, \ldots, n, \tag{2.2}$$

where Y_i is a set of n *i.i.d.* observations, in which case we propose inference methods on α and β that are invariant to μ and σ. Invariance, which we prove analytically in finite samples, ensures adequate size despite such irregularities.

As it is well known, a simple closed-form expression is not available for stable distributions (except in special cases); for a review of the properties of these distributions, see Samorodnitsky and Taqqu (1994, Chapter 1) and Rachev and Mittnik (2000, Chapter 2). Here we exploit the following limit result which characterizes the tail of a stable random variable $Y \sim S(\mu, \sigma, \alpha, \beta)$: for $0 < \alpha < 2$,

$$\lim_{\lambda \to \infty} \lambda^{\alpha} P[Y > \lambda] = C_{\alpha}\left(\frac{1+\beta}{2}\right)\sigma^{\alpha}, \tag{2.3}$$

$$\lim_{\lambda \to \infty} \lambda^{\alpha} P[Y < -\lambda] = C_{\alpha}\left(\frac{1-\beta}{2}\right)\sigma^{\alpha}, \tag{2.4}$$

$$C_{\alpha} = \left(\int_0^{\infty} x^{-\alpha}\sin x\,dx\right)^{-1} = \begin{cases} 1/[\Gamma(2-\alpha)\cos(\pi\alpha/2)], & \text{for } \alpha \neq 1, \\ 2/\pi, & \text{for } \alpha = 1; \end{cases} \tag{2.5}$$

see see Samorodnitsky and Taqqu (1994, pages 16-17). We use this expression for one class of statistics we introduce, to approximate the tail distribution of a standardized stable distribution as follows:

$$F_\infty(x;\alpha,\beta) = 1 - G_\infty(x;\alpha,\beta) = \begin{cases} C_\alpha\left(\frac{1+\beta}{2}\right)/\left|x^\alpha\right|, & x>0 \\ C_\alpha\left(\frac{1-\beta}{2}\right)/\left|x^\alpha\right|, & x<0 \end{cases}. \tag{2.6}$$

Alternative expressions are available and may be considered for special cases, including $\beta = 1$; see Samorodnitsky and Taqqu (1994, Chapter 1). Results in this paper rely on (2.6); our methodology [subject to some conditions discussed in section 3] can however be extended to alternative approximations for the tail probabilities.

Random variables with stable distributions can easily be simulated; see Chambers et al. (1976) and Weron (1996). All simulations performed in this paper apply Weron (1996), which we reproduce here for completeness. Generate, independently, a random variable $\mathscr{V}$, uniformly distributed over $(-\pi/2,\pi/2)$, and an exponential random variable $\mathscr{W}$ with mean 1, and set

$$\mathscr{B}_{\alpha,\beta} = \frac{\arctan\left(\beta\tan\left(\frac{\pi\alpha}{2}\right)\right)}{\alpha}, \quad \mathscr{S}_{\alpha,\beta} = \left[1+\beta^2\tan^2\left(\frac{\pi\alpha}{2}\right)\right]^{1/(2\alpha)}.$$

Then,

$$\mathscr{Y} = \begin{cases} \mathscr{S}_{\alpha,\beta}\times\frac{\sin\left(\alpha\left(\mathscr{V}+\mathscr{B}_{\alpha,\beta}\right)\right)}{\left(\cos(\mathscr{V})\right)^{1/\alpha}}\times\left(\frac{\cos\left(\mathscr{V}-\alpha\left(\mathscr{V}+\mathscr{B}_{\alpha,\beta}\right)\right)}{\mathscr{W}}\right)^{(1-\alpha)/\alpha}, & \text{for } \alpha\neq 1, \\ \frac{2}{\pi}\left[\left(\frac{\pi}{2}+\beta\mathscr{V}\right)\tan\mathscr{V}-\beta\log\left(\frac{\mathscr{W}\cos(\mathscr{V})}{\frac{\pi}{2}+\beta\mathscr{V}}\right)\right], & \text{for } \alpha = 1, \end{cases}$$

provides a draw from the $S(0,1,\alpha,\beta)$ distribution.

3. Inference methods

We develop a comprehensive approach for joint estimation and GF. Formally, we build confidence sets by *inverting* a test for the null hypothesis (2.2) where

$$H_0(\alpha_0,\beta_0):\alpha=\alpha_0\,,\,\beta=\beta_0 \tag{3.1}$$

where α_0 and β_0 are given. The joint confidence set for α and β involves the pairs (α_0,β_0) which are not rejected by the test applied.

The tests we introduce for this purpose are modifications of GF tests. Procedures based on moments are studied in Dufour et al. (2003) and Beaulieu et al. (2005), where we noticed the difficulty of making inference on the asymmetry parameter β. To improve inference, we exploit here different GF, quantile and EDF-based approaches. Even though the location and scale parameters (μ and σ) are also unknown, it turns out these can be eliminated. This is done by replacing the original data by appropriately normalized observations. More precisely, we consider a standardization based on the sample median and interquartile range:

$$\hat{y}_i = \frac{Y_i - Y[50]}{Y[75]-Y[25]}, \quad i=1,\ldots,n, \tag{3.2}$$

where $Y[x]$ refers to the xth quantile of Y_i, $i = 1, \ldots, n$.

Theorem 3.1 LOCATION AND SCALE INVARIANCE. *In the context of* (2.1)-(2.2)*, the joint distribution of the standardized observations* $\hat{y}_i,\ i = 1, \ldots, n,$ *defined in* (3.2) *does not depend on* μ *and* σ.

PROOF. If we denote by $y[x]$ the xth quantile of y_i, $i = 1, \ldots, n$ where, as defined by (2.1), $y_i = (Y_i - \mu)/\sigma$, then clearly $y[x] = (Y[x] - \mu)/\sigma$, so

$$\hat{y}_i = \frac{(Y_i - \mu)/\sigma - (Y[50] - \mu)/\sigma}{(Y[75] - \mu)/\sigma - (Y[25] - \mu)/\sigma} = \frac{y_i - y[50]}{y[75] - y[25]}, \quad i = 1, \ldots, n. \tag{3.3}$$

By (2.2), $y_i \overset{i.i.d.}{\sim} \mathscr{S}(0, 1, \alpha, \beta)$, and (3.3) implies that the distribution of $\hat{y}_i$, $i = 1, \ldots, n$, does not depend on μ and σ. $\square$

For testing $H_0(\alpha_0, \beta_0)$, Theorem **3.1** entails that any statistic which depends on the data only through $\hat{y}_i,\ i = 1, \ldots, n,$ does not depend on nuisance parameters, since its distribution is completely determined by the distribution of $y_i \overset{i.i.d.}{\sim} \mathscr{S}(0, 1, \alpha, \beta),\ i = 1, \ldots, n$. For a similar invariance results with stable distribution, see Proposition 1 in Dufour and Kurz-Kim (2010). Clearly, the sample mean and standard deviation lead to a similar result. We rely on the median and interquartile range for power considerations, as will be illustrated in section 4. Let $\hat{y}_{(i)}$, $i = 1, \ldots, n,$ denote the order statistics corresponding to $\hat{y}_i$, $i = 1, \ldots, n$, and $\hat{y}[x]$ refer to xth quantile of $\hat{y}_i$, $i = 1, \ldots, n$.

3.1. Goodness-of-fit test inversion

We consider here three classes of tests. First, we extend the order-statistic based normality tests of Shapiro and Wilk (1965) and Filliben (1975) to stable distributions. Second, we define GF measures using the estimators of McCulloch (1986). Finally, we propose EDF-based methods. Besides global EDF procedures, we consider criteria focusing on the tail of the distribution. All criteria compare sample measures, defined below, with their hypothesized (population) values under $H_0(\alpha_0, \beta_0)$, and exact p-values are obtained via the MC test method. Hypothesized measures are approximated via a preliminary simulation. This feature is fully taken into account by the MC method, so the level of the tests remains controlled. , but using the latter does not affect the exactness of MC p-values.

Formally, given a test S with observed value S_0 and setting the number of MC simulations to N so that $\alpha_*(N+1)$ is an integer, we obtain MC p-values denoted $\hat{p}_N(S_0)$ or $\tilde{p}_N(S_0)$ depending on whether the distribution of S is continuous or not, such that for finite n and finite N,

$$\mathsf{P}\left[\hat{p}_N(S_0) \leq \alpha_*\right] = \alpha_* \text{ or } \mathsf{P}\left[\tilde{p}_N(S_0) \leq \alpha_*\right] = \alpha_*.$$

Details and algorithms are given in section 3.2.

The tests are inverted to build confidence set as follows. We assemble, numerically, the pairs (α, β) which are not rejected by each of the proposed tests at level α_*. We used a grid search and

$\alpha_* = 5\%$. Formally, for each test we invert, we obtain a subset of $\mathbb{R}^2$, denoted $\mathrm{CS}(\alpha, \beta; \alpha_*)$, such that

$$\mathsf{P}\big[(\alpha, \beta) \in \mathrm{CS}(\alpha, \beta; \alpha_*)\big] \geq 1 - \alpha_* \quad \text{for finite } n. \tag{3.4}$$

The shape of confidence regions so obtained is non-standard and there is no reason to expect convexity; *e.g.*, the union of two disjoint sets cannot be ruled out. Because the parameter spaces for α and β are bounded, the confidence regions will not be unbounded; this is worth noting since confidence sets obtained by test inversion are not bounded by construction. *Ex ante*, there are no theoretical grounds to describe the resulting regions in any further specific way.

Moving from $\mathrm{CS}(\alpha, \beta; \alpha_*)$ to individual confidence sets for each of α and β is achieved by projecting this region. By definition, a projection-based confidence set can be obtained for any function $g(\alpha, \beta)$ by minimizing and maximizing the function $g(\alpha, \beta)$ over the α and β values included in $\mathrm{CS}(\alpha, \beta; \alpha_*)$. Confidence intervals so obtained are simultaneous, in the sense that valid inference on any arbitrary number of transformations of the (α, β) pair is feasible ensuring overall level control; see Miller (1981), Dufour (1989), Abdelkhalek and Dufour (1998) or Bolduc, Khalaf and Yelou (2010). Formally, for any set of m continuous real valued functions of the (α, β) pair, $g_i(\alpha, \beta) \in \mathbb{R}$, $i = 1, ..., m$, let $g_i\big(\mathrm{CS}(\alpha, \beta; \alpha_*)\big)$ denote the image of $\mathrm{CS}(\alpha, \beta; \alpha_*)$ by the function g_i. Clearly,

$$(\alpha, \beta) \in \mathrm{CS}(\alpha, \beta; \alpha_*) \Rightarrow g_i(\alpha, \beta) \in g_i\big(\mathrm{CS}(\alpha, \beta; \alpha_*)\big), i = 1, \ldots, m$$

hence

$$\mathsf{P}\big[g_i(\alpha, \beta) \in g_i\big(\mathrm{CS}(\alpha, \beta; \alpha_*)\big), \quad i = 1, \ldots, m\big] \geq \mathsf{P}\big[\alpha, \beta \in \mathrm{CS}(\alpha, \beta; \alpha_*)\big]. \tag{3.5}$$

Then equation (3.4) implies that

$$\mathsf{P}\big[g_i(\alpha, \beta) \in g_i\big(\mathrm{CS}(\alpha, \beta; \alpha_*)\big), \quad i = 1, \ldots, m\big] \geq 1 - \alpha_*, \quad \forall (\alpha, \beta). \tag{3.6}$$

It also follows that if $\mathrm{CS}(\alpha, \beta; \alpha_*)$ is empty, then (2.1)-(2.2), can be rejected at the considered test level, that is α_*.

3.1.1. Regression-based Shapiro-Wilks type criteria

The regression-based GF approach may be traced back to Shapiro and Wilk (1965) for the problem of testing normality. It consists in regressing the observed (sample) order statistics on a constant and the series of their means under the normality null hypothesis; tests for the significance of the regression slope serve to assess GF. Filliben (1975), again restricting focus to normality tests, suggested to replace, in the latter regression, the population means of order statistics by their population medians. These tests are left-tailed, for large values support the hypothesized distribution.

The (population) means or medians of order statistics for stable distributions are not available.[2]

[2]Beyond few special cases, for example, the Gaussian distribution for which Shapiro and Wilks provided specialized tables for given sample sizes, expected values and population medians of order statistics are unavailable. This literature acknowledges difficulties with various approximations even with Gaussian distributions.

We thus rely on simulation-based approximations, under the null hypothesis (3.1) which fixes α and β to given values α_0 and β_0. This is done as follows.

A1. Draw N_0 *i.i.d.* samples of size n from a stable distribution imposing (3.1);

A2. For each sample drawn, construct the order statistics; these yield N_0 realizations of each of the order statistics.

A3. The vector of empirical means (averages), denoted

$$\bar{s}(\alpha_0, \beta_0) = [\bar{s}_1(\alpha_0, \beta_0), \ldots, \bar{s}_n(\alpha_0, \beta_0)]',$$

and the vector of empirical medians

$$\tilde{s}(\alpha_0, \beta_0) = [\tilde{s}_1(\alpha_0, \beta_0), \ldots, \tilde{s}_n(\alpha_0, \beta_0)]',$$

of the N_0 simulated values for each order statistic yield the desired approximation.

Comparing $\bar{s}(\alpha_0, \beta_0)$ or $\tilde{s}(\alpha_0, \beta_0)$ which impose (3.1) to the vector of observed order statistics $\hat{y}_{(i)}$, $i = 1, \ldots, n$, allows one to assess the acceptability of (3.1). For this purpose, we propose to use the coefficient of determination [R^2], denoted $\rho\left[\hat{y}_{(i)}, \bar{s}_i(\alpha_0, \beta_0)\right]$, associated with regressing $\hat{y}_{(1)}, \ldots, \hat{y}_{(n)}$ on a constant and $\bar{s}_1(\alpha_0, \beta_0), \ldots, \bar{s}_n(\alpha_0, \beta_0)$. Alternatively, we consider the R^2, denoted $\rho\left[\hat{y}_{(i)}, \tilde{s}_i(\alpha_0, \beta_0)\right]$, from the regression of $\hat{y}_{(1)}, \ldots, \hat{y}_{(n)}$ on a constant and $\tilde{s}_1(\alpha_0, \beta_0), \ldots, \tilde{s}_n(\alpha_0, \beta_0)$. For convenience, we subtract the coefficients of determination from one, to obtain the right-tailed tests:

$$SW(\alpha_0, \beta_0) = 1 - \rho\left[\hat{y}_{(i)}, \bar{s}_i(\alpha_0, \beta_0)\right], \tag{3.7}$$
$$FB(\alpha_0, \beta_0) = 1 - \rho\left[\hat{y}_{(i)}, \tilde{s}_i(\alpha_0, \beta_0)\right]. \tag{3.8}$$

3.1.2. Quantile-based criteria

We next consider two test statistics based on the estimators of McCulloch (1986):

$$\hat{\phi}_1(\alpha_0, \beta_0) = \left|\phi_1 - \bar{\phi}_1(\alpha_0, \beta_0)\right|, \quad \phi_1 = \frac{\hat{y}[95] - \hat{y}[5]}{\hat{y}[75] - \hat{y}[25]}, \tag{3.9}$$
$$\hat{\phi}_2(\alpha_0, \beta_0) = \left|\phi_2 - \bar{\phi}_2(\alpha_0, \beta_0)\right|, \quad \phi_2 = \frac{\hat{y}[95] + \hat{y}[5] - 2\hat{y}[50]}{\hat{y}[95] - \hat{y}[5]}, \tag{3.10}$$

where $\bar{\phi}_1(\alpha_0, \beta_0)$ and $\bar{\phi}_2(\alpha_0, \beta_0)$ are the hypothesized values for ϕ_1 and ϕ_2 imposing (3.1). We estimate $\bar{\phi}_1(\alpha_0, \beta_0)$ and $\bar{\phi}_2(\alpha_0, \beta_0)$ by simulation, so we do not need any specialized tables to conduct these tests. The algorithm we use may be summarized as follows.

B1. Draw N_0 *i.i.d.* samples of size n from a stable distribution imposing (3.1).

B2. For each sample drawn, construct the quantiles which appear in the formulas for ϕ_1 and ϕ_2; these yield N_0 realizations of the measures under consideration.

B3. The average across the N_0 simulated values of each measure yield $\bar{\phi}_1(\alpha_0, \beta_0)$ and $\bar{\phi}_2(\alpha_0, \beta_0)$.

3.1.3. Empirical distribution function measures

Tests based on the EDF are naturally described as GF tests. Here we extend to the case of (3.1), three of the most popular EDF criteria of the Kolmogorov-Smirnov and Anderson-Darling type form, which we adapt to target the tail behavior of the distribution as follows.

Let $\hat{F}_n(x)$ refer to the EDF of the sample $\hat{y}_1, \ldots, \hat{y}_n$, that is, $\hat{F}_n(x)$ equals the proportion of observations $\hat{y}_1, \ldots, \hat{y}_n$ that are less than or equal to x. We consider the following three statistics, which differ regarding the weights attributed to the observed and hypothesized distributions:

$$KSA_1(\alpha_0, \beta_0) = \max_x \left\{ \sqrt{n} \left| \hat{F}_n(x) - \bar{F}_0(x; \alpha_0, \beta_0) \right| \right\}, \tag{3.11}$$

$$KSA_2(\alpha_0, \beta_0) = \max_x \left\{ \sqrt{n} \left| \frac{\hat{F}_n(x) - \bar{F}_0(x; \alpha_0, \beta_0)}{\bar{F}_0(x; \alpha_0, \beta_0) \left[1 - \bar{F}_0(x; \alpha_0, \beta_0) + 1/n\right]} \right| \right\}, \tag{3.12}$$

$$KSA_3(\alpha_0, \beta_0) = \max_x \left\{ \sqrt{n} \left| \frac{\hat{F}_n(x) - \bar{F}_0(x; \alpha_0, \beta_0)}{\hat{F}_n(x) \left[1 - \hat{F}_n(x)\right] + 1/n} \right| \right\}, \tag{3.13}$$

where $\bar{F}_0(x; \alpha_0, \beta_0)$ is a simulation based estimate of the hypothesized distribution, derived as follows.

C1. Draw one *i.i.d.* sample of size $\bar{n}$, denoted $y_1(\alpha_0, \beta_0), \ldots, y_{\bar{n}}(\alpha_0, \beta_0)$, from a standard stable distribution imposing (3.1).[3]

C2. For any x, define $\bar{F}_0(x; \alpha_0, \beta_0)$ as the proportion of the simulated observations $y_1(\alpha_0, \beta_0), \ldots, y_{\bar{n}}(\alpha_0, \beta_0)$ that are less than or equal to x.

Given our interest on heavy tailed distributions, we also consider variants of the latter EDF statistics which focus on the tail of the distribution:

$$KSC_1(\alpha_0, \beta_0) = \max_{x \in \Lambda(x)} \left\{ \sqrt{n} \left| \hat{F}_n(x) - \bar{F}_0(x; \alpha_0, \beta_0) \right| \right\}, \tag{3.14}$$

$$KSC_2(\alpha_0, \beta_0) = \max_{x \in \Lambda(x)} \left\{ \sqrt{n} \left| \frac{\hat{F}_n(x) - \bar{F}_0(x; \alpha_0, \beta_0)}{\bar{F}_0(x; \alpha_0, \beta_0) \left[1 - \bar{F}_0(x; \alpha_0, \beta_0) + 1/n\right]} \right| \right\}, \tag{3.15}$$

$$KSC_3(\alpha_0, \beta_0) = \max_{x \in \Lambda(x)} \left\{ \sqrt{n} \left| \frac{\hat{F}_n(x) - \bar{F}_0(x; \alpha_0, \beta_0)}{\hat{F}_n(x) \left[1 - \hat{F}_n(x)\right] + 1/n} \right| \right\}, \tag{3.16}$$

where

$$\Lambda(x) = \left\{ x \leq \bar{\lambda}_1 \text{ or } x \geq \bar{\lambda}_2 \right\}$$

[3]The size of the simulated sample, $\bar{n}$, is not necessarily equal to the size of the observed sample, n. A large $\bar{n}$ is recommended. Our simulation study uses $\bar{n} = 1000$ for all considered values of n. Our empirical analysis uses $\bar{n} = 2000$.

and $\bar{\lambda}_1$ and $\bar{\lambda}_2$ correspond to a cut-off points which allow us to focus on a specific region (here, in the tail) of the hypothesized distribution. In our simulation and empirical illustrations, we use the 5th and 95th percentiles of the *simulated* sample $y_1(\alpha_0, \beta_0), \dots, y_{\bar{n}}(\alpha_0, \beta_0)$.

Alternative approximations for the tail distribution function which underlie (3.14)–(3.16) may be also used. We consider the $F_\infty(x; \alpha_0, \beta_0)$ approximation as defined in (2.6), which leads to:

$$
\begin{aligned}
KST_1(\alpha_0, \beta_0) &= \max_{x \in \Lambda(x)} \left\{ \sqrt{n} \left| \hat{F}_n(x) - F_\infty(x; \alpha_0, \beta_0) \right| \right\}, && (3.17) \\
KST_2(\alpha_0, \beta_0) &= \max_{x \in \Lambda(x)} \left\{ \sqrt{n} \left| \frac{\hat{F}_n(x) - F_\infty(x; \alpha_0, \beta_0)}{F_\infty(x; \alpha_0, \beta_0)\left[1 - F_\infty(x; \alpha_0, \beta_0)\right] + 1/n} \right| \right\}, && (3.18) \\
KST_3(\alpha_0, \beta_0) &= \max_{x \in \Lambda(x)} \left\{ \sqrt{n} \left| \frac{\hat{F}_n(x) - F_\infty(x; \alpha_0, \beta_0)}{\hat{F}_n(x)\left[1 - \hat{F}_n(x)\right] + 1/n} \right| \right\}. && (3.19)
\end{aligned}
$$

Statistics so obtained would be less costly and do not call for a first-stage simulation. Interestingly, the MC technique [described below] allows us to obtain exact p-values for the latter statistics, even though limiting distributions are used in their formulation. For further reference on the unconventional use of asymptotic p-values to derive exact tests, see Dufour, Khalaf and Beaulieu (2010).

3.2. Finite sample p-values

Clearly, the test criteria introduced above complex null distributions which may be difficult to establish analytically in both finite and large samples. Yet these distributions can be easily simulated which justifies the application of Monte Carlo tests [Dufour (2006)]. The general MC test methodology proceeds as follows.

Let S_0 denote the test statistic calculated from the observed data set; generate N replications $S_1, \dots, S_N$ of the test statistic S so that $S_0, S_1, \dots, S_N$ are **exchangeable**. Given the latter series, compute $\hat{p}_N(S_0)$ where

$$
\hat{p}_N(x) = \frac{N\hat{G}_N(x) + 1}{N + 1}, \quad \hat{G}_N(x) = \hat{G}_N[x; S(N)] = \frac{1}{N} \sum_{i=1}^{N} \mathbb{1}(S_i \geq x), \tag{3.20}
$$

$S(N) = (S_1, \dots, S_N)'$ and $\mathbb{1}(C)$ is the indicator function associated with condition C:

$$
\begin{aligned}
\mathbb{1}(C) &= 1, \text{ if condition } C \text{ holds} \\
&= 0, \text{ otherwise.}
\end{aligned}
$$

In other words, $N\hat{G}_N(S_0)$ is the number of simulated values greater than or equal to S_0. The MC critical region is: $\hat{p}_N(S_0) \leq \alpha_*$, $0 < \alpha_* < 1$.

If the distribution of S is continuous and $\alpha_*(N+1)$ is an integer, then

$$
\mathsf{P}\left[\hat{p}_N(S_0) \leq \alpha_*\right] = \alpha_*.
$$

Some of the statistics we consider, particularly the truncated EDF-based ones, have possibly discontinuous distributions. The technique of MC tests can be adapted for discrete distributions using the following randomized tie-breaking procedure [for proofs and further references, see Dufour (2006)].

Draw $N+1$ uniformly distributed variates $Z_0, Z_1, \ldots, Z_N$ independently of $S(N)$ and arrange the pairs (S_j, Z_j) following the lexicographic order:

$$(S_i, Z_i) \geq (S_j, Z_j) \Leftrightarrow [S_i > S_j \quad or \quad (S_i = S_j \quad and \quad Z_i \geq Z_j)]\,. \tag{3.21}$$

This leads to the MC p-value $\tilde{p}_N(S_0)$ where

$$\tilde{p}_N(x) = \frac{N\tilde{G}_N(x)+1}{N+1}, \tag{3.22}$$

$$\begin{aligned}\tilde{G}_N(x) &= \tilde{G}_N\left[x; Z_0, S(N), Z(N)\right] \\ &= 1 - \frac{1}{N}\sum_{i=1}^{N} \mathbb{1}(S_i \leq x) + \frac{1}{N}\sum_{i \in E_N(x)}^{N} \mathbb{1}(Z_i \leq Z_0)\end{aligned} \tag{3.23}$$

and $Z(N) = (Z_1, \ldots, Z_N)'$, $E_N(x) = \{i : S_i = x,\ 1 \leq x \leq N\}$. The resulting critical region is $\tilde{p}_N(S_0) \leq \alpha_*$, $0 < \alpha_* < 1$. If $\alpha_*(N+1)$ is an integer, then

$$\mathsf{P}\left[\hat{p}_N(S_0) \leq \alpha_*\right] \leq \mathsf{P}\left[\tilde{p}_N(S_0) \leq \alpha_*\right] = \alpha_*.$$

When applied to the above GF criteria, the MC test technique can be summarized as follows. Note that step D1 is not needed for the statistics (3.17) - (3.18) - (3.19).

D1. We obtain the above described approximations for the population measures underlying all considered statistics. Specifically, we implement algorithm A1 - A3 to derive $\bar{s}(\alpha_0, \beta_0)$ or $\tilde{s}(\alpha_0, \beta_0)$, algorithm B1-B3 to derive $\bar{\phi}_1(\alpha_0, \beta_0)$ and $\bar{\phi}_2(\alpha_0, \beta_0)$, and algorithm C1-C2 which serves to construct the function $\bar{F}_0(x; \alpha_0, \beta_0)$, using the reference sample $y_1(\alpha_0, \beta_0)$, ... , $y_{\bar{n}}(\alpha_0, \beta_0)$. All these population measures are generated *only once*, so the next steps are conditional on these estimates.

D2. Applying (3.7)-(3.8), (3.9)-(3.10) and (3.11)-(3.19) to the data, we find the observed value of each test statistic.

D3. Independently of the step D1, we draw N *i.i.d.* samples of size n from a stable distribution under (3.1), and standardize the simulated observations for each draw, using the median and interquartile range of each simulated samples.

D4. Using the same population measures derived in D1, and applying (3.7)-(3.8), (3.9)-(3.10) and (3.11) - (3.19) to the simulated data, we obtain N simulated values for each test statistic considered.

D5. We can then compute a simulated p-value, for any one of the test statistics, using the rank of the observed statistic, relative to its simulated counterpart; see (3.20) or (3.22). The null hypothesis is rejected at level α_* by each of the test considered if the MC p-value so obtained is less than or equal to α_*.

Because the above MC test procedure involves two levels of simulations (a first one to approximate the population measures, and a second one to get the test statistics) for all statistics except (3.17) - (3.18) - (3.19), we call it a *two-stage MC test*. It is important to emphasize a key step in the above algorithm: the observed and simulated statistics must rely on the *same* approximated population measures; in this way, the observed and simulated statistics are (by construction) exchangeable under the null hypothesis, which yields size control; see Dufour (2006) and Dufour et al. (2003) or Beaulieu, Dufour and Khalaf (2007). The underlying simulations are non-independent but remain exchangeable, which is sufficient to ensure exactness as shown by Dufour (2006). In addition to minimizing noise which may affect power, using the same approximated population measures implies important execution cost savings. The fact that exchangeability is sufficient from a finite-sample perspective is worth pointing out here since all statistics we propose rely on just one preliminary simulation.

3.3. Combined statistics

The MC test technique can also be applied to combine the above statistics; see Dufour et al. (2003), Dufour and Khalaf (2002*a*), Dufour et al. (2004), Dufour, Khalaf and Beaulieu (2010), Beaulieu, Dufour and Khalaf (2013). Combining our modified version of McCulloch's (1986) statistics $\hat{\phi}_1(\alpha_0, \beta_0)$ and $\hat{\phi}_2(\alpha_0, \beta_0)$ is the most relevant question, since the former is originally designed to focus on α_0 and the latter on β_0. To avoid relying on Boole-Bonferroni rules for this purpose, we use the following combined statistics:

$$\hat{\phi}(\alpha_0, \beta_0) = 1 - \min\left\{\hat{p}_N\big(\hat{\phi}_1(\alpha_0, \beta_0)\big),\, \hat{p}_N\big(\hat{\phi}_2(\alpha_0, \beta_0)\big)\right\}, \tag{3.24}$$

$$\tilde{\phi}(\alpha_0, \beta_0) = 1 - \min\left\{\tilde{p}_N\big(\hat{\phi}_1(\alpha_0, \beta_0)\big),\, \tilde{p}_N\big(\hat{\phi}_2(\alpha_0, \beta_0)\big)\right\}. \tag{3.25}$$

Such a combination method allows us to reject the null hypothesis if at least one of the individual tests is significant; for convenience, we subtract the minimum p-value from one to obtain a right-sided test. The MC test technique may once again be applied to obtain a test based on the combined statistic; details of the algorithm can be summarized as follows, for the case of $\hat{\phi}(\alpha_0, \beta_0)$. The algorithm can be easily adapted to the case of $\tilde{\phi}(\alpha_0, \beta_0)$ replacing the survival function $\hat{G}_N[x; S(N)]$ by $\tilde{G}_N[x; S(N)]$ in what follows.

- E1. According to steps B1-B3, generate $\bar{\phi}_1(\alpha_0, \beta_0)$ and $\bar{\phi}_2(\alpha_0, \beta_0)$; conformably, calculate the observed value of $\hat{\phi}_1(\alpha_0, \beta_0)$ and $\hat{\phi}_2(\alpha_0, \beta_0)$ [denoted $\hat{\phi}_1^0(\alpha_0, \beta_0)$ and $\hat{\phi}_2^0(\alpha_0, \beta_0)$ respectively], and the N corresponding simulated statistics using the same $\bar{\phi}_1(\alpha_0, \beta_0)$ and $\bar{\phi}_2(\alpha_0, \beta_0)$.
- E2. For each test statistic, obtain the "survival function" $\hat{G}_N[x; S(N)]$ defined in (3.20) determined by the simulated statistics.
- E3. Independently of the previous simulations and the data, generate N additional *i.i.d.* realizations from a stable distribution under (3.1) each of size n, and standardize the simulated observations for each draw, using the median and interquartile range of the simulated samples.

E4. Using $\bar{\phi}_1(\alpha_0, \beta_0)$ and $\bar{\phi}_2(\alpha_0, \beta_0)$, and the N draws generated at step E3, compute the corresponding simulated statistics: $\hat{\phi}_1^l(\alpha_0, \beta_0)$ and $\hat{\phi}_2^l(\alpha_0, \beta_0) \quad l = 1, \ldots, N$.

E5. Using the survival functions obtained at step E2, evaluate the simulated p-values for the observed and the N additional simulated statistics; specifically, obtain $\hat{G}_N\left[\hat{\phi}_1^l(\alpha_0, \beta_0); S(N)\right]$, $l = 0, 1, \ldots, N$, and $\hat{G}_N\left[\hat{\phi}_2^l(\alpha_0, \beta_0); S(N)\right]$ $l = 0, 1, \ldots, N$, using for $S(N)$ the same simulated series described in step E1; these lead to the p-values $\hat{p}_N[\hat{\phi}_1^l(\alpha_0, \beta_0)]$ and $\hat{p}_N[\hat{\phi}_2^l(\alpha_0, \beta_0)]$, $l = 0, 1, \ldots, N$.

E6. From the latter, compute the corresponding values of the combined test statistics:

$$\hat{\phi}^l(\alpha_0, \beta_0) = 1 - \min\left\{\hat{p}_N[\hat{\phi}_1^l(\alpha_0, \beta_0)], \hat{p}_N[\hat{\phi}_2^l(\alpha_0, \beta_0)]\right\}, \quad l = 0, 1, \ldots, N. \tag{3.26}$$

It is easy to see that the vectors $\hat{\phi}^l(\alpha_0, \beta_0), l = 0, 1, \ldots, N,$ are exchangeable.

E7. Applying $\hat{G}_N\left[\hat{\phi}^l(\alpha_0, \beta_0); \left(\hat{\phi}^1(\alpha_0, \beta_0), \ldots, \hat{\phi}^N(\alpha_0, \beta_0)\right)'\right]$ leads to the desired combined p-value.

The test based on the combined p-value described in steps E1-E7 has the correct level because the variables $\hat{\phi}^l(\alpha_0, \beta_0),\ l = 0, 1, \ldots, N,$ are exchangeable under the null hypothesis. We call this three nested simulation procedure a triple or three stage MC test. Here again, maintaining the *same* approximated population measures throughout ensures exchangeability, hence exactness.

4. Monte Carlo study

We conduct a MC study to assess the performance of the above proposed tests. We design the experiment as follows. Simulated samples with $n = 25$, 100, 250 and 1000 are drawn from model (2.1)-(2.2) with $\mu = 0$ and $\sigma = 1$. For inference on β, samples are generated with $\beta = 0$ and two choices for α: 1.5 and 1.9. For inference on α, samples are generated with $\alpha = 1.75$ and two choices for β: 0 and .5.

Size results associated in turn with $\beta = 0$ and $\alpha = 1.75$ are reported in Table 1. The power study for inference on β is reported in Table 2 for $\alpha = 1.5$ and 3 for $\alpha = 1.9$. In each of these tables, the hypothesized value for α is set to the value which was used to generate the samples (1.5 or 1.9, respectively) while the hypothesized value for β varies from .1 to 1. The power study for inference on α is reported in Table 4 for $\beta = 0$ and 5 for $\beta = .5$. In each of these tables, the hypothesized value for β is set to the value which was used to generate the samples (0 or .5, respectively) while the hypothesized value for α varies from .5 to 1.99.

$N = 199$ replications are considered for the MC tests. $N_0 = 1000$ and $\bar{n} = 1000$ is considered for all values of n. We set $\bar{\lambda}_1$ and $\bar{\lambda}_2$ at 5th and 95th percentiles of the simulated samples $y_1(\alpha_0, \beta_0), \ldots, y_{\bar{n}}(\alpha_0, \beta_0)$ defined in algorithm C1-C2. All tables report empirical rejections associated with a nominal size of 5% and 1000 replications. Results can be summarized as follows.

TABLE 1: Size of GF tests

Tests on β	$\alpha = 1.5; \beta = 0$				$\alpha = 1.9; \beta = 0$			
n	25	100	250	1000	25	100	250	1000
SW	.056	.052	.050	.044	.060	.050	.046	.046
FB	.059	.042	.044	.044	.058	.049	.050	.045
KSA_1	.047	.045	.058	.047	.064	.043	.050	.056
KSA_2	.045	.052	.056	.052	.039	.051	.049	.036
KSA_3	.055	.045	.053	.061	.052	.047	.047	.048
KSC_1	.040	.059	.061	.047	.044	.051	.057	.050
KSC_2	.043	.053	.056	.052	.039	.051	.049	.036
KSC_3	.045	.045	.052	.061	.056	.053	.048	.048
KST_1	.055	.050	.043	.039	.064	.058	.052	.040
KST_2	.055	.050	.043	.039	.064	.058	.052	.040
KST_3	.055	.050	.043	.039	.064	.058	.052	.040
$\hat{\phi}_1$	.058	.055	.039	.054	.056	.049	.047	.440
$\hat{\phi}_2$	.053	.042	.059	.054	.049	.058	.052	.062
Tests on α	$\alpha = 1.75; \beta = 0$				$\alpha = 1.75; \beta = .5$			
n	25	100	250	1000	25	100	250	1000
SW	.064	.048	.047	.046	.064	.046	.047	.045
FB	.059	.047	.047	.043	.054	.045	.043	.045
KSA_1	.054	.045	.060	.053	.056	.038	.054	.059
KSA_2	.043	.055	.045	.042	.039	.057	.052	.038
KSA_3	.052	.055	.049	.057	.054	.060	.052	.057
KSC_1	.042	.050	.064	.047	.039	.045	.057	.046
KSC_2	.041	.055	.045	.042	.040	.058	.052	.038
KSC_3	.045	.054	.050	.057	.044	.057	.051	.057
KST_1	.061	.052	.048	.035	.061	.051	.050	.046
KST_2	.061	.052	.048	.035	.061	.051	.050	.038
KST_3	.061	.052	.048	.035	.061	.051	.050	.057
$\hat{\phi}_1$	.056	.068	.045	.046	.053	.065	.040	.048
$\hat{\phi}_2$	.053	.043	.055	.038	.061	.038	.055	.045

Notes – SW and FB are our extensions of the Shapiro and Wilk (1965) and Filliben (1975) tests defined in section 3.1.1, which involve regressing the observed order statistics on a constant and the series of their population means (SW) or medians (FB) imposing the null hypothesis (3.1). KSA_i, $i = 1, ..., 3$ are our full sample weighted [KSA_2 and KSA_3] and standard [KSA_1] Kolmogorov-Smirnov type tests defined in section 3.1.3. KSC_i, $i = 1, ..., 3$ are their tail-based counterparts, for which the population tail distribution is simulation-based. KST_i, $i = 1, ..., 3$ rely on the asymptotic tail approximation (2.3) instead. $\hat{\phi}_1$ and $\hat{\phi}_2$ defined in section 3.1.2 assess the distance between the estimators of McCulloch (1986) and their hypothesized values imposing the null hypothesis (3.1), the latter approximated by simulation.

TABLE 2: Power, GF tests on β

	$\alpha=1.5;\beta=.3$				$\alpha=1.5;\beta=.5$				$\alpha=1.5;\beta=.7$			
n	25	100	250	1000	25	100	250	1000	25	100	250	1000
SW	.082	.073	.081	.072	.127	.135	.149	.150	.219	.278	.286	.281
FB	.076	.066	.060	.054	.094	.092	.087	.086	.128	.156	.155	.172
KSA_1	.065	.072	.117	.277	.074	.135	.225	.623	.079	.191	.425	.886
KSA_2	.080	.124	.135	.113	.131	.302	.389	.377	.203	.581	.783	.759
KSA_3	.090	.103	.108	.112	.114	.151	.159	.164	.135	.179	.189	.209
KSC_1	.072	.143	.201	.496	.106	.295	.225	.896	.148	.488	.817	.992
KSC_2	.080	.123	.135	.113	.128	.301	.389	.377	.201	.582	.783	.759
KSC_3	.066	.096	.107	.112	.074	.129	.159	.164	.086	.148	.188	.209
KST_1	.077	.085	.067	.059	.103	.120	.101	.088	.160	.179	.174	.152
KST_2	.077	.085	.067	.059	.103	.120	.101	.088	.160	.179	.174	.152
KST_3	.077	.085	.067	.059	.103	.120	.101	.088	.160	.179	.174	.152
$\hat{\phi}_1$	.073	.071	.044	.049	.086	.080	.049	.050	.113	.085	.052	.049
$\hat{\phi}_2$	.067	.149	.314	.809	.114	.377	.669	.992	.188	.642	.906	1.00
	$\alpha=1.5;\beta=.9$				$\alpha=1.5;\beta=.99$				$\alpha=1.5;\beta=1.0$			
n	25	100	250	1000	25	100	250	1000	25	100	250	1000
SW	.362	.467	.477	.496	.419	.546	.573	.613	.423	.558	.579	.623
FB	.209	.275	.326	.375	.254	.361	.422	.528	.262	.367	.429	.542
KSA_1	.085	.281	.620	.974	.091	.334	.715	.986	.087	.332	.728	.989
KSA_2	.281	.867	.992	.986	.308	.935	.999	.999	.308	.932	.999	1.00
KSA_3	.151	.194	.213	.243	.165	.209	.211	.238	.166	.210	.212	.224
KSC_1	.194	.647	.957	.999	.209	.697	.976	1.00	.209	.701	.978	1.00
KSC_2	.281	.867	.992	.986	.307	.935	.999	.999	.307	.932	.999	1.00
KSC_3	.096	.150	.208	.241	.096	.149	.205	.236	.097	.145	.205	.224
KST_1	.391	.459	.452	.406	.604	.977	.995	.993	.066	.054	.053	.053
KST_2	.391	.459	.452	.406	.604	.977	.995	.993	.066	.054	.053	.053
KST_3	.391	.459	.452	.406	.604	.977	.995	.993	.066	.054	.053	.053
$\hat{\phi}_1$	.159	.084	.051	.042	.199	.085	.051	.042	.203	.084	.050	.041
$\hat{\phi}_2$	.384	.818	.977	1.00	.541	.863	.985	1.00	.552	.868	.987	1.00

Notes – See Table 1.

TABLE 3: Power, GF tests on β - continued

	$\alpha = 1.9; \beta = .3$				$\alpha = 1.9; \beta = .5$				$\alpha = 1.9; \beta = .7$			
n	25	100	250	1000	25	100	250	1000	25	100	250	1000
SW	.065	.055	.049	.042	.069	.062	.060	.050	.079	.077	.068	.068
FB	.060	.049	.050	.046	.065	.051	.053	.046	.074	.058	.060	.053
KSA_1	.066	.050	.048	.062	.065	.050	.053	.062	.066	.053	.053	.083
KSA_2	.039	.070	.067	.055	.052	.088	.097	.125	.055	.113	.143	.267
KSA_3	.061	.070	.074	.103	.063	.095	.097	.146	.067	.109	.123	.189
KSC_1	.047	.058	.063	.080	.049	.061	.077	.125	.047	.076	.094	.171
KSC_2	.040	.070	.067	.055	.052	.088	.097	.125	.055	.113	.143	.267
KSC_3	.067	.077	.075	.103	.066	.082	.096	.146	.070	.093	.122	.189
KST_1	.076	.082	.074	.063	.086	.113	.103	.089	.109	.165	.171	.146
KST_2	.076	.082	.074	.063	.086	.113	.103	.089	.109	.165	.171	.146
KST_3	.076	.082	.074	.063	.086	.113	.103	.089	.109	.165	.171	.146
$\hat{\phi}_1$	.065	.050	.045	.040	.071	.049	.045	.038	.077	.053	.048	.041
$\hat{\phi}_2$	.058	.055	.067	.115	.073	.071	.085	.210	.088	.085	.110	.351
	$\alpha = 1.9; \beta = .9$				$\alpha = 1.9; \beta = .99$				$\alpha = 1.9; \beta = 1$			
n	25	100	250	1000	25	100	250	1000	25	100	250	1000
SW	.092	.097	.100	.118	.099	.120	.127	.160	.101	.122	.131	.173
FB	.079	.065	.067	.066	.082	.066	.071	.075	.083	.067	.071	.075
KSA_1	.074	.058	.115	.261	.074	.063	.070	.120	.074	.062	.071	.119
KSA_2	.055	.134	.226	.474	.063	.143	.282	.573	.065	.149	.288	.579
KSA_3	.075	.121	.149	.230	.078	.124	.165	.253	.077	.123	.166	.255
KSC_1	.053	.081	.060	.101	.051	.080	.134	.317	.053	.081	.136	.324
KSC_2	.055	.134	.226	.474	.063	.143	.282	.573	.065	.149	.288	.579
KSC_3	.072	.096	.147	.230	.073	.103	.167	.253	.073	.104	.167	.255
KST_1	.133	.295	.405	.383	.139	.366	.590	.930	.048	.053	.053	.053
KST_2	.133	.295	.405	.383	.139	.366	.590	.930	.048	.053	.053	.053
KST_3	.133	.295	.405	.383	.139	.366	.590	.930	.048	.053	.053	.053
$\hat{\phi}_1$	.083	.053	.047	.039	.086	.053	.050	.034	.088	.053	.051	.033
$\hat{\phi}_2$	.099	.109	.157	.496	.114	.122	.181	.559	.114	.123	.187	.569

Notes – See Table 1.

TABLE 4: Power, GF tests on α

	$\alpha=.5;\beta=0$				$\alpha=1;\beta=0$				$\alpha=1.25;\beta=0$			
n	25	100	250	1000	25	100	250	1000	25	100	250	1000
SW	.080	.174	.204	.235	.146	.430	.476	.552	.045	.287	.405	.498
FB	.281	.671	.740	.760	.003	.003	.072	.209	.010	.004	.005	.007
KSA_1	.016	.996	1.0	1.0	.062	.450	.954	1.0	.117	.133	.351	.922
KSA_2	.001	.013	.044	.104	.001	.001	.010	.038	.004	0	.002	.015
KSA_3	.707	1.0	1.0	1.0	.244	.901	.989	1.0	.138	.591	.874	.986
KSC_1	0	.596	.995	1.0	0	.416	.829	1.0	.060	.208	.598	.982
KSC_2	0	0	.005	.102	0	0	.005	.035	.004	0	.002	.015
KSC_3	.157	.999	1.0	1.0	.157	.903	.976	1.0	.002	.629	.843	.986
KST_1	0	0	0	0	.001	0	0	0	.005	.002	.002	0
KST_2	0	0	0	0	.001	0	0	0	.005	.002	.002	0
KST_3	0	0	0	0	.001	0	0	0	.005	.002	.002	0
$\hat{\phi}_1$	.942	.016	.703	1.0	.129	.282	1	1.0	.025	.151	.911	1.0
$\hat{\phi}_2$	0	0	0	0	.005	0	0	0	.008	.0	.001	0
	$\alpha=1.5;\beta=0$				$\alpha=1.9;\beta=0$				$\alpha=1.99.0;\beta=0$			
n	25	100	250	1000	25	100	250	1000	25	100	250	1000
SW	.023	.022	.067	.103	.145	.162	.151	.156	.318	.686	.872	.957
FB	.016	.010	.013	.012	.148	.166	.158	.160	.328	.700	.877	.957
KSA_1	.055	.053	.092	.207	.053	.043	.052	.064	.047	.039	.056	.078
KSA_2	.011	.006	.003	.003	.065	.186	.384	.764	.082	.295	.711	.995
KSA_3	.079	.244	.415	.665	.039	.016	.006	.003	.032	.011	0	0
KSC_1	.017	.048	.144	.705	.053	.091	.139	.224	.064	.107	.171	.400
KSC_2	.011	.006	.003	.003	.065	.186	.385	.764	.082	.295	.711	.995
KSC_3	.080	.245	.411	.665	.040	.014	.006	.003	.033	.006	0	0
KST_1	.013	.006	.006	.004	.122	.163	.176	.171	.230	.543	.740	.842
KST_2	.013	.006	.006	.004	.122	.163	.176	.171	.230	.543	.740	.842
KST_3	.013	.006	.006	.004	.122	.163	.176	.171	.230	.543	.740	.842
$\hat{\phi}_1$	.013	.036	.219	.892	.124	.139	.168	.330	.212	.193	.281	.621
$\hat{\phi}_2$	.017	.005	.012	.006	.108	.093	.087	.075	.167	.107	.095	.091

Notes – See Table 1.

TABLE 5: Power, GF tests on α - continued

	$\alpha = .5; \beta = .5$				$\alpha = 1; \beta = .5$				$\alpha = 1.25; \beta = .5$			
n	25	100	250	1000	25	100	250	1000	25	100	250	1000
SW	.029	.114	.142	.131	.075	.240	.350	.372	.040	.170	.283	.343
FB	.001	0	.001	0	.009	.006	.013	.043	.011	.008	.014	.024
KSA_1	.109	.729	.972	1.0	.136	.525	.940	1.0	.088	.214	.546	.971
KSA_2	.167	.495	.539	.315	.025	.004	.014	.040	.019	.006	.005	.013
KSA_3	.940	1.0	1.0	1.0	.513	.923	.987	1.0	.263	.656	.864	.984
KSC_1	.001	.263	.952	1.0	.015	.255	.717	.987	.029	.149	.465	.976
KSC_2	0	0	.005	.099	.008	.0	.006	.034	.018	.005	.003	.013
KSC_3	.206	.999	1.0	1.0	.175	844	.960	1.0	.142	.599	.819	.984
KST_1	0	0	0	0	.001	0	0	0	.005	.003	.002	0
KST_2	0	0	0	0	.001	0	0	0	.005	.003	.002	0
KST_3	0	0	0	0	.001	0	0	0	.005	.003	.002	0
$\hat{\phi}_1$	.892	.008	.311	1.0	.089	.135	.971	1.0	.016	.083	.705	1.0
$\hat{\phi}_2$	0	.096	.998	1.0	.005	.161	.909	1.0	.009	.106	.631	1.0
	$\alpha = 1.5; \beta = .5$				$\alpha = 1.9; \beta = .5$				$\alpha = 1.99; \beta = .5$			
n	25	100	250	1000	25	100	250	1000	25	100	250	1000
SW	.027	.051	.099	.143	.147	.161	.158	.148	.334	.703	.880	.952
FB	.018	.012	.018	.013	.151	.162	.164	.152	.341	.717	.878	.943
KSA_1	.070	.078	.116	.379	.045	.038	.069	.102	.044	.042	.099	.233
KSA_2	.028	.020	.006	.006	.057	.150	.327	.735	.071	.279	.761	.999
KSA_3	.126	.292	.426	.673	.028	.007	.004	.003	.021	.003	.002	.001
KSC_1	.033	.070	.178	.667	.048	.074	.143	.336	.052	.117	.278	.714
KSC_2	.029	.018	.006	.006	.057	.150	.327	.735	.071	.279	.761	.999
KSC_3	.096	.262	.416	.673	.032	.015	.002	.003	.023	.014	.002	.001
KST_1	.016	.007	.008	.004	.100	.154	.168	.158	.134	.361	.572	.808
KST_2	.016	.007	.008	.004	.100	.154	.168	.158	.134	.361	.572	.808
KST_3	.016	.007	.008	.004	.100	.154	.168	.158	.134	.361	.572	.808
$\hat{\phi}_1$	.011	.033	.144	.824	.113	.126	.171	.352	.162	.162	.277	.614
$\hat{\phi}_2$	.025	.059	.203	.835	.117	.117	.176	.415	.163	.181	.349	.793

Notes – See Table 1.

All empirical sizes conform to the nominal level of 5%. Of course, this is expected because the procedures are provably size correct. With regards to power, our results do not reveal a uniformly dominant criterion. We thus analyze power ranking within each test class considered as well as from a global perspective.

Tests based on order statistics. The SW test outperforms the FB criterion throughout except in one noteworthy case: the FB test dominates for inference on α when $\alpha = .5$; see Table 4. Since moments do not exist for this case, discrepancies between observed and calibrated medians of order statistics make more sense than the distance between their observed and calibrated means. Both the SW and FB statistics are dominated by the other criteria we introduce except with very small sample sizes ($n = 25$) or when the tested distribution is close to Gaussian: see the $\alpha = 1.99$ and $\beta = 0$ case in Table 4.

Kolmogorov-Smirnov-type tests. Focusing on the tail improves, for inference on β, the power of the unweighted Kolmogorov type EDF statistic, as may be seen from comparing the performance of KSA_1 relative to KSC_1. In contrast, focusing on the tail costs power for inference on α unless α exceeds 1.25. The weighted statistics KSC_2 or KSC_3 may or may not outperform KSA_2 and KSA_3, so focus on the tail does not warrant power improvements for such statistics. On balance, we find that weighing may be preferable to truncation, although a uniformly dominant weighting scheme did not emerge.

The procedure considered to approximate the tail distribution, *i.e.*, whether by simulation or via an asymptotic argument, has important implications for test power, as may be seen from comparing the performance of KST_i, relative to KSC_i, $i = 1, 2, 3$. For inference on β, such effects vary with α. In particular, simulation outperforms asymptotics for $\alpha = 1.5$, whereas asymptotics seems preferable with $\alpha = 1.9$ as long as $\beta < 1$. However, power drops sharply even with a sample size of 1000 for $\beta = 1$, which reflects the inadequacy of the considered approximation for this case. Such a severe discontinuity illustrates the advantages of our proposed two-stage exact procedures for approximating the tail on which the statistic restricts focus as well as the statistic's p-value.[4] Results for inference on α in tables 4 - 5 reinforce this conclusion: the KST_i criteria perform poorly and are almost degenerate [have zero empirical rejections] in many cases.

Quantile-based tests. For inference on β, despite being dominated in the above discussed counter examples, $\hat{\phi}_2$ performs steadily well whereas $\hat{\phi}_1$ has low power. For tests on α and in sharp contrast with $\hat{\phi}_1$ which performs quite well, the power of $\hat{\phi}_2$ is low with $\beta = 0$ [see Table 4] yet it picks up remarkably well for the considered asymmetric case [see Table 5], enough to outperform $\hat{\phi}_1$ in a number of cases.[5] This result is worth noting since $\hat{\phi}_2$ was originally designed to focus on β: in contrast, we find that unless the tested distribution is symmetric, $\hat{\phi}_2$ holds useful information on α as well.

[4]Recall that the tests we construct are exact in terms of size control using both approximation methods, so power discrepancies can soundly be analyzed.

[5]The reported value of β_s in both Tables 4-5 is maintained under the null and alternative hypothesis (while of course α_s varies).

General observations. Several important conclusions can be drawn from tables 2 - 5 interpreted collectively. In contrast to conventional wisdom, quality inference with n as small as 25 is feasible. Examples include testing a value of $\beta > .9$ with $\alpha = 1.5$ for which power with *e.g.* $\hat{\phi}_2$ ranges from around 38-50%, or testing a value of $\alpha = .5$ regardless of the considered β for which power using again $\hat{\phi}_2$ as an example, ranges from 89-94%, which is remarkable with just 25 observations.

If kurtosis is low, all statistics have limited power to detect low-to-medium skewness except for $n = 1000$ where we observe some power. In parallel, kurtosis is harder to detect with symmetric distributions for all sample sizes. These results illustrate the non-separability of inference on α and β and provide further motivation for the joint inference approach we follow in this paper. The fact that $\hat{\phi}_2$ provides information on both α and a non-zero β further supports joint inference.

While we do not expect to pin down a uniformly most powerful criterion, we found that power ranking differ sizeably within and between alternatives. Given their somewhat steady performance, one may recommend the quantile-based criteria. These are however dominated by one of the EDF-based criterion in a number of cases, which suggests that focusing on specific quantiles is not without cost. Then again, aside from ruling out asymptotic-based tail approximations, we do not find grounds for recommending one EDF criterion over another. Recall that variations in α and β entail important differences in the shape of distributions which, for EDF-based statistics, may explain disparities in power ranking across the parameter space. On balance, results suggest combining various statistics. To illustrate the usefulness of such an approach, our empirical analysis implements the combined statistic $\tilde{\phi}(\alpha_0, \beta_0)$ as defined in 3.25. It is worth noting that any set of statistics, and not just $\hat{\phi}_1$ and $\hat{\phi}_2$ can be combined in the same way.

To conclude, we note that we have experimented with an alternative data standardization using the sample mean and standard deviation for inference on β. We find that using empirical means and variances (except of course in the case of $\hat{\phi}_1$ and $\hat{\phi}_2$) cost serious power losses, even with very large sample sizes, particularly for the EDF statistics based on a simulation-based approximation of stable distribution. For instance, with a sample size of $n = 100$ and for $\alpha = 1.5$ and $\beta = .7$, empirical rejections with KSC_1, KSC_2 and KSC_3 are 1.8, 8.2 and 5.0%; with $\beta = .9$, empirical rejections for these statistics are 1.4%, 9.0% and and 4.4%; power does not improve for these statistics when the sample increases to 250 observations.

We have also considered an alternative choice for $\bar{\lambda}_1$ and $\bar{\lambda}_2$, namely we set the 10th and 90th percentiles of the simulated samples $y_1(\alpha_0, \beta_0), \ \ldots \ , y_{\bar{n}}(\alpha_0, \beta_0)$, whereas reported results pertain to the 5th and 95th percentiles. Test powers are affected although not importantly, and no choice uniformly dominates for the cases analyzed.

5. Application to electricity prices

To illustrate the usefulness of the proposed procedures, and in particular the non-separable nature of the inference problem, we apply our set estimation method to electricity prices. Electricity prices have been regulated up to the beginning of the year 2000. In many countries, the trend since then has been to let the electricity market clear on its own. In that context, electricity prices have become very volatile which can be attributed to the fact that electricity is a non-storable commodity and to the characteristics of its market. The demand side is very inelastic while its supply side is affected by location of generators, their market concentration as well as the transmission structure. Given the importance of electricity in the commodity market [Bessembinder and Lemmon (2006)] and the increased risk for those who need to position themselves in that market leading to an increased use of derivatives, there have been a lot of developments in the literature on the modeling of electricity prices.

Indeed, models of electricity prices include precise features in order to find the best match for the empirical distribution. They include mean reversion, time of day and week day effects, seasonal effects, time-varying volatility and volatility clustering and extreme values. Yet it appears that despite modeling such features, normality or log-normality do not represent the data accurately in their inability to capture very large changes in prices [Knittel and R. (2005)]. Here we propose to analyze the fit of stable distributions on a standard series of electricity prices, and to obtain confidence set estimates for the associated skewness and tail index parameters. The spot prices under consideration were extracted from Reuters. They are the on-peak electricity spot price expressed in US dollars per megawatt hour. They were initially provided by ICAP US. The daily data series denoted P_t, $t = 1, \ldots T,$ starts from January 3, 2001 to May 15, 2006; the sample size is 1399 observations. We analyze the associated return series, *i.e.* $\ln(P_t) - \ln(P_{t-1})$.

We derive joint confidence regions for both skewness and tail index parameters. As explained in section 3, each confidence set is obtained by collecting all pairs of (α, β) values which are not rejected by each test applied. A grid search is applied over the range $0 < \alpha \leq 2$ and $-1 \leq \beta \leq 1$, and 95% level confidence sets are constructed by retaining the pairs of (α, β) for which (in turn) each test p-value [calculated using the MC test method as shown above] is greater than 5%. It is important to ensure that the same random draws at all stages of the MC procedures are maintained for each pair of values tested, so each test applied will depend on the same random variates throughout, and the sequence of tests applied thus differ only via the pair (α, β) values under test. We use $N = 999$ and $N_0 = \bar{n} = 2000$. We set $\bar{\lambda}_1$ and $\bar{\lambda}_2$ at the 5th and 95th percentiles of the underlying simulated sample $y_1(\alpha_0, \beta_0), \ldots, y_{\bar{n}}(\alpha_0, \beta_0)$.

Results are reported in graphical form, where we plot the regions associated with the non-rejected pairs for each test inverted at the 5% level. The grid search we implemented used a step of .05 for both parameters. Figures 1-12 report β as a function of α for all non-rejected (at the 5% level) pairs. Confidence intervals for each parameters conveniently obtain from the latter joint region set by projection.

As may be checked from Figures 1-12, the 95% confidence sets obtained differ dramatically depending on the tests inverted. This result is in line with our power study. Several statistics are quite uninformative particularly regarding the skewness coefficient. Nevertheless, three statistics lead to very concise set estimates, namely KSA_1, KSC_1 and the test which combines $\hat{\phi}_1$ and $\hat{\phi}_2$. Confidence sets based on these tests lead to the following projection-based intervals, respectively: [1.35,1.7] for α and [0,0.58] for β, using KSA_1; [1.4,1.6] for α and [0.2,0.5] for β, using KSC_1; and [1.3,1.5] for α and [0.18,0.62] for β, using the combined $\hat{\phi}_1$-$\hat{\phi}_2$ test. These results suggest that heavy kurtosis and asymmetries are evident in the series analyzed.

For comparison, we ran the tests on the price series rather than returns. Interestingly, in this case, the tests have lead to dramatically different confidence sets (at the 95% level). In particular, the sets based on KSA_2 and KSC_2 are completely uninformative on both parameters (the confidence sets practically covers the full parameter space), whereas the sets based on KSA_1 and KSC_1 are empty, leading to reject the family of stable distributions. The Bonferroni test which combines $\hat{\phi}_1$ and $\hat{\phi}_2$ leads to an interval which covers (and is quite concentrated around) zero for β, whereas the associated interval for α does not differ importantly from the intervals obtained using returns.

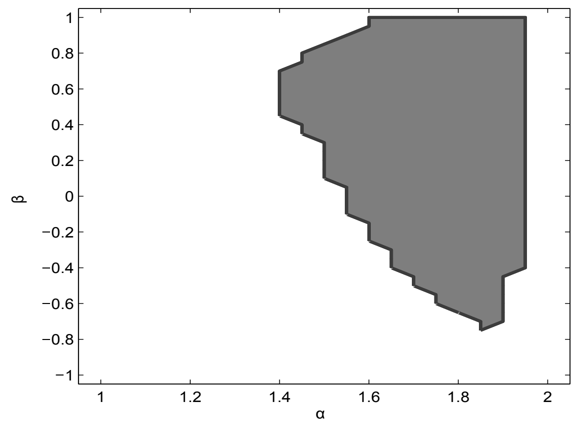

Figure 1. Joint 95% confidence set for (α, β) based on *SW*

6. Conclusion

In this paper, we have proposed a class of exact procedures for testing goodness-of-fit of the stable distribution in location-scale models. Our procedure extends usual GF tests as well as the quantile based criteria proposed by McCulloch (1986). The statistics null distributions are analytically intractable, so the tests are implemented using Monte Carlo test methods. By inverting these test statistics, we solve the problem of estimating the skewness and tail parameters. The properties of our proposed procedures were illustrated via a simulation study and an empirical application on electricity prices.

Our approach clearly has widespread applications beyond the specific class of distributions considered, and provides some insight into the type of simulation-based GF testing that we are likely to see much more of in the future.

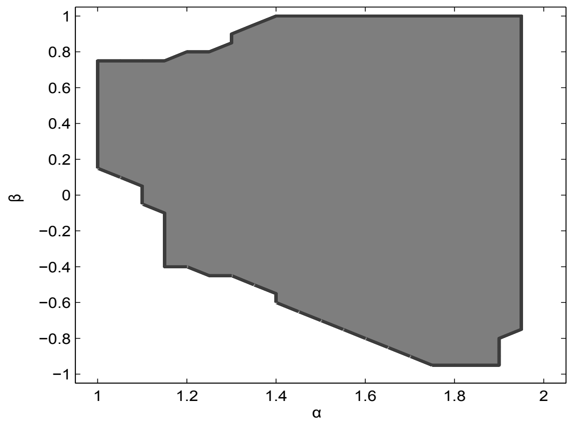

Figure 2. Joint 95% confidence set for (α, β) based on FB

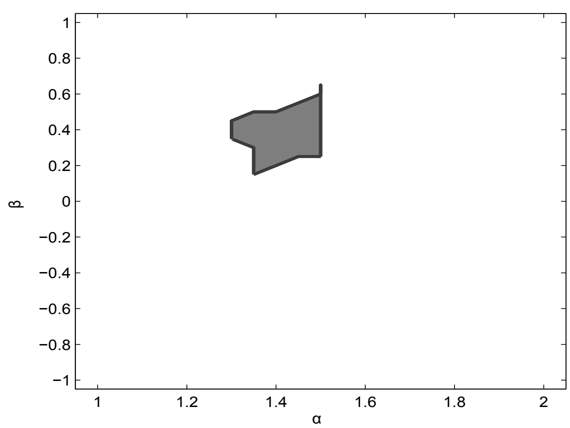

Figure 3. Joint 95% confidence set for (α, β) based on combining $\hat{\phi}_1$ and $\hat{\phi}_2$ using $\tilde{\phi}$

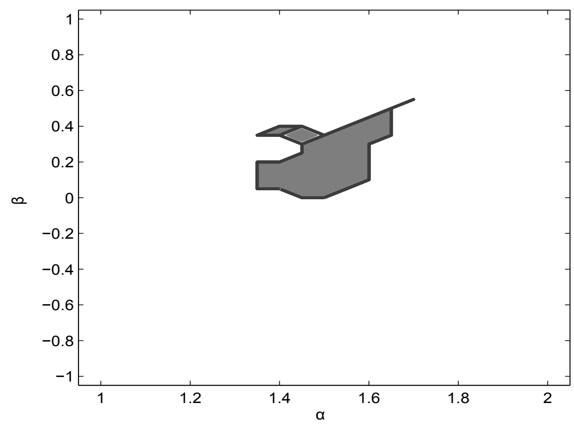

Figure 4. Joint 95% confidence set for (α, β) based on KSA_1

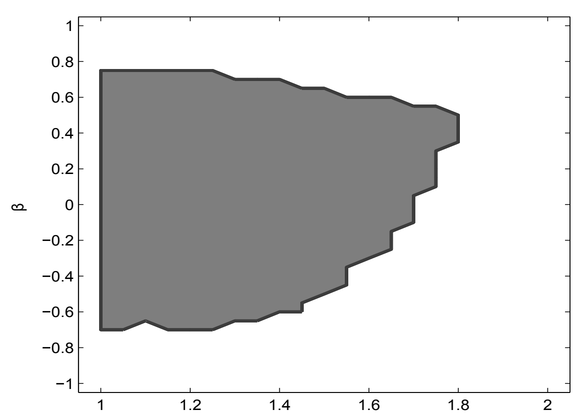

Figure 5. Joint 95% confidence set for (α, β) based on KSA_2

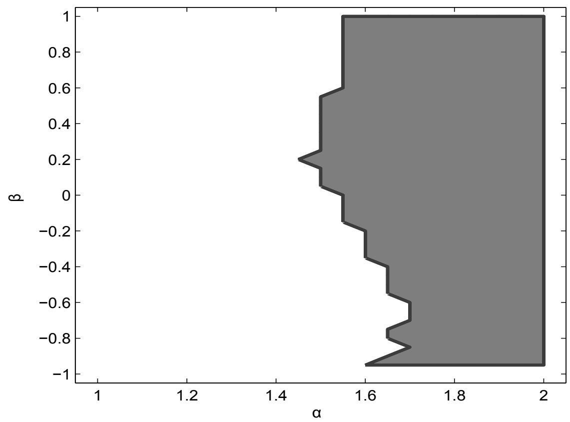

Figure 6. Joint 95% confidence set for (α, β) based on KSA_3

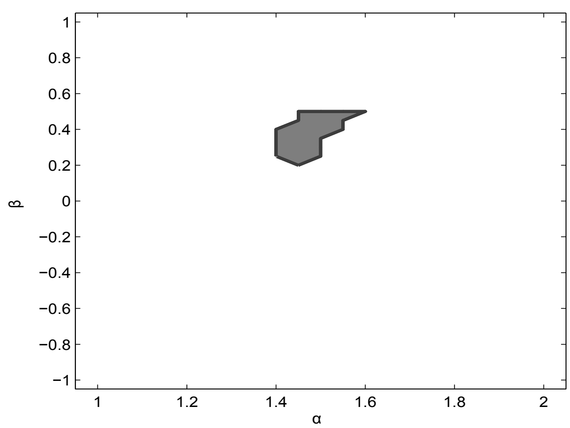

Figure 7. Joint 95% Confidence Set based on KSC_1

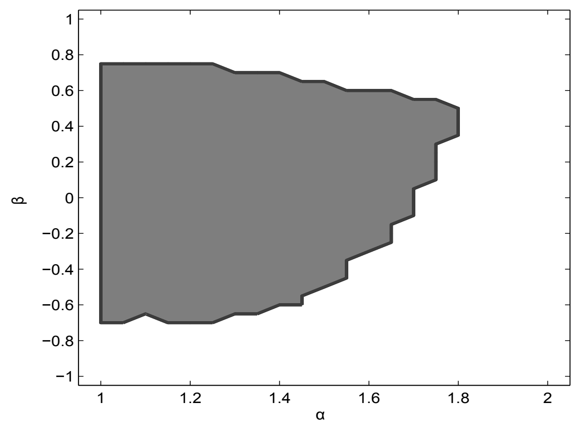

Figure 8. Joint 95% confidence set for (α, β) based on KSC_2

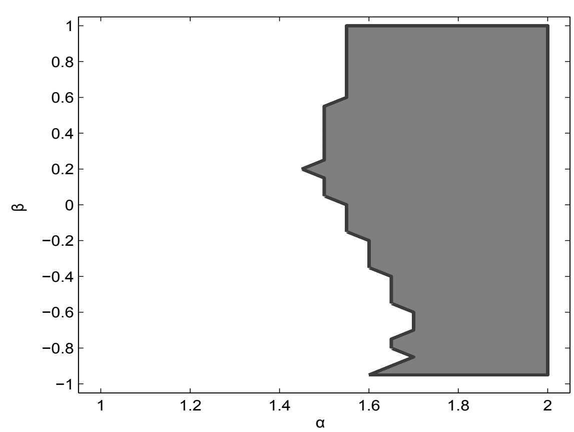

Figure 9. Joint 95% confidence set for (α, β) based on KSC_3

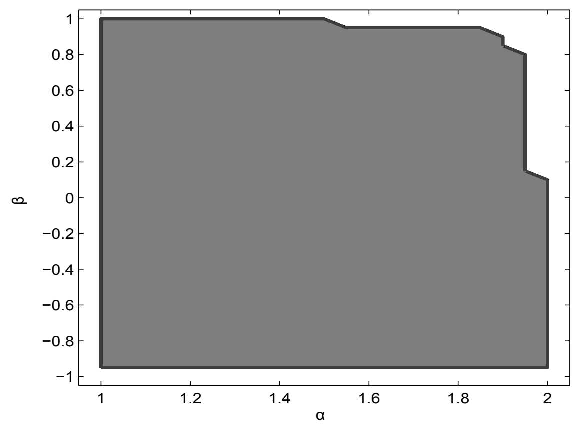

Figure 10. Joint 95% confidence set for (α, β) based on KST_1

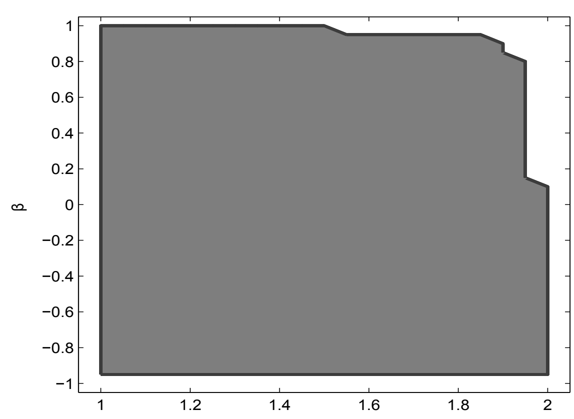

Figure 11. Joint 95% confidence set for (α, β) based on KST_2

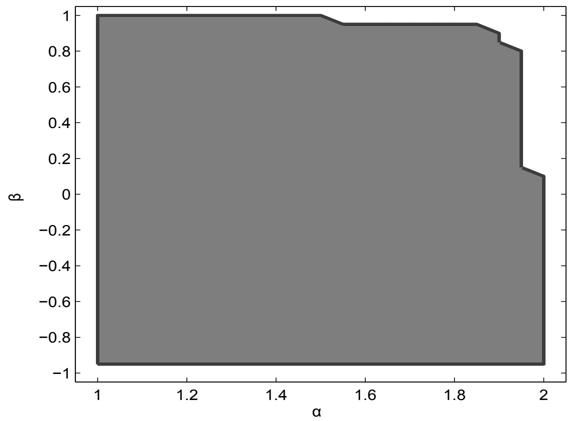

Figure 12. Joint 95% confidence set for (α, β) based on KST_3

References

Abdelkhalek, T. and Dufour, J.-M. (1998), 'Statistical inference for computable general equilibrium models, with application to a model of the Moroccan economy', *Review of Economics and Statistics* **LXXX**, 520–534.

Barnard, G. A. (1963), 'Comment on 'The spectral analysis of point processes' by M. S. Bartlett', *Journal of the Royal Statistical Society, Series B* **25**, 294.

Beaulieu, M.-C., Dufour, J.-M. and Khalaf, L. (2005), Exact multivariate tests of asset pricing models with stable asymmetric distributions, *in* M. Breton and H. Ben Ameur, eds, 'Numerical Methods in Finance', Kluwer, Dordrecht, The Netherlands.

Beaulieu, M.-C., Dufour, J.-M. and Khalaf, L. (2007), 'Multivariate tests of mean-variance efficiency with possibly non-Gaussian errors: An exact simulation-based approach', *Journal of Business and Economic Statistics* **25**, 398–410.

Beaulieu, M.-C., Dufour, J.-M. and Khalaf, L. (2013), 'Identification-robust estimation and testing of the Zero-Beta CAPM', *The Review of Economics Studies* **80**, 892–924.

Bessembinder, H. and Lemmon, M. L. (2006), 'Gains from trade under uncertainty: The case of electric power markets', *Journal of Business* **79**, 1755–1782.

Bolduc, D., Khalaf, L. and Yelou, C. (2010), 'Identification robust confidence sets methods for inference on parameter ratios with application to discrete choice models', *Journal of Econometrics* **157**, 317–327.

Chambers, J. M., Mallows, C. L. and Stuck, B. W. (1976), 'A method for simulating stable random variables', *Journal of the American Statistical Association* **71**, 340–344. Correction, 82, 704, 83, 581.

Dufour, J.-M. (1989), 'Nonlinear hypotheses, inequality restrictions, and non-nested hypotheses: Exact simultaneous tests in linear regressions', *Econometrica* **57**, 335–355.

Dufour, J.-M. (2006), 'Monte Carlo tests with nuisance parameters: A general approach to finite-sample inference and nonstandard asymptotics in econometrics', *Journal of Econometrics* **133**, 443–478.

Dufour, J.-M., Farhat, A., Gardiol, L. and Khalaf, L. (1998), 'Simulation-based finite sample normality tests in linear regressions', *The Econometrics Journal* **1**, 154–173.

Dufour, J.-M. and Khalaf, L. (2001), Monte Carlo test methods in econometrics, *in* B. Baltagi, ed., 'Companion to Theoretical Econometrics', Blackwell Companions to Contemporary Economics, Basil Blackwell, Oxford, U.K., chapter 23, pp. 494–519.

Dufour, J.-M. and Khalaf, L. (2002*a*), 'Exact tests for contemporaneous correlation of disturbances in seemingly unrelated regressions', *Journal of Econometrics* **106**, 143–170.

Dufour, J.-M. and Khalaf, L. (2002*b*), 'Simulation based finite and large sample tests in multivariate regressions', *Journal of Econometrics* **111**, 303–322.

Dufour, J.-M. and Khalaf, L. (2003), Finite sample tests in seemingly unrelated regressions, *in* D. E. A. Giles, ed., 'Computer-Aided Econometrics', Marcel Dekker, New York, chapter 2, pp. 11–35.

Dufour, J.-M., Khalaf, L. and Beaulieu, M.-C. (2003), 'Exact skewness-kurtosis tests for multivariate normality and goodness-of-fit in multivariate regressions with application to asset pricing models', *Oxford Bulletin of Economics and Statistics* **65**, 891–906.

Dufour, J.-M., Khalaf, L. and Beaulieu, M.-C. (2010), 'Multivariate residual-based finite-sample tests for serial dependence and GARCH with applications to asset pricing models', *Journal of Applied Econometrics* **25**(2), 263–285.

Dufour, J.-M., Khalaf, L., Bernard, J.-T. and Genest, I. (2004), 'Simulation-based finite-sample tests for heteroskedasticity and ARCH effects', *Journal of Econometrics* **122**(2), 317–347.

Dufour, J.-M. and Kiviet, J. F. (1996), 'Exact tests for structural change in first-order dynamic models', *Journal of Econometrics* **70**, 39–68.

Dufour, J.-M. and Kiviet, J. F. (1998), 'Exact inference methods for first-order autoregressive distributed lag models', *Econometrica* **66**, 79–104.

Dufour, J.-M., Kurz-Kim, J. and Palm, F., eds (2010), *Heavy Tails and Paretian Distributions in Empirical Finance. A Volume Honoring Benoît Mandelbrot*, Vol. 17 (No. 2) of *Journal of Empirical Finance (Special issue)*.

Dufour, J.-M. and Kurz-Kim, J.-R. (2010), ‘Exact inference and optimal invariant estimation for the stability parameter of symmetric stable distributions’, *Journal of Empirical Finance* **17**, 180–194.

Dwass, M. (1957), ‘Modified randomization tests for nonparametric hypotheses’, *Annals of Mathematical Statistics* **28**, 181–187.

Embrechts, P., Klüppelberg, C. and Mikosch, T. (1997), *Modelling Extremal Events for Insurance and Finance*, Springer-Verlag, Berlin.

Filliben, J. J. (1975), ‘The probability plot correlation coefficient test for normality’, *Technometrics* **17**, 111–117.

Ibragimov, I. A. and Linnik, Y. V. (1975), *Independent and Stationary Sequences of Random Variables*, Wolters-Nordhoff, Groningen, The Netherlands.

Kiviet, J. F. and Dufour, J.-M. (1997), ‘Exact tests in single equation autoregressive distributed lag models’, *Journal of Econometrics* **80**, 325–353.

Knittel, C. R. and R., R. M. (2005), ‘An empirical examination of restructured electricity prices’, *Energy Economics* **27**, 791–817.

Kolmogorov, A. N. (1933), ‘Sulla determinazione empirica di una legge di distribizione’, *Giornale dell’ Instituto Italiano degli Atturai* **4**, 83–91.

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

McCulloch, J. H. (1986), ‘Simple consistent estimators of stable distribution parameters’, *Communications in Statistics, Simulation and Computation* **15**, 1109–1136.

McCulloch, J. H. (1996), ‘On the parametrization of the afocal stable distributions’, *Bulletin of the London Mathematical Society* **28**, 651–655.

Miller, Jr., R. G. (1981), *Simultaneous Statistical Inference*, second edn, Springer-Verlag, New York.

Rachev, S., Kim, J.-R. and Mittnik, S. (1999*a*), ‘Stable paretian econometrics, part I’, *The Mathematical Scientist* **24**, 24–55.

Rachev, S., Kim, J.-R. and Mittnik, S. (1999*b*), ‘Stable paretian econometrics, part II’, *The Mathematical Scientist* **24**, 113–127.

Rachev, S. and Mittnik, S. (2000), *Stable Paretian Models in Finance*, John Wiley & Sons, New York.

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

Samuelson, P. (1967), 'On multivariate tests of the Capital Asset Pricing Model', *Journal of Financial and Quantitative Analysis* **2**, 107–122.

Shapiro, S. and Wilk, M. (1965), 'An analysis of variance tests for normality (complete samples)', *Biometrika* **52**, 591–611.

Smirnov, N. V. (1939), 'Sur les écarts de la courbe de distribution empirique (Russian/French summary)', *Matematičeskiĭ Sbornik N.S.* **6**, 3–26.

Thode, Jr., H. C. (2002), *Testing for Normality*, number 164 *in* 'Statistics: Textbooks and Monographs', Marcel Dekker, New York.

Weron, R. (1996), 'On the Chambers-Mallows-Stuck method for simulating skewed stable random variables', *Statistics and Probability Letters* **28**, 165–171.

Zolotarev, V. M. (1986), *One-Dimensional Stable Distributions*, Vol. 65 of *Translations of Mathematical Monographs*, American Mathematical Society, Providence, Rhose Island.