Industrial and Corporate Change, Volume 20, Number 4, pp. 991–1030
doi:10.1093/icc/dtr036
Advance Access published July 4, 2011

A new class of asymmetric exponential power densities with applications to economics and finance

Giulio Bottazzi*,† and Angelo Secchi**

We introduce a new five-parameter family of distributions, the asymmetric exponential power (AEP), able to cope with asymmetries and leptokurtosis and, at the same time, allowing for a continuous variation from non-normality to normality. We prove that the maximum likelihood (ML) estimates of the AEP parameters are consistent on the whole parameter space, and when sufficiently large values of the shape parameters are considered, they are also asymptotically efficient and normal. We derive the Fisher information matrix for the AEP and we show that it can be continuously extended also to the region of small shape parameters. Through numerical simulations, we find that this extension can be used to obtain a reliable value for the errors associated to ML estimates also for samples of relatively small size (100 observations). Moreover, we show that around this sample size, the bias associated with ML estimates, although present, becomes negligible. Finally, we present a few empirical investigations, using diverse data from economics and finance, to compare the performance of AEP with respect to other, commonly used, families of distributions.

JEL classification: C16, C46.

1. Introduction

A large and increasing number of empirical analyses in a variety of fields suggest that the assumption of normality of real data is quite often not tenable. Indeed, empirical densities characterized by heavy tails as well as by significant degree of

*Giulio Bottazzi, Scuola Superiore S.Anna, P.za Martiri della Liberta' 33, 56127 Pisa, Italy. e-mail: bottazzi@sssup.it

**Angelo Secchi, University of Pisa, Pisa, Italy; Universitè Paris 1 Panthèon-Sorbonne, Paris, France. e-mail: angelo.secchi@univ-paris1.fr

†Main author for correspondence.

asymmetry are often observed in many economic domains. In finance, since the seminal work of Mandelbrot, scholars and practitioners have become aware that the volatile dynamics that traditionally characterize financial markets cannot be properly described by using the Gaussian distribution; quite the contrary, almost every financial return series has been found to be characterized by the presence of fat tails (cf. the reviews in Mantegna and Stanley, 2000; McCauley, 2007, and the references therein). A number of recent studies have brought strong empirical support to the claim that fat tails are also a robust property of aggregate output growth rates distributions, both in cross-sections of different countries (Canning *et al.*, 1998; Castaldi and Dosi, 2009) and in within-country time series (Fagiolo *et al.*, 2008). At the micro-economic level, strong leptokurtosis has been identified in business companies growth rates in many developed countries, irrespectively of the proxy used to measure firm size and of the level of disaggregation considered (Stanley *et al.*, 1996; Bottazzi and Secchi, 2003, 2006a,b; Bottazzi *et al.*, 2007).

In all these domains, it is important to adopt flexible statistical models able to cope directly with skewness and leptokurtosis and, at the same time, to allow continuous variation from non-normality to normality (Huber, 1981; Azzalini, 1986; Hampel *et al.*, 1986). Both these aspects are captured by the asymmetric exponential power (AEP) family of distributions discussed in the present paper. As a further specific motivation for introducing it, we present three empirical exercises that show how it actually performs in describing those empirical distributions characterized jointly by significant degrees of skewness and fat tails. We compare the goodness of fit achieved by the AEP with those obtained with other commonly used distributions, namely the skewed exponential power (SEP), the α-Stable family, and the Generalized Hyperbolic (GHYP). Other examples of the successful and general applicability of the AEP are in Santoro (2006), Alfarano and Milakovic (2007), Fagiolo *et al.* (2008), and Sapio (2008).

The article is organized as follows. In the next section, the AEP family of distribution is introduced. In Section 3, we present some theoretical results on the maximum likelihood (ML) estimation of the AEP family and derive the elements of the Fisher's information matrix, discussing its domain of definition. In Section 3.1, we prove the consistency of the estimator in the whole parameter space and we discuss the asymptotic efficiency and normality for the case in which both parameters b_l and b_r are greater than two, while in Section 3.2, we show that, for some estimates, the domain of definition of the information matrix can be extended to the whole parameter space. Next, in Section 4, with the help of extensive numerical simulations, we analyze the bias of the ML estimator and their asymptotic behavior in the domain of the parameters space not covered by the analytical results. Finally, in Section 5 we compare the performance of the AEP with other, commonly adopted, families of distributions in three specific empirical exercises including electricity, foreign exchange, and stock market data.

2. The asymmetric exponential power distribution

Subbotin (1923) introduced a family of distribution, generally known as the exponential power (EP) distribution, characterized by a scale parameter $a > 0$, a shape parameter $b > 0$, and a location parameter m. The EP density reads

$$f_{\mathrm{EP}}(x;\, b, a, m) = \frac{1}{2ab^{1/b}\Gamma(1/b+1)}\; e^{-\frac{1}{b}\left|\frac{x-m}{a}\right|^{b}} \tag{1}$$

where $\Gamma(x)$ is the Gamma function. The Gaussian distribution is recovered when $b = 2$ while when $b < 2$, the distributions are heavy tailed: the lower is the shape parameter b the fatter the density tails. This model has been studied by many scholar: compare among others Box (1953), Turner (1960) and Vianelli (1963). Inferential aspects of the EP distribution inside the ML framework have been analyzed in Agró (1995) and Capobianco (2000). In order to deal with both fat tails and skewness, Azzalini (1986) considered the SEP distribution

$$f_{\mathrm{SEP}}(x\,;\, b, a, m, \lambda) = 2\,\Phi(\mathrm{sign}(z)|z|^{b/2}\,\lambda\,\sqrt{2/b})\, f_{\mathrm{EP}}(x;\, b, a, m) \tag{2}$$

where $z = (x - m)/\mathrm{a}$, $a, b > 0$, $-\infty < m < \infty$, $-\infty < x < \infty$, $-\infty < \lambda < \infty$ and Φ is the normal distribution function. It easy to see that f_{SEP} reduces to f_{EP} when $\lambda = 0$ so that the normal case is obtained when $(\lambda, b) = (0, 2)$. The ML inference problem for this distribution is discussed in details in DiCiccio and Monti (2004).

In the present article, we suggest an alternative way to tackle the presence of heavy tails and skewness. We propose a new five-parameters family of distributions, the AEP, characterized by two positive shape parameters b_r and b_l, describing the tail behavior in the upper and lower tail, respectively; two positive scale parameters a_r and a_l, associated with the distribution width above and below the modal value and

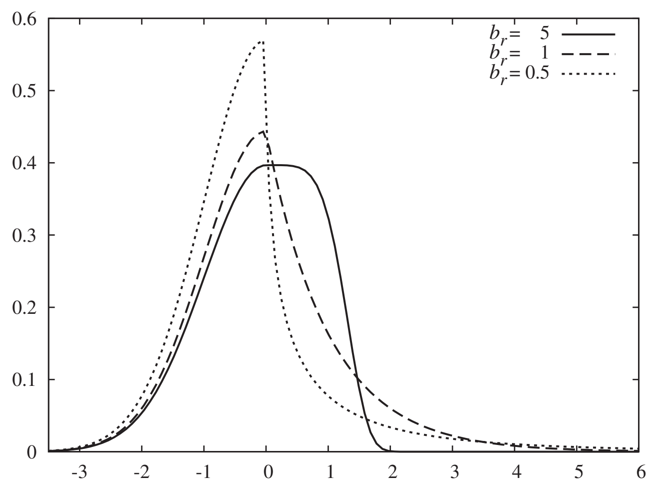

Figure 1 Densities of the AEP(1,2,1,b_r) with $b_r = 5$, $b_r = 1$, and $b_r = 0.5$.

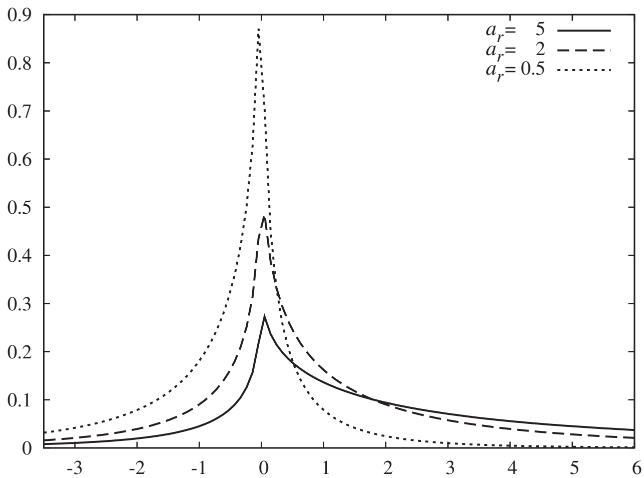

Figure 2 Densities of the AEP(1,0.5,a_r,0.5) with $a_r = 5$, $a_r = 2$, and $a_r = 0.5$.

one location parameter m, representing the mode. The AEP density presents the following functional form

$$f_{\mathrm{AEP}}(x;\mathbf{p}) = \frac{1}{C}\,\mathrm{e}^{-\left(\frac{1}{b_l}\left|\frac{x-m}{a_l}\right|^{b_l}\theta(m-x)+\frac{1}{b_r}\left|\frac{x-m}{a_r}\right|^{b_r}\theta(x-m)\right)} \tag{3}$$

where $\mathbf{p}=(b_l, b_r, a_l, a_r, m)$, $\theta(x)$ is the Heaviside theta function and where the normalization constant reads $C = a_l A_0(b_l) + a_r A_0(b_r)$ with

$$A_k(x) = x^{\frac{k+1}{x}-1}\Gamma\left(\frac{k+1}{x}\right). \tag{4}$$

Figure 1 and Figure 2 report AEP densities obtained by assuming different values for the right shape and width parameters, *br* and *ar*.

The AEP reduces to the EP when $a_l = a_r$ and $b_l = b_r$. The density in (3) can be easily integrated to obtain the distribution function

$$\begin{aligned} F_{\mathrm{AEP}}(x;\mathbf{p}) = {} & \frac{a_l A_0(b_l)}{C}\, Q\left(\frac{1}{b_l}, \left|\frac{x-m}{a_l}\right|^{b_l}\right)\theta(m-x) \\ & + \left(1 - \frac{a_r A_0(b_r)}{C}\, Q\left(\frac{1}{b_r}, \left|\frac{x-m}{a_r}\right|^{b_r}\right)\right)\theta(x-m), \end{aligned} \tag{5}$$

where $Q(\alpha, x)$ is the regularized upper incomplete gamma function $Q(\alpha, x) = \Gamma(\alpha, x)/\Gamma(\alpha)$.

The mean μ_{AEP} and the variance σ^2_{AEP} of the AEP distribution can be straightforwardly derived

$$\mu_{\mathrm{AEP}} = m + \frac{1}{C}\left(a_r^2 A_1(b_r) - a_l^2 A_1(b_l)\right) \quad \sigma^2_{\mathrm{AEP}} = \frac{a_r^3}{C} A_2(b_r) + \frac{a_l^3}{C} A_2(b_l). \tag{6}$$

Moreover, it is possible to express the generic h-th central moment M_h as a finite series reading

$$M_h = \sum_{q=0}^{h} \binom{h}{q} \frac{1}{C^{h-q+1}} \left(a_r^{q+1} A_h(b_r) + a_l^{q+1} A_h(b_l) \right) \left(a_r^2 A_1(b_r) - a_l^2 A_1(b_l) \right)^{h-q}. \tag{7}$$

The AEP constitutes a natural extension of the family originally proposed by Subbotin, hence the results derived in the present article apply also to the latter.

3. ML estimation

Consider a set of N observations $\{x_1, \ldots, x_N\}$ and assume that they are independently drawn from the AEP distribution with parameters $\mathbf{p_0}$. We are interested in the estimation of $\mathbf{p}$ from that sample. The ML estimate $\hat{\mathbf{p}}$ is obtained maximizing the empirical likelihood or, equivalently, minimizing the negative log-likelihood, computed taking the logarithm of the likelihood function and changing its sign

$$\hat{\mathbf{p}} = \arg\min_{\mathbf{p}} \sum_{i=1}^{N} L_{\mathrm{AEP}}(x_i; \mathbf{p_0}) \quad \text{where} \quad L_{\mathrm{AEP}}(x; \mathbf{p_0}) = -\log f_{\mathrm{AEP}}(x; \mathbf{p_0}). \tag{8}$$

The Cramer–Rao lower bound for the estimates standard error in the case of unbiased estimators is provided by the 5×5 information matrix $J(\mathbf{p_0})$, defined as the expected value of the cross-derivative

$$J_{i,j}(\mathbf{p_0}) = \mathrm{E}_{\mathbf{p}_0}\left[\partial_i L_{\mathrm{AEP}}(x; \mathbf{p_0})\, \partial_j L_{\mathrm{AEP}}(x; \mathbf{p_0}) \right], \tag{9}$$

where $E_{\mathbf{p}_0}[.]$ is the theoretical expectation computed using the true values $\mathbf{p_0}$ and where the indexes i and j runs over the five elements of $\mathbf{p}$, (b_l, b_r, a_l, a_r, m). In practice, one usually assumes $\mathbf{p_0} = \hat{\mathbf{p}}$. In the next sections, we will show that, notwithstanding the presence of finite-sample biases and of analytical problems in extending the definition of J to small values of b_l and b_r, the elements of this matrix can be used to characterize the statistical errors associated to ML estimates on a large part of the parameters space. The expression of the elements of the Fisher information matrix for the AEP distribution are provided in the following

Theorem 3.1 (information matrix of AEP density): The elements of the Fisher information matrix $J(\mathbf{p})$ of the AEP distribution (3) are

$$J_{b_l b_l} = \frac{1}{C} a_l B_0''(b_l) - \frac{1}{C^2} a_l^2 (B_0'(b_l))^2 + \frac{a_l}{C b_l} B_2(b_l) - \frac{2a_l}{C b_l^2} B_1(b_l) + \frac{2a_l}{C b_l^3} B_0(b_l)$$

$$J_{b_l b_r} = -\frac{1}{C^2} a_l a_r B_0'(b_l) B_0'(b_r)$$

$$
\begin{aligned}
J_{b_l a_l} &= \frac{1}{C} B_0'(b_l) - \frac{1}{C^2} a_l B_0(b_l) B_0'(b_l) - \frac{1}{C} B_1(b_l) \\
J_{b_l a_r} &= -\frac{1}{C^2} a_l B_0(b_r) B_0'(b_l) \\
J_{b_l m} &= \frac{1}{b_l C} (\log b_l - \gamma) \\
J_{b_r b_r} &= \frac{1}{C} a_r B_0''(b_r) - \frac{1}{C^2} a_r^2 (B_0'(b_r))^2 + \frac{a_r}{C b_r} B_2(b_r) - \frac{2 a_r}{C b_r^2} B_1(b_r) + \frac{2 a_r}{C b_r^3} B_0(b_r) \\
J_{b_r a_l} &= -\frac{1}{C^2} a_r B_0(b_l) B_0'(b_r) \\
J_{b_r a_r} &= \frac{1}{C} B_0'(b_r) - \frac{1}{C^2} a_r B_0(b_r) B_0'(b_r) - \frac{1}{C} B_1(b_r) \\
J_{b_r m} &= -\frac{1}{b_r C} (\log b_r - \gamma) \\
J_{a_l a_l} &= -\frac{1}{C^2} B_0^2(b_l) + \left(\frac{b_l + 1}{a_l}\right) \frac{1}{C} B_0(b_l) \\
J_{a_l a_r} &= -\frac{1}{C^2} B_0(b_l) B_0(b_r) \\
J_{a_l m} &= -\frac{b_l}{C a_l} \\
J_{a_r a_r} &= -\frac{1}{C^2} B_0^2(b_r) + \left(\frac{b_r + 1}{a_r}\right) \frac{1}{C} B_0(b_r) \\
J_{a_r m} &= \frac{b_r}{C a_r} \\
J_{mm} &= \frac{b_l^{-1/b_l + 1}}{a_l C} \Gamma\left(\frac{2 b_l - 1}{b_l}\right) + \frac{b_r^{-1/b_r + 1}}{a_r C} \Gamma\left(\frac{2 b_r - 1}{b_r}\right)
\end{aligned}
\tag{10}
$$

where γ is the Euler-Mascheroni constant and, for any integer k, it is

$$
B_k(x) = x^{\frac{1}{x} - k} \sum_{h=0}^{k} \binom{k}{h} \log^h x \, \Gamma^{(k-h)}\left(1 + \frac{1}{x}\right), \tag{11}
$$

where $\Gamma^{(k)}$ stands for the k-th derivative of the Gamma function.

Proof: See Appendix A.

In principle, the elements of the inverse information matrix J^{-1} can be directly obtained from the expressions in (10). None of these elements, however, is identically zero, nor any easy simplification can be found. For these reasons, we decided not to report here their cumbersome expressions. In general, for practical purposes, it is much more convenient to compute the elements of J and obtain the elements of J^{-1} by numerical inversion. The situation changes if one considers the original symmetric EP obtained when $a_l = a_r = a$ and $b_l = b_r = b$. For this case, the information matrix has been derived in Agró (1995). To ease the comparison of the general and the particular case, we report the result here using our notation.[1] One has:

Theorem 3.2 (information matrix of EP density): Consider the EP distribution defined in (1) for the set of parameters (b, a, m). The Fisher information matrix $\bar{J}(b, a, m)$ defined as

$$\bar{J}_{i,j}(b, a, m) = \mathrm{E}_{b,a,m}\left[\partial_i L_{\mathrm{EP}}(x;\, b, a, m)\, \partial_j L_{\mathrm{EP}}(x;\, b, a, m)\right], \tag{12}$$

where $L_{\mathrm{EP}}(x; b, a, m) = -\log f_{\mathrm{EP}}(x; b, a, m)$ is found to be

$$\begin{bmatrix} \frac{1}{b^3}\left[\psi(1+1/b)+\log b\right]^2 + \frac{\psi'(1+1/b)}{b^3}\left(1+\frac{1}{b}\right) - \frac{1}{b^3} & -\frac{1}{ab}\left[\log b + \psi\left(1+\frac{1}{b}\right)\right] & 0 \\ -\frac{1}{ab}\left[\log b + \psi\left(1+\frac{1}{b}\right)\right] & \frac{b}{a^2} & 0 \\ 0 & 0 & \frac{b^{-2/b+1}\Gamma(2-1/b)}{a^2\Gamma(1+1/b)} \end{bmatrix} \tag{13}$$

and its inverse reads

$$\begin{bmatrix} \frac{b^4}{-b+(1+b)\psi'\left(1+\frac{1}{b}\right)} & \frac{ab^2\left[\log b+\psi\left(1+\frac{1}{b}\right)\right]}{-b+(1+b)\psi'\left(1+\frac{1}{b}\right)} & 0 \\ \frac{ab^2\left[\log b+\psi\left(1+\frac{1}{b}\right)\right]}{-b+(1+b)\psi'\left(1+\frac{1}{b}\right)} & \frac{a^2\left[b(-1+\log^2 b)+(1+b)\psi'\left(1+\frac{1}{b}\right)+2b\psi\left(1+\frac{1}{b}\right)\log b+b\psi^2\left(1+\frac{1}{b}\right)\right]}{b\left[-b+(1+b)\,\psi'\left(1+\frac{1}{b}\right)\right]} & 0 \\ 0 & 0 & \frac{a^2 b^{2/b-1}\,\Gamma\left(1+\frac{1}{b}\right)}{\Gamma\left(2-\frac{1}{b}\right)} \end{bmatrix} \tag{14}$$

Proof: Since $L_{\mathrm{EP}}(x;\, b, a, m) = L_{\mathrm{AEP}}(x;\, \bar{\mathbf{p}})$ where $\bar{\mathbf{p}} = (b, b, a, a, m)$, the elements of (13) can be easily found starting from the elements of the AEP reported in Theorem 3.1. Consider, for instance, the shape parameter b. The derivative with respect to b of L_{EP} is the sum of the derivatives with respect to b_l and b_r of L_{AEP}. In other terms, in computing the elements of the

[1]Notice that the expansion of the element $\bar{J}^{-1}_{b,a}$ of the inverse information matrix reported in Agró (1995) contains a mistake: the term $[\log b + \psi(1 + \frac{1}{b})]$ in the numerator is incorrectly squared.

Fisher information matrix for the EP distribution, one has to consider the substitution $\frac{\partial}{\partial b} \leftrightarrow \frac{\partial}{\partial b_l} + \frac{\partial}{\partial b_r}$ *so that, for instance,*

$$\begin{aligned}\bar{J}_{a,b}(b,a,m) &= \mathrm{E}[\partial_a L_{\mathrm{EP}}\ \partial_b L_{\mathrm{EP}}] = \mathrm{E}\big[\big(\partial_{b_l} L_{\mathrm{AEP}} + \partial_{b_r} L_{\mathrm{AEP}}\big)\big(\partial a_l L_{\mathrm{AEP}} + \partial_{a_r} L_{\mathrm{AEP}}\big)\big] \\ &= J_{a_l,b_l}(\bar{\mathbf{p}}) + J_{a_l,b_r}(\bar{\mathbf{p}}) + J_{a_r,b_l}(\bar{\mathbf{p}}) + J_{a_r,b_r}(\bar{\mathbf{p}}).\end{aligned}$$

The other elements are obtained in an analogous way. ∎

3.1 Properties of the estimators

We investigate now, form an analytical point of view, the sufficient conditions for consistency, asymptotic normality, and asymptotic efficiency of the AEP maximum likelihood estimators. The behavior of these estimators are different whenever the parameter m ought to be estimated or can be consider known. We analyze the two cases separately, starting with the case of unknown m.

From the definition of AEP in (3), the parameters $\mathbf{p} = (b_l, b_r, a_l, a_r, m)$ belong to the open set $\mathbf{D} = (0,+\infty) \times (0,+\infty) \times (0,+\infty) \times (0,+\infty) \times (-\infty,+\infty)$. Let $\mathbf{p_0}$ be the true parameters value, then:

Theorem 3.3 (Consistency): For any $\mathbf{p_0} \in \mathbf{D}$ maximum likelihood estimator $\hat{\mathbf{p}}$ is consistent, that is $\hat{\mathbf{p}}$ converges in probability to its true value $\mathbf{p_0}$.

Proof: For any $\mathbf{p_0} \in \mathbf{D}$ *there exists a compact* $\mathbf{P} \subset \mathbf{D}$ *such that:*

(1) $\mathbf{p_0} \in \mathbf{P}$
(2) $\forall \mathbf{p} \neq \mathbf{p_0}$, $\mathbf{p} \in \mathbf{P}$, it is $f(x_i|\mathbf{p}) \neq f(x_i|\mathbf{p_0})$
(3) $\forall \mathbf{p} \in \mathbf{P}$, $\log f(x_i|\mathbf{p})$ is continuous
(4) $\mathrm{E}[\sup_{\mathbf{P}} |\log f(x_i|\mathbf{p})|] < \infty$.

According to Theorem 2.5 in Newey and McFadden (1994: 2131, Chapter 36), these four conditions are sufficient to prove the statement. ∎

While consistency is easy to prove in general, finding sufficient conditions for asymptotic normality and efficiency is much more difficult. However, both can be found to apply for sufficiently large values of the shape parameters.

Theorem 3.4 (asymptotic normality and efficiency): If $b_l, b_r \geq 2$ *the unique a solution* $\hat{\mathbf{p}}$ *of the maximum likelihood problem (8) is asymptotically normal and efficient in the sense that* $\sqrt{N}(\hat{\mathbf{p}} - \mathbf{p_0})$ *converges in distribution to* $\mathcal{N}\{0, [J(\mathbf{p})]^{-1}\}$.

Proof: For the proof see Appendix 8.

Analogous results were derived in Agró (1995) for the symmetric EP distribution (1). The reason why the asymptotic efficiency and normality of the ML estimator can only be proved when $b_l, b_r \geq 2$ is due to the presence of singularities in the derivatives of L_{AEP} with respect to the parameter m. When this parameter is considered known, the situation becomes much simpler. In this case, the vector of unknown parameters $\mathbf{p} = (b_l, b_r, a_l, a_r)$ belongs to the open set $\mathbf{D} = (0, +\infty) \times (0, +\infty) \times (0, +\infty) \times (0, +\infty)$. Let $\mathbf{p_0}$ be the true parameters value, then the following holds:

Theorem 3.5 (consistency, asymptotic normality, and efficiency): If m is known, the solution $\hat{\mathbf{p}}$ of the maximum likelihood problem (8) converges in probability to its true value $\mathbf{p_0}$; $\hat{\mathbf{p}}$ is also asymptotically normal and efficient in the sense that $\sqrt{N}(\hat{\mathbf{p}} - \mathbf{p_0})$ converges in distribution to $\mathcal{N}\{0, [J(\mathbf{p})]^{-1}\}$.

Proof: The proof follows directly from the proofs of the previous theorems. Indeed, when m *is known no discontinuities in the derivatives of* $\partial \log\ f(x_i|\boldsymbol{p})/\partial p_j$ *emerge and hence the conditions required by Theorems 3.3 and 3.4 are always satisfied.* ■

Basically, the previous Theorem guarantee that when m is known, the ML estimates of $\mathbf{p}$ are consistent, asymptotically efficient, and normal on the whole parameter space. Of course, the same thing also applies to the symmetric EP density (Agró, 1995).

3.2 Extending the Fisher information matrix

The presence of singularities that forbids the extension of the results of Theorem 3.4 to small values of b's also affects the domain of definition of the elements of the Fisher matrix J.

The function $B_k(x)$ defined in (11) and all its derivatives are defined for $x > 0$ and for any k. Consequently, all the elements of J in (10), apart from J_{mm}, are defined on the whole parameter space. The latter element, on the contrary, is only defined when both b_l and b_r are greater than 0.5. When b_l or b_l move toward 0.5, the gamma function contained in that element encounters a pole (in $x = 0$) so that J_{mm} diverges. Of course, this phenomenon does not happen when the parameter m can be considered known. In that case, the 4×4 Fisher matrix (upper left block of J) is defined for any value of b_l and b_r and, according to Theorem 3.5, this matrix can be used to characterize the asymptotic error of the estimates over the whole parameter space. The presence of a pole in J_{mm} seems to suggest that, when m is unknown, the Fisher information matrix cannot be used to obtain a theoretical benchmark of the asymptotic errors involved in the ML estimation for small value of b. It turns out that this is not true. Indeed, the only estimates whose error diverges is $\hat{m}$.

To see how this mechanism works, consider the symmetric case in (13). In this case, the Fisher matrix $\bar{J}$ has a block diagonal structure, so that the value of the bottom right block, $\bar{J}_{m,m}$, does not affect the computation of the inverse of the upper left block, which contains the standard error of the estimates $\hat{a}$ and $\hat{b}$ and their

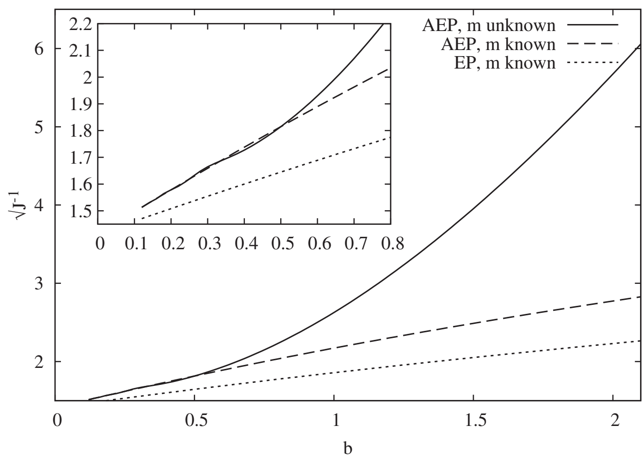

Figure 3 Relative asymptotic error $J_{b_l,b_l}^{-1/2}/b$ for AEP(b,b,1,1,0) as a function of b. Both the case with m known and unknown are displayed, together with the symmetric (EP) case $\bar{J}_{b,b}^{-1/2}/b$.

cross-correlation. Due to this block diagonal structure, the fact that m is known or not, does not have any effect on the asymptotic error of the estimates of the first two parameters. Hence, one can imagine that the upper left block of the Fisher information matrix can be used to obtain a theoretical values for the standard deviations σ_b and σ_a also for $b<0.5$.

In the asymmetric case, the block-diagonal structure of the Fisher information matrix disappears. In general, the fact that m is known or that its value has to be estimated does have an effect on the elements of the inverse information matrix associated with the standard error of the a's and b's estimates. Nonetheless, a peculiar cancellation in the computation of the elements of J^{-1} allows to recover a result analogous to the one found in the symmetric case. More precisely, when b_l or b_r goes toward 0.5, the element $J_{m,m}$ diverges and, correspondingly, $J_{m,m}^{-1}$ goes to 0, but, at the same time, the covariance terms of J^{-1} involving m tend to 0, so that the elements in the 4×4 upper left block remains finite. In fact, the 4×4 upper-left block of J^{-1} become positive definite and is equal to the 4×4 inverse Fisher information matrix obtained in the case in which m is known. Hence, analogously to the symmetric case, the elements of J can be used to recover a theoretical benchmark for the error of the estimated b's and a's on the whole parameters space. To illustrate the described behavior, the error on $\hat{b}$ and $\hat{a}$ estimated as the square root of the diagonal elements of J^{-1} are reported in Figures 3 and 4, respectively. For comparisons, both the case with m known and unknown are considered, and the associated element of the EP case $\bar{J}^{-1/2}$ is also reported. As can be clearly seen from the insets, when $b\rightarrow0.5$ the element of J^{-1} for the case of m unknown case are indistinguishable for the same elements computed assuming m known.

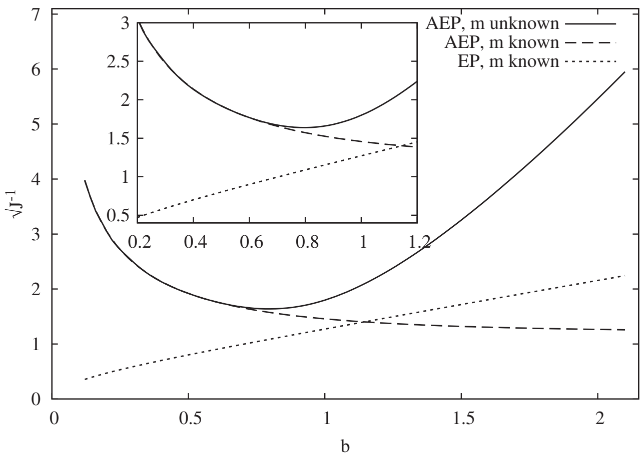

Figure 4 Asymptotic error $J_{a_l,a_l}^{-1/2}$ for AEP(b,b,1,1,0) as a function of b. Both the case with m known and unknown are displayed, together with the symmetric (EP) case $\bar{J}_{a,a}^{-1/2}$.

The same behavior can be observed also when only one parameter between b_l and b_r converges to 0.5.

What is the meaning of the inverse Fisher information matrix for values of b lower than 0.5? Can we exploit the continuation of the upper-left block of J^{-1} to investigate asymptotic efficiency and normality of ML estimators also in the region of the parameter space where b is low? Using extensive numerical simulations, we will try to answer these questions in the next section.

4. Numerical analyses

The analyses of this section focus on two aspects of the ML estimation of the symmetric exponential power and AEP distribution. First, we analyze the presence of bias in the estimates. We know from Theorem 3.3 that this bias progressively disappears when the sample becomes larger, but we are interested in characterizing its magnitude for relatively small samples. Second, we address the issue of the estimate errors, analyzing their behaviors for small samples, and trying to describe their asymptotic dynamics. These investigations are performed using numerical simulation. For a given set of parameters $\mathbf{p_0}$, we generate a large number of *i.i.d.* samples of size N, then for each parameter $p \in \mathbf{p_0}$, we compute the sample mean of the estimated value $\bar{p}(N; \mathbf{p_0}) = \mathrm{E}_N\left[\hat{p}|\mathbf{p_0}\right]$, where the expectation is computed over all the generated samples, and the associated bias $\tilde{p}(N; \mathbf{p_0}) = \bar{p}(N; \mathbf{p_0}) - p_0$.

This value is an estimate of the bias of $\hat{p}$ and, in general, depends on the true value $\mathbf{p_0}$. Since the ML estimates are consistent on the whole parameter space, we expect that $\lim_{N\to+\infty} \tilde{p}(N; \mathbf{p_0}) = 0$. The second measure that we consider is the sample variance of the estimated values, that is $\sigma_p^2(N;\mathbf{p_0}) = \mathrm{E}_N\big[(\hat{p}-\bar{p})^2|\mathbf{p_0}\big]$. Notice that the previous two quantities together define the root mean squared error (RMSE) of the estimate $p_{\mathrm{RMSE}}(N;\mathbf{p_0}) = \sqrt{\mathrm{E}_N\big[(\hat{p}-p_0)^2|\mathbf{p_0}\big]} = \sqrt{\tilde{p}^2+\sigma_p^2}$.

4.1 *Symmetric EP distribution*

Consider the symmetric EP distribution. In Table 1, we report the values of the bias and the estimates standard deviation for the three parameters a, b, and m computed using 10,000 independent samples of size N, with N running from 100 to 6400 and for different values of b. For the present qualitative discussion, the value of the parameters a and m is irrelevant; hence, we fix their value to 1 and 0, respectively. The values of the bias and the estimates standard deviation for the parameters a and b in the case of m known are reported in Table 2.

Since we consider 10,000 replications, the standard error on the reported bias estimation is nothing but the estimator standard deviation over $\sqrt{10,000}$. The bias estimates that results 2 SD away from 0 are reported in bold face in Tables 1 and 2. Looking at the first column of Table 1 for each estimate, one observes that the ML estimates of a and b are sometimes biased, while the estimated bias for m is never significantly different from 0. Notice that in all cases in which it is present, the bias seems to decrease proportionally to $1/N$ (for both known and unknown m). For the parameter a, the bias stops to be significantly different from 0 also for medium-sized samples (N around 400) while for b it is in general significant until largest sample sizes are reached. It is worthwhile to notice that, when the parameter m is considered known, the bias of the estimated values of a and b tends to increase, irrespectively of the true value of b.

Let us consider now the estimated standard errors $\sigma_p(N)$ in Table 1. The first thing to notice is that they are always at least one order of magnitude greater that the estimated biases, so that the contribution of the latter to the estimates RMSE is in general negligible. This means that, for any practical purposes, the ML estimates of the symmetric Power Exponential distribution can be considered *unbiased*. This is also true if one consider the case with m known, reported in Table 2. Indeed the values of the estimates standard error are practically identical for the two cases with only a couple of exceptions when N is small and b large. In this cases (see, e.g. $N=100$ and $b=1.4$), the standard error is much bigger when also m has to be estimated.

The second thing to notice is that the estimated standard errors seem to decrease with the inverse squared root of N. Indeed in Figure 5, we report for three different values of b, $\sqrt{N}\sigma_a(N)$ and $\sqrt{N}\sigma_b(N)$, for m unknown (Figure 5a and c) and known (Figure 5b and d). Notwithstanding the presence of noticeable small sample effects, these products always converge toward an asymptotic value. Since the convergence is from above, the efficiency of the estimator for small sample is lower than

Table 1 Bias and standard deviation of $\hat{b}$, $\hat{b}$, $\hat{a}$ and $\hat{m}$ estimated on 10,000 samples drawn from a power exponential distribution

N	$\tilde{b}/b$	σ_b/b	$\tilde{a}/a$	σ_a/a	$\tilde{m}$	σ_m	K
(b,a,m) = (0.4,1,0)							
100	−**0.018288**	0.177637	−**0.019566**	0.178384	−0.000365	0.059433	0
200	−**0.007221**	0.118821	−**0.008976**	0.122441	−0.000642	0.035281	0
400	−**0.004860**	0.081781	−**0.004822**	0.086703	−0.000240	0.021029	0
800	−**0.002362**	0.057095	−**0.002149**	0.061403	−0.000071	0.012641	0
1600	−**0.000950**	0.040103	−0.000650	0.043213	−0.000054	0.007717	0
3200	−0.000500	0.028149	−0.000387	0.030772	−0.000060	0.004570	0
6400	−0.000710	0.019966	−0.000173	0.021858	0.000006	0.002715	0
(b,a,m) = (0.8,1,0)							
100	**0.024698**	0.217721	−**0.005042**	0.141531	0.000457	0.102071	0
200	**0.010619**	0.137288	−**0.002619**	0.097276	−0.000158	0.068417	0
400	**0.004350**	0.091226	−**0.001645**	0.068244	0.000521	0.047679	0
800	**0.002038**	0.063613	−**0.000996**	0.047803	−0.000023	0.032717	0
1600	**0.000972**	0.044655	−0.000196	0.033742	0.000129	0.022560	0
3200	0.000426	0.031728	−0.000006	0.024025	−0.000123	0.015543	0
6400	0.000013	0.021858	−0.000119	0.016879	0.000014	0.010769	0
(b,a,m) = (1.4,1,0)							
100	**0.123678**	5.325462	**0.005878**	0.125171	−0.001145	0.112919	0
200	**0.030093**	0.161387	**0.002007**	0.085312	0.000602	0.077747	0
400	**0.013300**	0.106216	0.000311	0.059140	0.000302	0.055068	0
800	**0.006123**	0.072968	0.000307	0.041433	0.000249	0.038259	0
1600	**0.003050**	0.050587	0.000355	0.028948	−0.000124	0.026960	0
3200	**0.000927**	0.035539	−0.000204	0.020489	0.000240	0.019192	0
6400	0.000280	0.024811	−0.000176	0.014431	0.000081	0.013594	0
(b,a,m) = (2.2,1,0)							
100	**0.491071**	12.614268	**0.012540**	0.120088	−0.000602	0.099523	0
200	**0.049846**	0.194413	**0.005017**	0.078570	−0.000744	0.069450	0
400	**0.024967**	0.126713	**0.003576**	0.054255	−0.000774	0.047950	0
800	0.011329	0.084521	**0.001311**	0.037981	−0.000272	0.033816	0
1600	0.005102	0.058735	**0.000547**	0.026772	0.000015	0.023958	0
3200	0.002471	0.040739	0.000322	0.018683	0.000100	0.016927	0
6400	0.001520	0.028629	0.000298	0.013257	−0.000000	0.012098	0

K is the number of times the ML procedure did not converge.

Bias estimates 2 standard deviations away from 0 are reported in boldface.

Table 2 Bias and standard deviation of $\hat{b}$, $\hat{b}$, $\hat{a}$ and $\hat{m}$ estimated on 10,000 samples drawn from a power exponential distribution when m is known

N	$\tilde{b}/b$	σ_b/b	$\tilde{a}/a$	σ_a/a	K
(*b*,*a*) = (0.4,1)					
100	**0.040468**	0.174889	**0.018407**	0.180738	0
200	**0.018971**	0.118363	**0.007964**	0.123157	0
400	**0.008160**	0.081851	**0.003515**	0.086975	0
800	**0.004253**	0.057183	**0.002026**	0.061492	0
1600	**0.002472**	0.040050	**0.001478**	0.043217	0
3200	**0.001256**	0.028099	**0.000692**	0.030777	0
6400	0.000170	0.019822	0.000363	0.021830	0
(*b*,*a*) = (0.8,1)					
100	**0.054497**	0.207635	**0.014160**	0.138900	0
200	**0.025469**	0.134228	**0.006792**	0.096496	0
400	**0.011932**	0.090158	**0.003114**	0.068023	0
800	**0.005788**	0.063193	**0.001341**	0.047691	0
1600	**0.002764**	0.044496	**0.000928**	0.033709	0
3200	**0.001323**	0.031615	**0.000552**	0.024005	0
6400	**0.000482**	0.021620	0.000168	0.016814	0
(*b*,*a*) = (1.4,1)					
100	**0.074693**	0.260163	**0.013868**	0.121101	0
200	**0.033730**	0.157512	**0.006150**	0.084261	0
400	**0.015243**	0.104988	**0.002404**	0.058833	0
800	**0.007109**	0.072519	**0.001331**	0.041282	0
1600	**0.003590**	0.050498	**0.000879**	0.028906	0
3200	**0.001153**	0.035471	**0.000042**	0.020489	0
6400	0.000381	0.024579	0.000057	0.014364	0
(*b*,*a*) = (2.2,1)					
100	**0.152469**	5.046575	**0.014395**	0.113174	0
200	**0.046257**	0.187227	**0.006733**	0.077362	0
400	**0.023759**	0.124730	**0.004466**	0.053871	0
800	**0.010726**	0.083782	**0.001735**	0.037794	0
1600	**0.004872**	0.058559	**0.000779**	0.026715	0
3200	**0.002375**	0.040666	**0.000445**	0.018663	0
6400	**0.001438**	0.028421	**0.000352**	0.013206	0

K is the number of times the ML procedure did not converge.

Bias estimates 2 standard deviations away from 0 are reported in boldface.

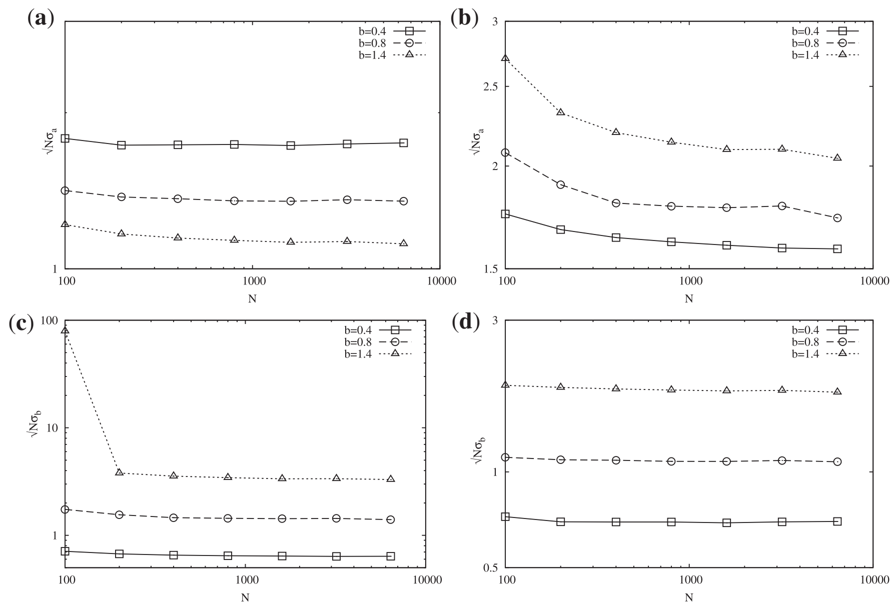

Figure 5 Rescaled standard error of the estimates of the parameter a (a and b) and b (c and d) as a function of the sample size N for the symmetric Subbotin distribution with $a = 1$, $m = 0$ and for different values of b. (a) $\sqrt{N}\sigma_a(N)$ where m is unknown; (b) $\sqrt{N}\sigma_a(N)$ where m is known; (c) $\sqrt{N}\sigma_b(N)$ where m is unknown; (d) $\sqrt{N}\sigma_b(N)$ where m is known.

the Cramer–Rao bound, implying a small sample inefficiency. Notice, however, that this inefficiency is in general of modest size.

For the case of unknown m, in order to compare the asymptotic behavior of the Monte Carlo estimates of the standard error with the theoretical prediction we consider the large samples limit

$$\lim_{N\to\infty} \sqrt{N}\sigma_p(N; \mathbf{p_0}) = \sigma_p^{\mathrm{ASY}}(\mathbf{p_0}). \tag{15}$$

We compute these values by extrapolating the three observations relative to the largest values of N estimating with Ordinary Least Squares the intercept of the following linear relation

$$\sqrt{N}\sigma_p \sim \alpha + \beta\frac{1}{N}. \tag{16}$$

The results for the different values of b are reported in Table 3 together with the theoretical prediction obtained from $\bar{J}^{-1}$ in (13). As expected, the agreement is extremely good, with discrepancies ~0.5%, in the region $b \geq 2$, where the Theorem 3.4 applies. In this region, the ML estimators of the EP density are, indeed, asymptotically efficient, so that the observed agreement serves as a consistency check of our extrapolation procedure. The same degree of agreement, however, is also observable

Table 3 Extrapolated values for the asymptotic (large N) estimates standard errors together with the theoretical Cramer–Rao values

	b		a		m	
b	σ^{ASY}	J^{-1}	σ^{ASY}	J^{-1}	σ^{ASY}	J^{-1}
0.2	0.3012	0.3016	2.3418	2.3519	0.0186	–
0.4	0.6366	0.6400	1.7547	1.7489	0.1921	–
0.6	1.0105	1.0134	1.4849	1.4994	0.5628	0.4130
0.8	1.4024	1.4198	1.3550	1.3604	0.8499	0.8134
1.0	1.8608	1.8574	1.2654	1.2715	1.0041	1.0000
1.2	2.2602	2.3244	1.2100	1.2095	1.0808	1.0700
1.4	2.7697	2.8194	1.1550	1.1639	1.0912	1.0817
1.6	3.3065	3.3411	1.1195	1.1287	1.0762	1.0651
1.8	3.8407	3.8883	1.0928	1.1008	1.0480	1.0353
2.0	4.4819	4.4599	1.0900	1.0779	1.0036	1.0000
2.2	4.9894	5.0550	1.0536	1.0587	0.9674	0.9632

in the region $0.5 < b < 2$, where the Fisher information matrix is defined, but no theoretical results guarantee the efficiency of the estimator for large samples. Moreover, quite surprising, the agreement remains high, for the a and b estimators, also in the region $b < 0.5$, where the Fisher information matrix cannot be defined according to equation (12) but can be analytically continued, as discussed in Section 3.2.

In conclusions, the previous numerical investigation extends in many respect to the analytical findings of the existing literature. We have shown that for the symmetric EP distribution:

(1) the bias of the ML estimators, being very small, can be safely ignored at least for samples with more than 100 observations;
(2) the ML estimators of a, b, and m are asymptotically efficient, independently of the value of the true parameters and of the fact that the value of m is known or unknown; and
(3) the continuation of the Fisher information matrix to the region with $b < .5$ can be used to obtain a reliable measure of the error involved in the ML estimation of parameters a and b.

4.2 AEP distribution

This section extends the numerical analysis to the case of AEP distribution. For the sake of clarity, we split our analysis in two steps. First, we analyze the asymptotic behavior of the ML estimates when the true parameters have symmetric values. Second, we comment on the observed effects when different degrees of asymmetry characterize the true values of the shape parameters b_l and b_r.

In Table 4, we report the values of the bias and the estimates standard deviation for the five parameters a_l, a_r, b_l, b_r, and m computed using 10,000 independent samples of size N, with N running from 100 to 6400. The samples are randomly generated from equation (3) considering different values for the parameters $b_l = b_r$. Again the exact value of the a's and m parameters is irrelevant for the present discussion and we set $a_l = a_r = 1$ and $m = 0$ for all simulations. As can be seen, the picture that emerges is identical to the symmetric case. The bias is in general present for small samples, apart for the estimate $\hat{m}$ which seems in general unbiased. When present, the bias tends to decrease proportionally to $1/N$ and, for the parameters a_l and a_r it becomes statistically indistinguishable from zero with the increase of the sample size. Notice that for $N > 100$, the bias is always at least one order of magnitude smaller than the standard deviation. Consequently, also in the case of AEP distribution, when the true parameters are symmetric, and for sufficiently large samples ($N > 100$), the ML estimates can be considered, for any practical purposes, *unbiased*. Also the behavior of the estimates standard deviation is substantially identical to what observed in the case of symmetric distribution. Indeed, the plots in Figure 6a and c confirm that the rescaled estimates $\sqrt{N}\sigma_p(N)$ approach flat lines when N becomes large, making the asymptotic efficiency apparent. However, the small sample effect seems to last a little longer: when one consider small values of b (see Figure 6a), it is still noticeable for sample as large as 1000 observations.

In Table 5, we report the values of the bias and the estimates standard deviation for the four parameters b_l, b_r, a_l, and a_r, obtained with the Monte Carlo procedure illustrated above, in the case in which the parameter m is assumed known. No large differences are observed in the behavior of biases and standard deviations with respect to the case of unknown m. The general increase of the bias level, already observed for the symmetric distribution, is still there. Concerning the estimates standard errors, in Figure 6 notice that the display behavior similar to what observed in the Figure 6a and c, confirming that the deviations from the Cramer–Rao bound is essentially due to small sample effect. In the case of m known, these effects tend to disappear completely when $N > 400$.

In order to judge the reliability of J^{-1} in estimating the observed errors, we compute the asymptotic values of the standard errors σ_p^{ASY} extrapolating the three estimates obtained with the largest samples ($N = 1600$, 3200, 6400) following the same procedure used above [cf. equation (16)]. The results are reported in Table 6 (upper part). Again, the agreement between the values extrapolated from numerical simulations and the theoretical values obtained from the inverse information matrix J^{-1} is remarkably high: discrepancies are around 1% both in the region of high and low b's, confirming that J^{-1} can be used to obtain a value of the asymptotic standard errors of the estimates also in the region in which Theorem 3.4 does not apply.

Table 4 Bias and standard deviation of $\hat{b}_l$, $\hat{b}_r$, $\hat{a}_l$, $\hat{a}_r$ and $\hat{m}$ estimated on 10,000 samples drawn from an AEP distribution

N	$\tilde{b}_l/b_l$	σ_{b_l}/b_l	$\tilde{b}_r/b_r$	σ_{b_r}/b_r	$\tilde{a}_l/a_l$	σ_{a_l}/a_l	$\tilde{a}_r/a_r$	σ_{a_r}/a_r	$\tilde{m}$	σ_m	K
$(b_l, b_r, a_l, a_r, m) = (0.5,0.5,1,1,0)$											
100	**0.026188**	0.281091	**0.020557**	0.271076	**0.014968**	0.215253	**0.014931**	0.210935	**0.003749**	0.166962	1
200	**0.012562**	0.162519	**0.012789**	0.161660	**0.005921**	0.140722	**0.006872**	0.141735	0.000388	0.091752	0
400	**0.007066**	0.107014	**0.005707**	0.107006	0.001429	0.096847	**0.003919**	0.098412	0.000393	0.056794	0
800	**0.003648**	0.072630	**0.003716**	0.074012	0.001149	0.068157	**0.002622**	0.069134	−0.000345	0.034856	0
1600	**0.001486**	0.049725	0.000821	0.049235	0.000266	0.048057	**0.001194**	0.047838	0.000002	0.020220	1
3200	0.000433	0.034397	0.000309	0.034407	−0.000006	0.034113	0.000448	0.034070	−0.000090	0.012452	0
6400	0.000306	0.023751	0.000086	0.024056	0.000160	0.024499	0.000474	0.024146	0.000011	0.007887	0
$(b_l, b_r, a_l, a_r, m) = (1.5,1.5,1,1,0)$											
100	**0.138699**	0.707531	**0.130697**	0.830274	**0.041225**	0.376155	**0.042109**	0.371139	0.000928	0.553390	45
200	**0.059863**	0.364016	**0.049007**	0.350531	**0.021834**	0.255554	**0.016018**	0.252260	0.005378	0.385947	0
400	**0.025145**	0.226582	**0.023601**	0.224548	**0.009361**	0.176657	**0.008696**	0.177574	0.000974	0.274766	0
800	**0.012233**	0.154245	**0.011369**	0.153025	**0.004094**	0.124075	**0.004513**	0.124694	−0.000187	0.194852	0
1600	**0.006437**	0.106212	**0.004958**	0.104984	**0.002698**	0.087034	0.001332	0.086825	0.001153	0.137088	0
3200	**0.002850**	0.072848	**0.002355**	0.073090	0.001223	0.060221	0.000308	0.060127	0.000990	0.094983	0
6400	0.001065	0.050449	**0.001670**	0.050608	0.000367	0.041679	0.000469	0.041504	0.000036	0.065446	0
$(b_l, b_r, a_l, a_r, m) = (2.5,2.5,1,1,0)$											
100	**0.216104**	1.077383	**0.194571**	0.988308	**0.052892**	0.540990	**0.051839**	0.537115	0.001134	0.730692	357
200	**0.105139**	1.287989	**0.096703**	0.752724	**0.032009**	0.432849	**0.036462**	0.432766	−0.003785	0.593991	8
400	**0.048444**	0.382355	**0.036445**	0.375708	**0.024977**	0.345262	**0.017779**	0.342416	0.003945	0.477221	0
800	**0.020174**	0.270658	**0.019085**	0.269044	**0.010986**	0.262583	**0.012462**	0.262840	−0.001216	0.367170	0
1600	**0.009100**	0.192912	**0.011360**	0.191377	**0.005337**	0.193851	**0.008018**	0.193535	−0.001951	0.272406	0
3200	**0.004226**	0.136708	**0.006990**	0.134924	0.002167	0.140358	**0.005429**	0.139778	−0.002423	0.197663	0
6400	**0.002709**	0.095266	**0.003138**	0.094106	0.001603	0.098212	**0.002417**	0.097851	−0.000599	0.138287	0

K is the number of times the ML procedure did not converge.

Bias estimates 2 standard deviations away from 0 are reported in boldface.

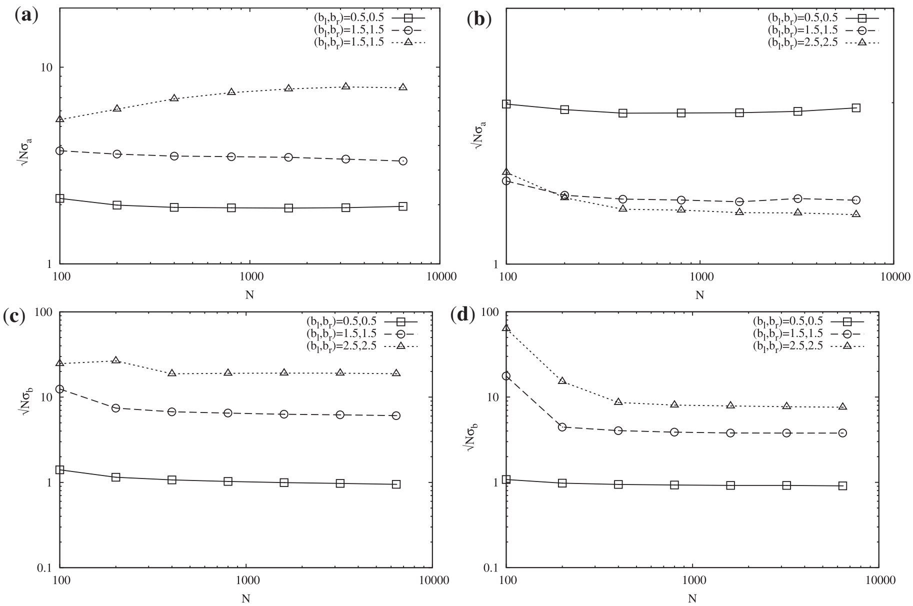

Figure 6 Rescaled standard error of the estimator of the parameters a_l (a and b) and b_l (c and d) as a function of the sample size N, for the asymmetric Subbotin distribution for $a_l = a_r = 1$, $m = 0$ and different (but equal) values of b_l and b_r. (a) $\sqrt{N}\sigma_a(N)$ where m is unknown; (b)$\sqrt{N}\sigma_a(N)$ where m is known; (c) $\sqrt{N}\sigma_b(N)$ where m is unknown; (d) $\sqrt{N}\sigma_b(N)$ where m is known.

Finally, we have explored the behavior of the ML estimator when the true values of the parameters b_l and b_r are different. Results are reported in Table 7 for a selection of different values of the two shape parameters. The most noticeable effect of the introduction of asymmetry in the true values of the parameters is an increase in the biases of their estimates. First, in this situation, also the estimate of location parameter m results biased. Second, the observed biases of the estimates of b remain statistically different from zero also for relatively large samples ($N = 6400$). Again, when the sample size increases, the biases still decrease proportionally to $1/N$. At the same time, the behavior of the estimates standard error σ_p resembles the ones observed in the previous cases: as the plots in Figure 7 show, all the rescaled standard errors defined accordingly to equation (15) asymptotically approach flat lines so that the ML estimator can be considered asymptotically efficient. The different asymptotic behaviors of the bias and the standard error imply that for sufficiently large samples, the contribution of the former to the estimates RMSEs becomes negligible. Indeed, it is already the case for sample sizes around 100 observations. As in the symmetric case, these results do not change when m is known (cf. Table 8).

Table 5 Bias and standard deviation of $\hat{b}_l$, $\hat{b}_r$, $\hat{a}_l$, $\hat{a}_r$ and $\hat{m}$ estimated on 10,000 samples drawn from an AEP distribution with μ known

N	$\tilde{b}_l/b_l$	σ_{b_l}/b_l	$\tilde{b}_r/b_r$	σ_{b_r}/b_r	$\tilde{a}_l/a_l$	σ_a/a_l	$\tilde{a}_r/a_r$	σ_{a_r}/a_r	K
$(b_l, b_r, a_l, a_r, m) = (0.5, 0.5, 1, 1, 0)$									
100	**0.064224**	0.215717	**0.064604**	0.216754	**0.021229**	0.198856	**0.022081**	0.197856	0
200	**0.031114**	0.138348	**0.031988**	0.138717	**0.010393**	0.137301	**0.011382**	0.139067	0
400	**0.015344**	0.094456	**0.014446**	0.093711	**0.003460**	0.095598	**0.005939**	0.097347	0
800	**0.007962**	0.065844	**0.007348**	0.065663	**0.002087**	0.067657	**0.003570**	0.068646	0
1600	**0.003681**	0.046000	**0.003035**	0.045963	0.000915	0.047896	**0.001879**	0.047672	0
3200	**0.001620**	0.032504	**0.001368**	0.032498	0.000343	0.034064	**0.000780**	0.034026	0
6400	**0.000878**	0.022711	**0.000713**	0.022942	0.000392	0.024454	**0.000653**	0.024135	0
$(b_l, b_r, a_l, a_r, m) = (1.5, 1.5, 1, 1, 0)$									
100	**0.170308**	0.909158	**0.170702**	1.173527	**0.019263**	0.142899	**0.021168**	0.141552	0
200	**0.061326**	0.216921	**0.058578**	0.209759	**0.008688**	0.095054	**0.009503**	0.094936	0
400	**0.027404**	0.134213	**0.027293**	0.134654	**0.003746**	0.066096	**0.004358**	0.066122	0
800	**0.013651**	0.091370	**0.012676**	0.091370	**0.001609**	0.046557	**0.001857**	0.046577	0
1600	**0.006274**	0.063041	**0.006320**	0.063171	**0.000687**	0.032683	**0.000711**	0.032923	0
3200	**0.002594**	0.044291	**0.003323**	0.044531	0.000038	0.023429	0.000279	0.023403	0
6400	**0.001237**	0.031043	**0.001905**	0.031479	0.000071	0.016446	0.000218	0.016504	0
$(b_l, b_r, a_l, a_r, m) = (2.5, 2.5, 1, 1, 0)$									
100	**0.498902**	3.420278	**0.411656**	2.536465	**0.030263**	0.148275	**0.027090**	0.147140	1
200	**0.099737**	0.381414	**0.098326**	0.430159	**0.011924**	0.094034	**0.011169**	0.093641	0
400	**0.043165**	0.175576	**0.037703**	0.172490	**0.006442**	0.063279	**0.005061**	0.063162	0
800	**0.018806**	0.116601	**0.016616**	0.113832	**0.002202**	0.044585	**0.002169**	0.044289	0
1600	**0.008874**	0.078796	**0.009164**	0.078615	**0.001305**	0.031190	**0.001403**	0.031516	0
3200	**0.005009**	0.054622	**0.005034**	0.054509	**0.001012**	0.022023	**0.000987**	0.021996	0
6400	**0.002764**	0.038202	**0.002561**	0.037959	**0.000642**	0.015458	**0.000659**	0.015617	0

K is the number of times the ML procedure did not converge.

Bias estimates 2 standard deviations away from 0 are reported in boldface.

Table 6 Extrapolated values for the asymptotic (large N) estimates standard errors of the EP together with the theoretical Cramer–Rao values

	σ^{ASY}		$\mathbf{J}^{-1}$		σ^{ASY}		$\mathbf{J}^{-1}$		σ^{ASY}	$\mathbf{J}^{-1}$
(b_l, b_r)	b_l	b_r	$b_l = b_r$		a_l	a_r	$a_l = a_r$		m	m
(0.4, 0.4)	0.7181	0.7083	0.6907		2.1407	2.1628	2.1341		0.3740	–
(0.5, 0.5)	0.9392	0.9565	0.9073		1.9636	1.9386	1.9199		0.5788	–
(0.75, 0.75)	1.6974	1.6811	1.6114		1.6557	1.6755	1.6458		1.4214	1.1146
(1.5, 1.5)	5.9582	6.0244	5.9308		3.2969	3.2845	3.2534		5.1804	5.1064
(2.5, 2.5)	19.0743	18.7929	19.2629		7.9499	7.9109	8.0497		11.2056	11.3643
(b_l, b_r)	b_l	b_r	b_l	b_r	a_l	a_r	a_l	a_r	m	m
(0.5, 1.5)	0.8709	3.8556	0.8174	3.5742	2.1005	1.5258	2.0572	1.3205	0.8588	–
(0.5, 2.5)	0.8802	7.2828	0.7991	6.9769	2.0958	1.4619	2.0710	1.1991	0.9164	–
(1.5, 2.5)	6.8920	14.3902	6.7661	14.1345	4.1304	5.3853	4.0050	5.2242	7.1248	6.9119

We conclude the section on the numerical analysis with some brief comment on the technical aspects of ML estimation. The solution of the problem in equation (8) is in general made difficult by the fact that both the AEP and EP densities are not analytic functions. The situation becomes more severe when small values of the shape parameter b are considered. In this case, the likelihood as a function of the location parameter m possesses many local maxima, located on the observations that compose the samples. In order to overcome this difficulties, the ML estimation presented above have been obtained with a three steps procedure: in each case, the negative likelihood minimization started with initial conditions obtained with a simple method of moments. Then a global minimization was performed in order to obtain a first ML estimate, which is later refined performing several separate minimizations in the different intervals defined by successive observations in the neighborhood of the first estimate. Even if this method is not guaranteed to provide the global minimum, we checked that in the whole range of parameters analyzed, discrepancies were always negligible.[2] For further details on the minimization methods utilized, the reader is referred to Bottazzi (2004).

As already observed in Agró (1995) for the EP distribution, when the value of the shape parameter b is large and the size of the sample relatively small, the minimization procedure can fail to converge. In the case of AEP distribution, the situation is in general worsened especially when the shape parameters b_l and b_r present largely

[2]Observed discrepancies were generally due to the presence of several clustered observations.

Table 7 Bias and standard deviation of $\hat{b}_l$, $\hat{b}_r$, $\hat{a}_l$, $\hat{a}_r$ and $\hat{m}$ estimated on 10,000 samples drawn from an AEP distribution

N	$\tilde{b}_l/b_l$	σ_{b_l}/b_l	$\tilde{b}_r/b_r$	σ_{b_r}/b_r	$\tilde{a}_l/a_l$	σ_{a_l}/a_l	$\tilde{a}_r/a_r$	σ_{a_r}/a_r	$\tilde{m}$	σ_m	K
$(b_l, b_r, a_l, a_r, m) = (0.5, 1.5, 1, 1, 0)$											
100	**0.016059**	0.251608	**0.066257**	0.403796	**0.026195**	0.228994	**−0.009739**	0.216587	**0.019185**	0.191960	84
200	**0.005344**	0.147271	**0.032755**	0.232989	**0.012207**	0.154975	**−0.003246**	0.136095	**0.006282**	0.109004	3
400	**0.002462**	0.096266	**0.016076**	0.145892	**0.006336**	0.106578	−0.001011	0.088222	**0.002936**	0.066112	1
800	0.000016	0.064622	**0.010703**	0.098329	**0.003381**	0.074980	0.001126	0.059925	−0.000526	0.042494	0
1600	−0.000799	0.045051	**0.006403**	0.068035	**0.002236**	0.052221	0.000876	0.041374	**−0.000907**	0.027879	0
3200	−0.000847	0.031354	**0.003399**	0.047031	**0.001514**	0.036679	0.000320	0.028286	**−0.000393**	0.017856	0
6400	−0.000348	0.021951	**0.001960**	0.032511	**0.000977**	0.026344	0.000344	0.019415	**−0.000313**	0.011392	0
$(b_l, b_r, a_l, a_r, m) = (0.5, 2.5, 1, 1, 0)$											
100	**0.022468**	0.255162	**0.101449**	0.555071	**0.020517**	0.225258	**−0.018580**	0.219204	**0.028914**	0.196187	423
200	**0.008303**	0.149654	**0.050432**	0.287029	**0.010281**	0.153611	**−0.004446**	0.138285	**0.010341**	0.112153	7
400	**0.004299**	0.098062	**0.020972**	0.169655	**0.005071**	0.106479	**−0.001899**	0.086841	**0.004974**	0.067606	2
800	**0.001987**	0.065114	**0.009224**	0.111832	**0.001813**	0.074475	**−0.001770**	0.057358	**0.002692**	0.042156	0
1600	0.000572	0.044927	**0.005221**	0.077055	**0.001262**	0.052684	−0.000442	0.039397	**0.001054**	0.026905	0
3200	0.000452	0.031767	**0.003277**	0.053408	**0.000906**	0.036877	0.000328	0.027017	0.000215	0.018008	0
6400	0.000171	0.022005	**0.001973**	0.036795	0.000444	0.026330	0.000501	0.018571	−0.000034	0.011815	0
$(b_l, b_r, a_l, a_r, m) = (1.5, 2.5, 1, 1, 0)$											
100	**0.172840**	0.807995	**0.163922**	1.018400	**0.083851**	0.413484	−0.003162	0.479259	**0.076579**	0.635499	238
200	**0.078985**	0.394488	**0.061394**	0.510150	**0.048404**	0.297385	**−0.008636**	0.354509	**0.050121**	0.472570	3
400	**0.038409**	0.257181	**0.019304**	0.311780	**0.027430**	0.215093	**−0.007662**	0.262142	**0.029973**	0.352509	0
800	**0.020593**	0.175969	**0.005227**	0.211818	**0.015980**	0.153095	**−0.007333**	0.189167	**0.019614**	0.254872	0
1600	**0.007903**	0.119389	**0.005257**	0.146614	**0.005724**	0.105444	−0.001113	0.133336	**0.006430**	0.178423	0
3200	**0.002899**	0.083172	**0.002837**	0.103493	**0.002151**	0.074641	0.000119	0.095920	0.002139	0.127786	0
6400	**0.001851**	0.057875	0.001033	0.072014	**0.001390**	0.051602	−0.000185	0.066737	0.001534	0.088487	0

K is the number of times the ML procedure did not converge

Bias estimates 2 standard deviations away from 0 are reported in boldface.

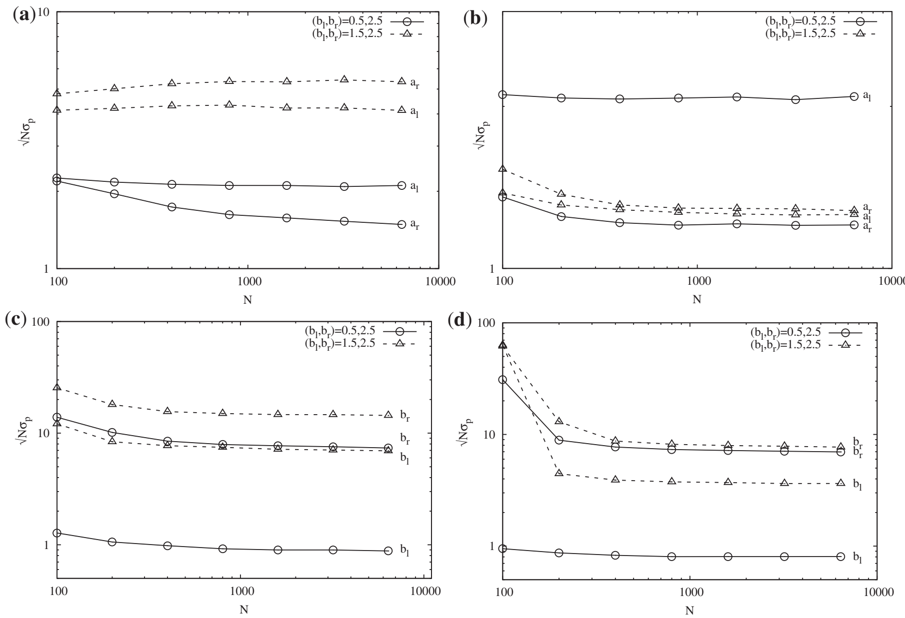

Figure 7 Standard error of the estimator of the parameters a_l, a_r (a and b) and b_l, b_r (c and d) as a function of the sample size N for the asymmetric Subbotin distribution for different values of b_l, $b_r = 2.5$, $a_l = a_r = 1$ and $m = 0$. (a) $\sqrt{N}\sigma_a(N)$ where m is unknown; (b)$\sqrt{N}\sigma_a(N)$ where m is known; (c) $\sqrt{N}\sigma_b(N)$ where m is unknown; (d) $\sqrt{N}\sigma_b(N)$ where m is known.

different true values (see e.g. $N = 100$, $b_l = 0.5$, and $b_r = 2.5$ in Table 4). The number of failures is reported in the columns "K" of the relevant Tables.

5. Empirical applications

In the present section, we test the ability of the AEP to fit empirical distributions obtained from different economic and financial datasets. We compare the AEP with the SEP, the α-Stable family, and the Generalized Hyperbolic (GHYP) estimating their parameters via ML procedures [for parametrization and details on the SEP, the α-Stable, and on the GHYP see DiCiccio and Monti (2004), Nolan (1998) and McNeil *et al.* (2005) respectively]. In order to evaluate the accuracy of the agreement between the empirical observed distributions and the theoretical alternatives, we consider two complementary measures of goodness-of-fit, the Kolmogorov–Smirnov D and the Cramer–Von Mises $W2$ defined as

$$D = \sup_n \left| F^{\text{Emp}}(x_n) - F^{\text{Th}}(x_n) \right| \quad W2 = \frac{1}{12n} + \sum_n \left(F^{\text{Emp}}(x_n) - F^{\text{Th}}(x_n) \right)^2, \quad (17)$$

Table 8 Bias and standard deviation of $\hat{b}_l$, $\hat{b}_r$, $\hat{a}_l$, $\hat{a}_r$ and $\hat{m}$ estimated on 10,000 samples drawn from an AEP distribution with μ known

N	$\tilde{b}_l/b_l$	σ_{b_l}/b_l	$\tilde{b}_r/b_r$	σ_{b_r}/b_r	$\tilde{a}_l/a_l$	σ_{a_l}/a_l	$\tilde{a}_r/a_r$	σ_{a_r}/a_r	K
$(b_l, b_r, a_l, a_r, m) = (0.5, 1.5, 1, 1, 0)$									
100	**0.053773**	0.195910	**0.125824**	0.837937	**0.008226**	0.210580	**0.019986**	0.139315	0
200	**0.025039**	0.125204	**0.051494**	0.195526	**0.004733**	0.147089	**0.009401**	0.094616	0
400	**0.011770**	0.084416	**0.024379**	0.126572	**0.002732**	0.103001	**0.004863**	0.066439	0
800	**0.005727**	0.058028	**0.011656**	0.086037	0.000728	0.072828	**0.001962**	0.046634	0
1600	**0.002342**	0.041046	**0.005938**	0.060213	0.000719	0.051191	**0.000677**	0.032976	0
3200	**0.000659**	0.028824	**0.003137**	0.042609	0.000707	0.035983	0.000243	0.023462	0
6400	**0.000484**	0.020419	**0.001537**	0.029969	0.000432	0.025943	0.000128	0.016550	0
$(b_l, b_r, a_l, a_r, m) = (0.5, 2.5, 1, 1, 0)$									
100	**0.049015**	0.189674	**0.228050**	1.238896	0.000973	0.210265	**0.022900**	0.135733	0
200	**0.023643**	0.122868	**0.072195**	0.251545	0.000192	0.146596	**0.010420**	0.088294	0
400	**0.011436**	0.082733	**0.031470**	0.154247	0.000626	0.103198	**0.005328**	0.060806	0
800	**0.005635**	0.056868	**0.014698**	0.103640	−0.000054	0.073261	**0.002103**	0.042548	0
1600	**0.002651**	0.040238	**0.007654**	0.071829	0.000320	0.052042	**0.001282**	0.030253	0
3200	**0.001697**	0.028480	**0.004188**	0.050021	0.000367	0.036385	**0.000941**	0.021258	0
6400	**0.000874**	0.020158	**0.002018**	0.034866	0.000088	0.026084	**0.000587**	0.015053	0
$(b_l, b_r, a_l, a_r, m) = (1.5, 2.5, 1, 1, 0)$									
100	**0.253803**	4.212897	**0.435188**	2.473012	**0.018725**	0.138093	**0.031128**	0.152805	0
200	**0.059715**	0.209753	**0.099552**	0.367232	**0.007405**	0.092740	**0.012295**	0.097120	0
400	**0.026696**	0.130166	**0.038787**	0.174597	**0.003372**	0.064278	**0.005117**	0.065592	0
800	**0.012453**	0.088677	**0.018056**	0.115543	**0.001334**	0.044944	**0.002241**	0.045771	0
1600	**0.006231**	0.061846	**0.009675**	0.079525	0.000511	0.031555	**0.001409**	0.032307	0
3200	**0.002890**	0.042806	**0.004814**	0.055465	0.000249	0.022223	**0.000740**	0.022808	0
6400	**0.001675**	0.030318	**0.002671**	0.038534	0.000268	0.015741	**0.000596**	0.016006	0

K is the number of times the ML procedure did not converge

Bias estimates 2 standard deviations away from 0 are reported in boldface.

where F^{Emp} and F^{Th} stands for the empirical and theoretical distribution, respectively. These two statistics can be considered complementary as they capture somehow different effects. The D statistics is indeed proportional to the largest observed absolute deviation of the theoretical form, the empirical distribution, while the $W2$ is intended to account for their "average" discrepancy over the entire sample.

Notice that the following discussion is not focused on assessing whether the deviation of the theoretical models from actual data can be considered a significant signal of misspecification. Rather, we are interested in evaluating the relative abilities of the different families to properly describe the behavior of the empirical distributions. Hence, all the figures associated with the different statistics should be regarded in comparative and not absolute terms.

5.1 French electricity market

As a first application, we analyze data from Powernext, the French power exchange. We consider a data set containing the day-ahead electricity prices, in different hours, from November 2001 to August 2006,[3] and we build the empirical distribution of the corresponding daily log returns. Then using the goodness-of-fit statistics defined in equation (17), we investigate the ability of the four competing families to reproduce the observed distributions. Results are reported in Table 9.

Two main evidences emerge from the reported figures. First, the AEP outperforms all the other distributions both in terms of the Kolmogorov–Smirnov and of the Cramer–Von Mises statistics. In particular, from Table 9, it is clear that while the observed Kolmogorov–Smirnov statistics D is, for the AEP, only slightly lower than the ones obtained for the other families the same appears not true in the case of the Cramer–Von Mises test. Indeed, the values of the $W2$ statistic are significantly lower for the AEP being always less than half of the average of the other three. In order to provide a more revealing, albeit qualitative, assessment of the relative ability of the different families in reproducing the empirical distribution we present, in Figure 8, two plots, for the AEP and the GHYP, respectively, of the function $\Delta(x)$ defined as

$$\Delta(x) = F^{\mathrm{Emp}}(x) - F^{\mathrm{Th}}(x). \tag{18}$$

Deviations of $\Delta(x)$ from the constant line $y=0$ represent the local discrepancy between the theoretical and the empirical distribution. This figure, while confirming in accordance with formal tests the better fit of the AEP, adds also some interesting insights: the AEP is clearly better in the whole central part of the distribution and in

[3]These prices are fixed on day, separately for the 24 individual hours, for delivery on the same day or on the following.

Table 9 Maximum likelihood estimates (standard errors in parentheses) of the shape parameters, b_l and b_r, of the AEP density together with the EDF goodness-of-fit statistics for four different families of distribution

			Goodness of fit — *W2*				Goodness of fit — *D*			
Hour	b_l	b_r	**AEP**	**GHYP**	**Stable**	**SEP**	**AEP**	**GHYP**	**Stable**	**SEP**
10.00 a.m.	0.565 (0.022)	0.893 (0.043)	**0.287**	1.365	1.436	1.339	**0.030**	0.053	0.051	0.042
12.00 a.m.	0.625 (0.026)	0.985 (0.051)	**0.155**	0.253	0.644	0.390	**0.022**	0.024	0.036	0.032
2.00 p.m.	0.600 (0.024)	0.999 (0.051)	**0.147**	0.752	1.016	0.573	**0.026**	0.040	0.044	0.035
5.00 p.m.	0.591 (0.023)	1.003 (0.051)	**0.193**	0.592	0.774	0.847	**0.027**	0.036	0.037	0.042
8.00 p.m.	0.650 (0.027)	0.912 (0.046)	**0.091**	0.178	0.576	0.239	**0.017**	0.024	0.033	0.022

Data are daily log returns of electricity prices from the French power exchange, Powernext. Goodness of fit statistics for the best performing family are reported in boldface.

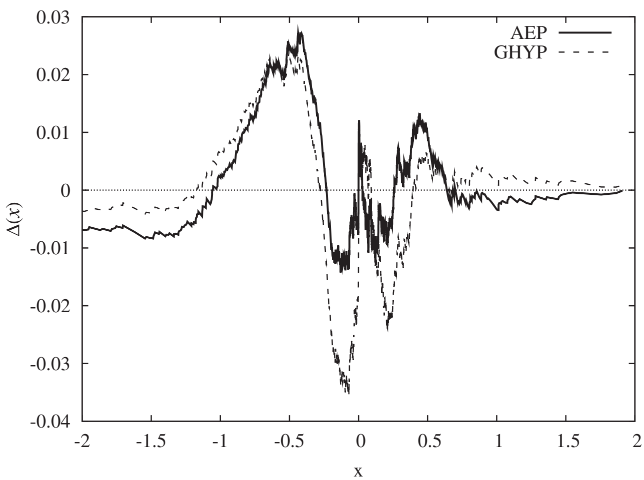

Figure 8 Deviations $\Delta(x)$ of the AEP and of the GHYP from the empirical distribution. Data are daily log returns of the French electricity price at 5 p.m.

its upper tail, while the opposite is true for the lower tail where the GHYP seems slightly preferable.[4]

The second evidence emerging from Table 9 regards the difference between the estimated values of the AEP shape parameters b_l and b_r, which suggests the presence

[4]For the sake of clarity, we do not report the function $\Delta(x)$ for the α-Stable and the SEP, since from Table 9 it is apparent that their ability to fit the empirical distribution is substantially worse.

Table 10 Maximum likelihood estimates (standard errors in parentheses) of the shape parameters of the AEP density together with the EDF goodness-of-fit statistics for four different families of distribution

			Goodness of fit — *W2*				Goodness of fit — *D*			
Currencies	b_l	b_r	AEP	GHYP	Stable	SEP	AEP	GHYP	Stable	SEP
usd4eu	1.193 (0.127)	1.503 (0.165)	**0.052**	0.073	0.351	3.420	**0.018**	0.022	0.036	0.107
usd4uk	1.385 (0.172)	1.688 (0.217)	**0.037**	0.044	0.214	0.120	**0.016**	0.019	0.035	0.026
sz4usd	1.455 (0.163)	1.374 (0.167)	**0.054**	0.060	0.339	0.078	**0.018**	0.019	0.039	0.021
si4usd	1.110 (0.119)	1.530 (0.153)	0.038	**0.033**	0.066	2.798	0.020	**0.016**	0.020	0.088
jp4usd	1.195 (0.125)	1.541 (0.176)	**0.019**	0.029	0.141	0.703	**0.014**	0.018	0.032	0.059

Data are daily log first difference on different exchange rates. *Source*: FRED® Federal Reserve Economic Data.

Goodness of fit statistics for the best performing family are reported in boldface.

of substantial asymmetries in the empirical distribution of electricity price returns. This finding is not a peculiar feature of the French market but applies to a number of different power exchanges, see Sapio (2008) for a broader analysis. As such, it provides a potent, empirically based, case for the development of class of distributions able to cope at the same time with fat tails and skewness.

To sum up, our evidence suggests that the AEP fits systematically better the skewed distribution function of the log returns of French electricity prices presenting, at the same time, the lowest overall discrepancy and the lowest maximum deviation from the corresponding empirical benchmark.

5.2 Exchange rates market

As a second application, we consider exchange rates data collected from FRED®, a database of over 15,000 U.S. economic time series available at the Federal Reserve Bank of St Louis. We select a data set containing five different exchange rates and we focus on the most recent 1000 observations.[5] We build empirical distributions of the (log) differenced exchange rates series and, as we did in the previous section, we test the relative ability of the four families under investigation to fit their observed counterpart.

Results of the goodness-of-fit test are reported in Table 10. Once again the AEP and the GHYP clearly show, when compared with the other two families, a better

[5]The exchange rates analyzed are as follows: U.S. Dollar to one Euro, U.S. Dollar to one U.K. Pound, Japanese Yen to one U.S. Dollar, Singapore Dollars to one U.S. Dollars and Swiss Francs to one U.S. Dollars. The time window goes from August 25, 2003 to August 14, 2007.

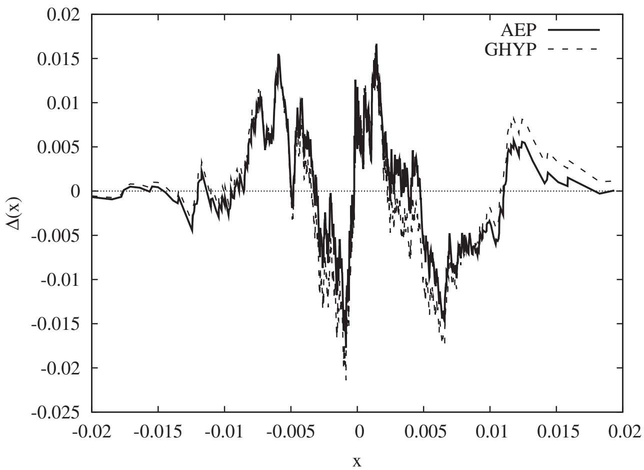

Figure 9 Deviations $\Delta(x)$ of the AEP and of the GHYP from the empirical distribution. Data are daily log first difference of the exchange rate between US Dollar and Euro.

ability to reproduce the empirical distributions with the former displaying the best results in four out of five sample considered. To add further evidence, Figure 9 reports the function $\Delta(x)$ for the exchange growth rates of US Dollar versus Euro: the difference between the two families appears, if compared with Figure 8, rather mild even if it is apparent the better capability of the AEP to fit the extreme upper tail of the empirical distribution.

5.3 *Stock markets*

As a last application, we consider daily log returns of a sample of 30 stocks, 15 from the London Stock Exchange (LSE), and 15 from the Milan Stock Exchange (MIB) chosen among the top ones in terms of capitalization and liquidity.[6]

The results of the goodness-of-fit tests performed using the D and $W2$ statistics is reported in Table 11. As can be seen, the obtained results are more ambiguous than in the previous two analyses on electricity power prices and exchange rates. While also in this case the AEP and the GHYP systematically outperform both the α-Stable and the SEP, it seems less clear how to rank them in terms of their capability to fit the empirical returns distributions. On the one hand, for the majority of the stocks, the Generalized Hyperbolic seems better in approximating the overall shape of the empirical density, as witnessed by the lower values of the $W2$ statistic. On the other hand, the highest observed deviation D is almost always lower for the AEP

[6]We use daily closing prices as retrieved from Bloomberg financial data service. The time window considered covers the period between June 1998 and June 2002.

Table 11 Maximum likelihood estimates (standard errors in parentheses) of the shape parameters of the AEP density together with the EDF goodness-of-fit statistics for four different families of distribution

	b_l	b_r	Goodness of fit — W2				Goodness of fit — D			
			AEP	GHYP	Stable	SEP	AEP	GHYP	Stable	SEP
LSE										
ARM	1.076 (0.092)	0.855 (0.063)	**0.0666**	0.0790	0.2042	0.4951	**0.0287**	0.0289	0.0392	0.0508
DXN	0.718 (0.053)	1.259 (0.096)	**0.0336**	0.0910	0.1605	0.2702	**0.0203**	0.0217	0.0374	0.0346
BG	1.110 (0.099)	0.983 (0.081)	0.0282	**0.0253**	0.1809	4.5531	**0.0214**	0.0225	0.0309	0.1173
BLT	1.315 (0.127)	0.896 (0.069)	0.0811	**0.0517**	0.0976	3.8995	**0.0224**	0.0258	0.0271	0.1190
ISY	0.714 (0.051)	1.125 (0.084)	**0.0336**	0.1666	0.2446	0.0665	**0.0237**	0.0333	0.0433	0.0247
CS	1.388 (0.137)	0.918 (0.073)	0.0652	**0.0646**	0.2244	1.6211	0.0385	**0.0379**	0.0453	0.0724
LGE	1.081 (0.092)	0.867 (0.065)	0.0714	**0.0616**	0.1896	0.0739	0.0385	0.0343	**0.0342**	0.0372
CNA	1.047 (0.089)	0.873 (0.065)	0.0589	**0.0345**	0.1680	1.8616	0.0318	**0.0305**	0.0367	0.0776
HSB	1.143 (0.105)	1.007 (0.085)	0.0544	**0.0162**	0.0864	0.3686	0.0203	**0.0168**	0.0202	0.0385
BT	1.197 (0.125)	1.328 (0.134)	**0.0354**	0.0454	0.1461	0.1509	**0.0143**	0.0179	0.0312	0.0282
TSC	1.142 (0.101)	0.895 (0.069)	0.0393	**0.0358**	0.2824	3.1644	**0.0224**	0.0258	0.0348	0.1043
SHE	1.325 (0.132)	1.188 (0.124)	0.0381	**0.0283**	0.0797	5.3933	**0.0181**	0.0184	0.0211	0.1163
BAR	1.026 (0.099)	1.447 (0.138)	**0.0201**	0.0265	0.1397	9.0418	**0.0160**	0.0174	0.0271	0.1721
BP	1.359 (0.130)	0.999 (0.089)	**0.0232**	0.0329	0.2276	4.2845	**0.0145**	0.0177	0.0341	0.1128
VOD	1.988 (0.253)	1.274 (0.158)	0.0625	**0.0511**	0.0789	0.6844	0.0215	**0.0191**	0.0271	0.0588
MIB30										
BIN	1.104 (0.096)	0.941 (0.076)	**0.0406**	0.0452	0.2742	0.2730	**0.0295**	0.0309	0.0369	0.0476
BUL	1.023 (0.092)	1.017 (0.081)	0.0802	**0.0734**	0.4221	0.1231	0.0283	**0.0275**	0.0490	0.0327
FNC	1.176 (0.119)	1.131 (0.101)	**0.0387**	0.0388	0.1364	0.0725	0.0217	**0.0181**	0.0297	0.0222
OL	0.941 (0.086)	1.354 (0.118)	**0.0394**	0.0605	0.1517	0.3213	**0.0172**	0.0208	0.0386	0.0396
ROL	0.891 (0.067)	0.841 (0.062)	0.0824	**0.0493**	0.1285	0.1381	**0.0286**	0.0294	0.0301	0.0310
SPM	1.072 (0.103)	1.211 (0.110)	0.0426	**0.0222**	0.1178	3.1962	0.0270	**0.0228**	0.0267	0.1066
UC	1.002 (0.083)	0.973 (0.079)	0.1182	**0.0616**	0.1077	0.1142	0.0371	**0.0368**	0.0393	0.0418
AUT	0.959 (0.074)	0.720 (0.047)	0.1204	**0.0941**	0.2442	12.5376	**0.0397**	0.0407	0.0467	0.1841
BPV	0.864 (0.063)	0.747 (0.051)	**0.0822**	0.1068	0.3362	0.1309	0.0344	**0.0342**	0.0491	0.0431
CAP	0.954 (0.077)	0.853 (0.062)	**0.0642**	0.0719	0.2164	1.1071	**0.0265**	0.0304	0.0467	0.0734
FI	0.891 (0.069)	0.915 (0.069)	0.0278	**0.0183**	0.1551	1.4545	**0.0161**	0.0161	0.0291	0.0731
MB	1.131 (0.100)	0.906 (0.071)	**0.0271**	0.0306	0.2008	0.0497	**0.0208**	0.0209	0.0276	0.0228
PRF	1.191 (0.107)	0.870 (0.065)	0.1571	**0.0971**	0.1570	0.7884	**0.0427**	0.0444	0.0480	0.0493
RI	1.109 (0.103)	1.024 (0.085)	0.0731	**0.0594**	0.1539	3.9919	**0.0221**	0.0222	0.0343	0.0943
STM	1.511 (0.197)	1.451 (0.170)	0.0471	**0.0391**	0.1112	0.0565	0.0162	**0.0158**	0.0243	0.0187

Data are daily log price returns. *Source*: London and Milan Stock Exchange.

Goodness of fit statistics for the best performing family are reported in boldface.

(cf. again Table 11). Anyway, one should be very cautious in ranking these two families, also because the respective values of D and $W2$ are very close to each other.

We can, however, obtain other interesting insights by analyzing in depth the unique case in which the AEP appears to performs substantially better than all the

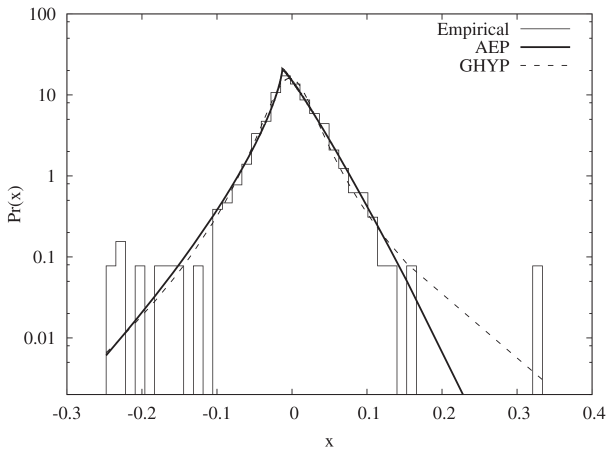

Figure 10 Empirical log return density together with the AEP and the GHYP fits. Data are daily log returns of the INVENSYS PLC stock listed at the London Stock Exchange.

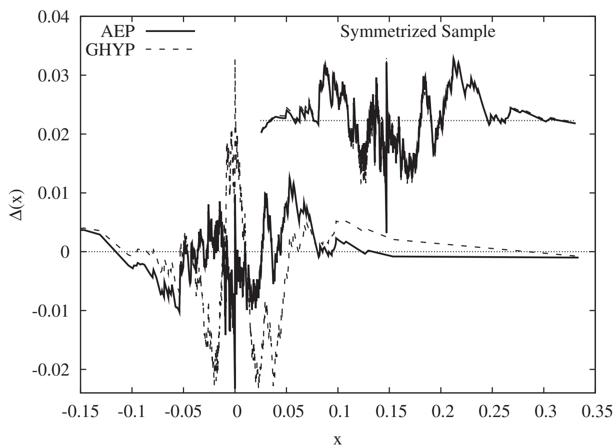

Figure 11 Deviations $\Delta(x)$ of the AEP and of the GHYP from the empirical distribution. Data are daily log returns of the INVENSYS PLC stock listed at the London Stock Exchange. $\Delta(x)$ for the symmetrized series.

other three families, GHYP included: the stock price returns of the INVENSYS PLC, a British company represented in the LSE by the abbreviation ISY. It turns out that in this case the log returns observed present two peculiar features: they display a significant degree of skewness and they include one rather anomalous observation in the upper tail, as can be seen from the empirical density displayed in Figure 10 together with the AEP (thick solid line) and GHYP (dashed line) fits. The function $\Delta(x)$ reported in Figure 11 shows that the quality of the fit provided by the GHYP is remarkably worse than the one obtained using the AEP. The impression is that the concomitant presence of a significant degree of skewness and very few anomalous observations negatively affects the ability of the GHYP to capture the observed distribution, notably worsening its fit. To further investigate this impression, we run the following experiment. From the original sample of the ISY stock returns, we removed the top 1% observations, thus inducing the original distribution to become more symmetric.[7] Then we replicate the goodness-of-fit analysis. We obtain values of both the Cramer–Von Mises and the Kolmogorov–Smirnov statistics that are very close to each other: 0.0327 and 0.0224, respectively, for the AEP and 0.0351 and 0.0186 for the GHYP. The reduction of the gap, in terms of goodness of fit, suggests that the GHYP is less robust when data are skewed and contains extreme outliers.

6. Conclusions

This paper introduces a new family of distributions, the AEP, able to cope with asymmetries and leptokurtosis and at the same time allowing for a continuous variation from non-normality to normality. We discuss the ML estimation of the AEP parameters, investigating the properties of their sampling distribution using both analytical and numerical methods.

We present a series of analytical results on the consistency, asymptotic efficiency, and asymptotic normality of the ML estimator of the AEP parameters. They are basically an extension of results previously known for the symmetric EP and prove that the estimator is consistent over the whole parameter space and that they are asymptotically efficient and normal when b_l and b_r are both greater or equal 2 (cf. Table 12 for a summary of these results). At the same time, we derive the Fisher information matrix of the AEP, showing that it is well defined in the parameter space where b_l and b_r are grater than 0.5. In this derivation, we obtain the result for the symmetric EP as a special case, fixing a mistake present in a previous work (Agró, 1995). Furthermore, we prove that a relevant part of the Fisher information matrix J can be continuously extended to the whole parameter space. Indeed, we show that even when b_l and b_r are smaller than 0.5, the upper-left 4×4 block of the inverse information matrix continues to be finite and positive definite. This suggests that the information matrix can be used to obtain theoretical asymptotic values for the estimates standard errors also when the values of the shape parameters are less than 0.5. We prove this conjecture numerically: using extensive Monte Carlo simulations we show that, first, ML estimators are always asymptotically efficient (i.e. scale with $\sqrt{N}$) even if, especially in the presence of strong asymmetries, small

[7]Coherently, the left and right estimated shape parameters of the AEP become more similar: on the symmetrized sample, b_l is found to be 1.029(0.099) while b_r is found equal to 1.085(0.089).

Table 12 Properties of the ML estimator of the AEP parameters

	Theoretical results		Numerical analysis	
	m known	*m* unknown	*m* known	*m* unknown
$b_l \geq 2,\ b_r \geq 2$	Consistent Asymp. normal Asymp. efficient	Consistent Asymp. normal Asymp. efficient	Biased[a]	Biased[a]
$0.5 < b_l < 2,\ 0.5 < b_r < 2$	Consistent Asymp. normal Asymp. efficient	Consistent *J* well defined	Biased[a]	Biased[a] Asymp. efficient
$b_l \leq 0.5,\ b_r \leq 0.5$	Consistent Asymp. normal Asymp. efficient	Consistent	Biased[a]	Biased[a] Asymp. efficient

[a]Bias contribution to RMSE is negligible for any practical application when the sample size N is greater than 100.

sample effects are present and, second, that the inverse information matrix provides accurate measures of the ML estimates also in the region of the parameter space where J is defined via analytic continuation, that is where $b_l, b_r < 0.5$. The numerical investigation of the asymptotic behavior of the ML estimator also shows that a bias is in general present, but due to its negligible contribution to the mean squared error of the estimates, it can safely be ignored for any practical purpose even when the sample size is relatively small (cf. again Table 12 for a summary of the results).

On the empirical side, our investigations provide rather strong motivations for the use of the AEP distribution for descriptive purposes. Indeed, using a selection of diverse economic and financial data, we show that the AEP performs better, in terms of its ability to approximate empirical distributions, than other commonly used families. Moreover, even in those situations in which its performance seems comparable to the one obtained with the best alternative available, namely the Generalized Hyperbolic, the AEP seems able to provide a more robust fitting framework in the presence of significant skewness and anomalous observations.

Two elements of the study of the inferential aspects of the AEP distribution are not discussed in the present contribution and still need to be investigated: the behavior of the ML estimator for small sample sizes and the characterization of the error associated with the estimate of the location parameter m when $b_l, b_r < 0.5$. We did not pursue these issues here because we consider them, from a practical point of view, of a secondary relevance. Indeed, in the large majority of applications in which the use of the AEP could result useful, one typically has at his disposal samples of several hundreds of observations and the shape parameters *b* rarely take values below 0.5.

Acknowledgement

The authors thanks Ivan Petrella, Sandro Sapio and Massimiliano Santoro for their helpful comments. Support from the Scuola Superiore Sant'Anna (grant E6006GB) and from the EU [Contract No 12410 (NEST)] is gratefully acknowledged.

References

Agró, G. (1995), 'Maximum likelihood estimation for the exponential power function parameters,' *Communications in Statistics - Simulation and Computation*, **24**, 523–536.

Alfarano, S. and M. Milakovic (2007), 'Does classical competition explain the statistical features of firm growth?' mimeo, University of Kiel.

Azzalini, A. (1986), 'Further results on a class of distributions which includes the normal ones,' *Statistica*, **12**, 171–178.

Bottazzi, G. (2004), 'Subbotools user's manual,' *LEM - Working Papers 2004–14*, Scuola Superiore Sant'Anna.

Bottazzi, G., E. Cefis, G. Dosi and A. Secchi (2007), 'Invariances and diversities in the patterns of industrial evolution: Some evidence from italian manufacturing industries,' *Small Business Economics*, **29**(1), 137–159.

Bottazzi, G. and A. Secchi (2003), 'Properties and sectoral specificities in the dynamics of U.S. manufacturing companies,' *Review of Industrial Organization*, **23**, 217–232.

Bottazzi, G. and A. Secchi (2006a), 'Explaining the distribution of firms growth rates,' *Rand Journal of Economics*, **37**, 234–263.

Bottazzi, G. and A. Secchi (2006b), 'Gibrat's law and diversification,' *Industrial and Corporate Change*, **15**(5), 847–875.

Box, G. (1953), 'a note on regions of kurtosis,' *Biometrika*, **40**, 465–468.

Canning, D., L. A. N. Amaral, Y. Lee, M. Meyer and H. E. Stanley (1998), 'Scaling the volatility of GDP growth rates,' *Economics Letters*, **60**(3), 335–341.

Capobianco, R. (2000), 'Robust aspect of the generalized normal distribution,' *Quaderni di Statistica*, **2**, 127–145.

Castaldi, C. and G. Dosi (2009), 'The patterns of output growth of firms and countries: Scale invariances and scale specificities,' *Empirical Economics*, **37**(3), 475–495.

DiCiccio, T. and A. Monti (2004), 'Inferential aspects of the skew exponential power distribution,' *Journal of the American Statistical Association*, **99**, 439–450.

Fagiolo, G., M. Napoletano and A. Roventini (2008), 'Are output growth-rate distributions fat-tailed? some evidence from oecd countries,' *Journal of Applied Econometrics*, **23**(5), 639–669.

Gradshteyn, I. and I. Ryzhyk (2000), *Table of Integrals, Series and Products*. Academic Press: New York.

Hampel, F., E. Rocchetti, P. Rousseeuw and W. Stahel (1986), *Robust Statistics: The Approach Based on the Influence Functions*. John Wiley and Sons: New York.

Huber, P. (1981), *Robust Statistics*. John Wiley and Sons: New York.

Lehmann, E. (1983), *Theory of Point Estimation*. John Wiley and Sons: New York.

Mantegna, R. and H. Stanley (2000), *An Introduction to Econophysics: Correlations and Complexity in Finance*. Cambridge University Press: Cambridge, UK.

McCauley, J. (2007), *Dynamics of Markets: Econophysics and Finance*. Cambridge University Press: New York, USA.

McNeil, A., R. Frey and P. Embrechts (2005), *Quantitative Risk Management: Concepts, Techniques, and Tools*. Princeton University Press: Princeton, USA.

Newey, W. and D. McFadden (1994), 'Large sample estimation and hypothesis testing,' in R. Engle and D. McFadden (eds), *Handbook of Econometrics*. North Holland: Amsterdam.

Nolan, J. (1998), 'Parametrizations and modes of stable distributions,' *Statistics and Probability Letters*, **38**, 187–195.

Santoro, E. (2006), 'Wax and wane: the us and uk business cycle from the cross section,' Department of Economics. *Working Papers 2006-06*, University of Trento.

Sapio, S. (2008), 'The two faces of electricity auctions. a joint analysis of price and volume growth distributions in day-ahead power exchanges,' in C. Deissemberg and G. Iori (eds), *Economic Complexity*. Elsevier Science: Oxford, UK.

Stanley, M., L. Amaral, S. Buldyrev, S. Havlin, H. Leschhorn, P. Maass, M. Salinger and H. Stanley (1996), 'Scaling behaviour in the growth of companies,' *Nature*, **379**, 804–806.

Subbotin, M. (1923), 'On the law of frequency of errors,' *Matematicheskii Sbornik*, **31**, 296–301.

Turner, M. (1960), 'On heuristic estimation methods,' *Biometrics*, **16**, 299–301.

Vianelli, S. (1963), 'La misura di variabilità condizionata in uno schema generale delle curve normali di frequenza,' *Statistica*, **23**, 447–473 [In italian].

Appendix A

Before deriving the information matrix J matrix for the AEP distribution, let us solve the following useful integral

$$I^l_{\lambda,k} = \int_{-\infty}^{m} dx f(x) \left(\frac{m-x}{a_l}\right)^{\lambda} \left(\log \frac{x-m}{a_l}\right)^{k} \quad k \in \mathbb{N},\ \lambda \in \mathbb{R}^+. \tag{A1}$$

Substituting (3) in (A1) and changing the variable to $t=\frac{1}{b_l}\left(\frac{x-m}{a_l}\right)^{b_l}$ one obtains

$$I^l_{\lambda,k}=\frac{a_l\, b_l^{\frac{\lambda+1}{b_l}-1-k}}{C}\int_0^{+\infty} dt\ e^{-t}\ t^{\frac{\lambda+1}{b_l}-1}\left(\log t+\log b_l\right)^k \tag{A2}$$

that expanding the summation becomes

$$I^l_{\lambda,k}=\frac{a_l\, b_l^{\frac{\lambda+1}{b_l}-1-k}}{C}\sum_{h=0}^{k}\binom{k}{h}\ \log^h b_l\int_0^{+\infty} dt\ \mathrm{e}^{-t}\ t^{\frac{\lambda+1}{b_l}-1}\log^{k-h} t \tag{A3}$$

and finally

$$I^l_{\lambda,k}=\frac{a_l\, b_l^{\frac{\lambda+1}{b_l}-1-k}}{C}\sum_{h=0}^{k}\binom{k}{h}\ \log^h b_l\ \Gamma^{(k-h)}\left(\frac{\lambda+1}{b_l}\right) \tag{A4}$$

where $\Gamma^{(i)}$ is the i-th derivative of the Gamma function and where we used [cf. Gradshteyn and Ryzhyk (2000) eq. 4.358] $\int_0^{+\infty} dx\ \log^n x\ x^{\nu-1}\ e^{-x}=\Gamma^{(n)}(x)$.

For instance, when $\lambda=b_l$ we get

$$I^l_{b_l,k}=\frac{a_l}{C}\ \ b_l^{\frac{1}{b_l}-k}\sum_{h=0}^{k}\binom{k}{h}\ \log^h b_l\ \Gamma^{(k-h)}\left(\frac{b_l+1}{b_l}\right)=\frac{a_l}{C}\ B_k(b_l) \tag{A5}$$

where $B_k(x)$ is defined in (11). When $\lambda=b_l-1$ one has

$$I^l_{b_l-1,k}=\frac{a_l}{C}\ \ b_l^{-k}\sum_{h=0}^{k}\binom{k}{h}\ \log^h b_l\ \Gamma^{(k-h)}(1) \tag{A6}$$

while when $\lambda=2b_l$ it is

$$I^l_{2b_l,k}=\frac{a_l}{C}\ \ b_l^{\frac{1}{b_l}+1-k}\sum_{h=0}^{k}\binom{k}{h}\ \log^h b_l\ \Gamma^{(k-h)}\left(\frac{2b_l+1}{b_l}\right). \tag{A7}$$

and when $k=0$ and $\lambda=h\in\mathbb{N}$ it is $I^l_{h,0}=\frac{a_l}{C}A_h(b_l)$ where $A_h(x)$ is defined in (4).

Correspondingly

$$I^r_{\lambda,k}=\int_m^{+\infty} dxf(x)\left(\frac{x-m}{a_r}\right)^{\lambda}\left(\log\frac{x-m}{a_r}\right)^k \tag{A8}$$

$$=\frac{a_r\, b_r^{\frac{\lambda+1}{b_r}-1-k}}{C}\sum_{h=0}^{k}\binom{k}{h}\ \log^h b_r\ \Gamma^{(k-h)}\left(\frac{\lambda+1}{b_r}\right)\qquad k\in\mathbb{N},\lambda\in\mathbb{R}^+ \tag{A9}$$

We provide below preliminary calculations needed to derive the Fisher information matrix J of $f(x;\hat{\mathbf{p}})$. They must be used in conjunction with equations (A5), (A6), (A7), and (A9) to obtain expressions in (10).

$$J_{b_l b_l} = \int_{-\infty}^{+\infty} dx\, f(x;\mathbf{p})\left(\frac{1}{C}a_l B_0'(b_l) + \left(-\frac{1}{b_l^2}\left|\frac{x-m}{a_l}\right|^{b_l} + \frac{1}{b_l}\left|\frac{x-m}{a_l}\right|^{b_l}\log\left|\frac{x-m}{a_l}\right|\right)\theta(m-x)\right)^2$$
$$= \frac{a_l}{C}B_0''(b_l) - \frac{a_l^2}{C^2}(B_0'(b_l))^2 + \frac{1}{b_l}I^l_{b_l,2} - \frac{2}{b_l^2}I^l_{b_l,1} + \frac{2}{b_l^3}I^l_{b_l,0}.$$

$$J_{b_l b_r} = \int_{-\infty}^{+\infty} dx\, f(x;\mathbf{p})\left(\frac{1}{C}a_l B_0'(b_l) + \left(-\frac{1}{b_l^2}\left|\frac{x-m}{a_l}\right|^{b_l} + \frac{1}{b_l}\left|\frac{x-m}{a_l}\right|^{b_l}\log\left|\frac{x-m}{a_l}\right|\right)\theta(m-x)\right)$$
$$\left(\frac{1}{C}a_r B_0'(b_r) + \left(-\frac{1}{b_r^2}\left|\frac{x-m}{a_r}\right|^{b_r} + \frac{1}{b_r}\left|\frac{x-m}{a_r}\right|^{b_r}\log\left|\frac{x-m}{a_r}\right|\right)\theta(x-m)\right)$$
$$= -\frac{a_l a_r}{C^2}B_0'(b_l)B_0'(b_r).$$

$$J_{b_l a_l} = \int_{-\infty}^{+\infty} dx\, f(x;\mathbf{p})\left(\frac{1}{C}a_l B_0'(b_l) + \left(-\frac{1}{b_l^2}\left|\frac{x-m}{a_l}\right|^{b_l} + \frac{1}{b_l}\left|\frac{x-m}{a_l}\right|^{b_l}\log\left|\frac{x-m}{a_l}\right|\right)\theta(m-x)\right)$$
$$\left(\frac{1}{C}B_0(b_l) - \left|\frac{x-m}{a_l}\right|^{b_l}\theta(m-x)\right) = \frac{1}{C}B_0'(b_l) - \frac{a_l}{C^2}B_0(b_l)B_0'(b_l) - \frac{1}{a_l}I^l_{b_l,1}.$$

$$J_{b_l a_r} = \int_{-\infty}^{+\infty} dx\, f(x;\mathbf{p})\left(\frac{1}{C}a_l B_0'(b_l) + \left(-\frac{1}{b_l^2}\left|\frac{x-m}{a_l}\right|^{b_l} + \frac{1}{b_l}\left|\frac{x-m}{a_l}\right|^{b_l}\log\left|\frac{x-m}{a_l}\right|\right)\theta(m-x)\right)$$
$$\left(\frac{1}{C}B_0(b_r) - \left|\frac{x-m}{a_r}\right|^{b_r}\theta(x-m)\right) = -\frac{a_l}{C^2}B_0(b_r)B_0'(b_l).$$

$$J_{b_l m} = \int_{-\infty}^{+\infty} dx\, f(x;\mathbf{p})\left(\frac{1}{C}a_l B_0'(b_l) + \left(-\frac{1}{b_l^2}\left|\frac{x-m}{a_l}\right|^{b_l} + \frac{1}{b_l}\left|\frac{x-m}{a_l}\right|^{b_l}\log\left|\frac{x-m}{a_l}\right|\right)\theta(m-x)\right)$$
$$\left(\frac{1}{a_l}\left|\frac{x-m}{a_l}\right|^{b_l-1}\theta(m-x) - \frac{1}{a_r}\left|\frac{x-m}{a_r}\right|^{b_r-1}\theta(x-m)\right) = \frac{1}{a_l}I^l_{b_l-1,1}.$$

$$J_{b_r b_r} = \int_{-\infty}^{+\infty} dx\, f(x;\mathbf{p})\left(\frac{1}{C}a_r B_0'(b_r) + \left(-\frac{1}{b_r^2}\left|\frac{x-m}{a_r}\right|^{b_r} + \frac{1}{b_r}\left|\frac{x-m}{a_r}\right|^{b_r}\log\left|\frac{x-m}{a_r}\right|\right)\theta(x-m)\right)^2$$
$$= \frac{a_r}{C}B_0''(b_r) - \frac{a_r^2}{C^2}(B_0'(b_r))^2 + + \frac{1}{b_r}I^r_{b_r,2} - \frac{2}{b_r^2}I^r_{b_r,1} + \frac{2}{b_r^3}I^r_{b_r,0}.$$

$$J_{b_r a_l} = \int_{-\infty}^{+\infty} dx\, f(x;\mathbf{p})\left(\frac{1}{C}a_r B_0'(b_r) + \left(-\frac{1}{b_r^2}\left|\frac{x-m}{a_r}\right|^{b_r} + \frac{1}{b_r}\left|\frac{x-m}{a_r}\right|^{b_r}\log\left|\frac{x-m}{a_r}\right|\right)\theta(x-m)\right)$$
$$\left(\frac{1}{C}B_0(b_r) - \left|\frac{x-m}{a_r}\right|^{b_r}\theta(x-m)\right) = -\frac{a_r}{C^2}B_0(b_l)B_0'(b_r).$$

$$J_{b_r a_r} = \int_{-\infty}^{+\infty} dx\, f(x;\mathbf{p}) \left(\frac{1}{C} a_r B_0'(b_r) + \left(-\frac{1}{b_r^2} \left| \frac{x-m}{a_r} \right|^{b_r} + \frac{1}{b_r} \left| \frac{x-m}{a_r} \right|^{b_r} \log \left| \frac{x-m}{a_r} \right| \right) \theta(x-m) \right) \left(\frac{1}{C} B_0(b_r) - \left| \frac{x-m}{a_r} \right|^{b_r} \theta(x-m) \right) = \frac{1}{C} B_0'(b_r) - \frac{a_r}{C^2} B_0(b_r) B_0'(b_r) - \frac{1}{a_r} I^r_{b_r,1}.$$

$$J_{b_r m} = \int_{-\infty}^{+\infty} dx\, f(x;\mathbf{p}) \left(\frac{1}{C} a_r B_0'(b_r) + \left(-\frac{1}{b_r^2} \left| \frac{x-m}{a_r} \right|^{b_r} + \frac{1}{b_r} \left| \frac{x-m}{a_r} \right|^{b_r} \log \left| \frac{x-m}{a_r} \right| \right) \theta(x-m) \right) \left(\frac{1}{a_l} \left| \frac{x-m}{a_l} \right|^{b_l-1} \theta(m-x) - \frac{1}{a_r} \left| \frac{x-m}{a_r} \right|^{b_r-1} \theta(x-m) \right) = -\frac{1}{a_r} I^r_{b_r-1,1}.$$

$$J_{a_l a_l} = \int_{-\infty}^{+\infty} dx\, f(x;\mathbf{p}) \left(\frac{1}{C} B_0(b_l) - \left| \frac{x-m}{a_l} \right|^{b_l} \theta(m-x) \right)^2 = -\frac{1}{C^2} B_0^2(b_l) + \frac{b_l+1}{a_l^2} I^l_{b_l,0}.$$

$$J_{a_l a_r} = \int_{-\infty}^{+\infty} dx\, f(x;\mathbf{p}) \left(\frac{1}{C} B_0(b_l) - \left| \frac{x-m}{a_l} \right|^{b_l} \theta(m-x) \right) \left(\frac{1}{C} B_0(b_r) - \left| \frac{x-m}{a_r} \right|^{b_r} \theta(x-m) \right) = -\frac{1}{C^2} B_0(b_l) B_0(b_r).$$

$$J_{a_l m} = \int_{-\infty}^{+\infty} dx\, f(x;\mathbf{p}) \left(\frac{1}{a_l} \left| \frac{x-m}{a_l} \right|^{b_l-1} \theta(m-x) - \frac{1}{a_r} \left| \frac{x-m}{a_r} \right|^{b_r-1} \theta(x-m) \right) \left(\frac{1}{C} B_0(b_l) - \left| \frac{x-m}{a_l} \right|^{b_l} \theta(m-x) \right) = -\frac{b_l}{a_l^2} I^l_{b_l-1,0}.$$

$$J_{a_r a_r} = \int_{-\infty}^{+\infty} dx\, f(x;\mathbf{p}) \left(\tfrac{1}{C} B_0(b_r) - \left| \frac{x-m}{a_r} \right|^{b_r} \theta(x-m) \right)^2 = -\frac{1}{C^2} B_0^2(b_r) + \frac{b_r+1}{a_r^2} I^r_{b_r,0}.$$

$$J_{a_r m} = \int_{-\infty}^{+\infty} dx\, f(x;\mathbf{p}) \left(\frac{1}{a_l} \left| \frac{x-m}{a_l} \right|^{b_l-1} \theta(m-x) - \frac{1}{a_r} \left| \frac{x-m}{a_r} \right|^{b_r-1} \theta(x-m) \right) \left(\frac{1}{C} B_0(b_r) - \left| \frac{x-m}{a_r} \right|^{b_r} \theta(x-m) \right) = -\frac{b_r}{a_r^2} I^r_{b_r-1,0}.$$

$$J_{mm} = \int_{-\infty}^{+\infty} dx\, f(x;\mathbf{p}) \left(\frac{1}{a_l} \left| \frac{x-m}{a_l} \right|^{b_l-1} \theta(m-x) - \frac{1}{a_r} \left| \frac{x-m}{a_r} \right|^{b_r-1} \theta(x-m) \right)^2 = \frac{1}{a_l^2} I^l_{2b_l-2,0} + \frac{1}{a_r^2} I^r_{2b_r-2,0}.$$

Appendix B

Consider a set of N observations $\{x_1, \ldots, x_N\}$ and assume that they are independently drawn from an AEP distribution of unknown parameters $\mathbf{p}_0$. According to Lehmann (1983), the ML estimates of these parameters $\hat{\mathbf{p}}$ obtained through (8) are asymptotically normal and efficient if the following four regularity conditions apply:

(A) there exists an open subset $\wp$ of $\mathbf{P}$ containing the true parameter point $\mathbf{p}_0$ such that for almost all x, the density $f_{\mathrm{AEP}}(x|\mathbf{p})$ admits all third derivatives $(\partial^3/\partial p_h \partial p_j \partial p_k) f_{\mathrm{AEP}}(x)$ for all $\mathbf{p} \in \wp$;

(B) the first and second logarithmic derivatives of f_{AEP} satisfy the equations

$$\mathrm{E}\left[\frac{\partial \log f_{\mathrm{AEP}}(x;\mathbf{p})}{\partial p_j}\right] = 0 \quad \forall j \tag{A10}$$

and

$$J_{jk}(\mathbf{p}) = H_{jk}(\mathbf{p}) \quad \forall j, k, \tag{A11}$$

where $H_{jk}(\mathbf{p}) = \mathrm{E}\left[\frac{-\partial^2 \log f_{\mathrm{AEP}}(x;\mathbf{p})}{\partial p_j \partial p_k}\right]$.

(C) the elements $J_{hj}(\mathbf{p})$ are finite and the matrix $J(\mathbf{p})$ is positive definite for all $\mathbf{p}$ in $\wp$;

(D) there exists functions M_{hjk} such that $\left|\frac{\partial^3}{\partial p_h \partial p_j \partial p_k} \log f_{\mathrm{AEP}}(x|\mathbf{p})\right| \leq M_{hjk}(x) \quad \forall p \in \wp$ where $m_{hjk} = \mathrm{E}_{\mathbf{p}_0}[M_{hjk}(x)] < \infty \ \forall h, j, k$.

Below we will prove that these four conditions are satisfied in the subset $\wp = [2, +\infty) \times [2, +\infty) \times (0, +\infty) \times (0, +\infty) \subset D$. In what follows we will denote f_{AEP} simply by f, the meaning being understood.

(1) Condition (A) is always satisfied since any derivative of f_{AEP} present, at most, a single discontinuity in correspondence of $x = m$.

(2) Since it is

$$\mathrm{E}\left[\frac{\partial \log f(x;\mathbf{p})}{\partial a_l}\right] = \int_{-\infty}^{+\infty} dx\, f(x;\mathbf{p})\left[-\frac{1}{C}B_0(b_l) + \left|\frac{x-m}{a_l}\right|^{b_l}\theta(m-x)\right]$$
$$= -\frac{1}{C}B_0(b_l) + \frac{1}{C}B_0(b_l) = 0.$$

$$\mathrm{E}\left[\frac{\partial \log f(x;\mathbf{p})}{\partial a_r}\right] = \int_{-\infty}^{+\infty} dx\, f(x;\mathbf{p})\left[-\frac{1}{C}B_0(b_r) + \left|\frac{x-m}{a_r}\right|^{b_r}\theta(x-m)\right]$$
$$= -\frac{1}{C}B_0(b_r) + \frac{1}{C}B_0(b_r) = 0.$$

$$\mathrm{E}\left[\frac{\partial \log f(x;\mathbf{p})}{\partial b_l}\right] = \int_{-\infty}^{+\infty} dx\ f(x;\mathbf{p})\left[-\frac{1}{C}a_l B_0'(b_l) + \left(\frac{1}{b_l^2}\left|\frac{x-m}{a_l}\right|^{b_l} - \frac{1}{b_l}\left|\frac{x-m}{a_l}\right|^{b_l}\log\left|\frac{x-m}{a_l}\right|\right)\right.$$
$$\left.\theta(m-x)\right] = \frac{a_l b_l^{1/b_l-2}}{C}[(\log(b_l)-1)\Gamma(1+1/b_l) + \psi(1+1/b_l)\Gamma(1+1/b_l)$$
$$+\Gamma(1+1/b_l) + -\log(b_l)\Gamma(1+1/b_l) - \psi(1+1/b_l)\Gamma(1+1/b_l)] = 0.$$

$$\mathrm{E}\left[\frac{\partial \log f(x;\mathbf{p})}{\partial b_r}\right] = \int_{-\infty}^{+\infty} dx\ f(x;\mathbf{p})\left[-\frac{1}{C}a_r B_0'(b_r) + \left(\frac{1}{b_r^2}\left|\frac{x-m}{a_r}\right|^{b_r} - \frac{1}{b_r}\left|\frac{x-m}{a_r}\right|^{b_r}\log\left|\frac{x-m}{a_r}\right|\right)\right.$$
$$\left.\theta(x-m)\right] = \frac{a_r b_r^{1/b_r-2}}{C}[(\log(b_r)-1)\Gamma(1+1/b_r) + \psi(1+1/b_r)\Gamma(1+1/b_r)$$
$$+\Gamma(1+1/b_r) - \log(b_r)\Gamma(1+1/b_r) - \psi(1+1/b_r)\Gamma(1+1/b_r)] = 0.$$

$$\mathrm{E}\left[\frac{\partial \log f(x;\mathbf{p})}{\partial m}\right] = \int_{-\infty}^{+\infty} dx\ f(x;\mathbf{p})\left[\frac{-1}{a_l}\left|\frac{x-m}{a_l}\right|^{b_l-1}\theta(m-x) + \frac{1}{a_r}\left|\frac{x-m}{a_r}\right|^{b_r-1}\theta(x-m)\right]$$
$$= \frac{-1}{C}B_0(b_l-1) + \frac{-1}{C}B_0(b_r-1) = 0.$$

the first part [Equation (28)] of Condition (B) is satisfied. Moreover it is
In order to prove (A11), notice that when $f(x;\mathbf{p})\ \partial\log f(x;\mathbf{p})/\partial p_j$ are continuous functions, this equation is a simple consequence of an integration by parts. Hence, it remains to prove (A11) only in those cases where a derivative with respect to the parameter m is involved. One has

$$H_{b_l m} = \int_{-\infty}^{+\infty} dx\ f(x)\left[\frac{1}{a_l}\left|\frac{x-m}{a_l}\right|^{b_l-1}\log\left|\frac{x-m}{a_l}\right|\theta(m-x)\right] = \frac{1}{a_l}I^l_{b_l-1,1} = J_{b_l m}$$

$$H_{b_r m} = \int_{-\infty}^{+\infty} dx\ f(x)\left[\frac{1}{a_r}\left|\frac{x-m}{a_r}\right|^{b_r-1}\log\left|\frac{x-m}{a_r}\right|\theta(x-m)\right] = -\frac{1}{a_r}I^r_{b_r-1,1} = J_{b_r m}$$

$$H_{a_l m} = -\int_{-\infty}^{+\infty} dx\ f(x)\left[\frac{b_l}{a_l^2}\left|\frac{x-m}{a_l}\right|^{b_l-1}\theta(m-x)\right] = -\frac{b_l}{a_l^2}I^l_{b_l-1,0} = J_{a_l m}$$

$$H_{a_r m} = -\int_{-\infty}^{+\infty} dx\ f(x)\left[\frac{b_r}{a_r^2}\left|\frac{x-m}{a_r}\right|^{b_r-1}\theta(x-m)\right] = -\frac{b_r}{a_r^2}I^r_{b_r-1,0} = J_{a_r m}$$

$$
\begin{aligned}
H_{mm} &= \int_{-\infty}^{+\infty} dx\, f(x) \left[\frac{b_l - 1}{a_l^2} \left| \frac{x-m}{a_l} \right|^{b_l - 2} \theta(m-x) + \frac{b_r - 1}{a_r^2} \left| \frac{x-m}{a_r} \right|^{b_r - 2} \theta(x-m) \right] \\
&= \frac{b_l - 1}{a_l^2} I^l_{b_l - 2,\, 0} + \frac{b_r - 1}{a_r^2} I^r_{b_r - 2,\, 0} = J_{mm}
\end{aligned}
$$

and (A11) is proved.

(3) According to Theorem 3.1, the matrix J exists and is positive definite for $b_l, b_r > 0.5$. When one of these two parameters moves toward the value 0.5 the element J_{mm} encounters a pole and the matrix is no longer defined.

(4) Consider the case when $p_h = p_j = p_k = m$. It is easy to show that

$$
\begin{aligned}
\frac{\partial^3}{\partial m^3} \log f(x|\mathbf{p}) = &\frac{(b_l - 1)(b_l - 2)}{a_l^3} \left| \frac{x-m}{a_l} \right|^{b_l - 3} \theta(m-x) \\
&- \frac{(b_r - 1)(b_r - 2)}{a_r^3} \left| \frac{x-m}{a_r} \right|^{b_r - 3} \theta(x-m).
\end{aligned}
\tag{A12}
$$

If one defines

$$
M_{mmm}(x) = \frac{(b_l - 1)(b_l - 2)}{a_l^3} \left| \frac{x-m}{a_l} \right|^{b_l - 3} + \frac{(b_r - 1)(b_r - 2)}{a_r^3} \left| \frac{x-m}{a_r} \right|^{b_r - 3} \tag{A13}
$$

it follows that $\left| \frac{\partial^3}{\partial m^3} \log f(x|\mathbf{p}) \right| \leq M_{mmm}(x) \quad \forall p \in \wp$. Moreover, for $b_l, b_r > 2$ it is E $[M_{mmm}] < \infty$. Using the same argument it is straightforward to prove that when $b_l, b_r > 2$ Condition (D) is satisfied also for all other cases. ■