Journal of Econometrics 148 (2009) 86–99

Contents lists available at ScienceDirect

Journal of Econometrics

journal homepage: www.elsevier.com/locate/jeconom

Properties and estimation of asymmetric exponential power distribution

Dongming Zhu [a], Victoria Zinde-Walsh [b,c,*]

[a] HSBC School of Business, Peking University, China
[b] Department of Economics, McGill University, Canada
[c] CIREQ, Canada

ARTICLE INFO

Article history:
Received 23 October 2007
Received in revised form
22 September 2008
Accepted 24 September 2008
Available online 15 October 2008

JEL classification:
C13
C16

Keywords:
Asymmetric distributions
Maximum likelihood estimation

ABSTRACT

The new distribution class, Asymmetric Exponential Power Distribution (AEPD), proposed in this paper generalizes the class of Skewed Exponential Power Distributions (SEPD) in a way that in addition to skewness introduces different decay rates of density in the left and right tails. Our parametrization provides an interpretable role for each parameter. We derive moments and moment-based measures: skewness, kurtosis, expected shortfall. It is demonstrated that a maximum entropy property holds for the AEPD distributions. We establish consistency, asymptotic normality and efficiency of the maximum likelihood estimators over a large part of the parameter space by dealing with the problems created by non-smooth likelihood function and derive explicit analytical expressions of the asymptotic covariance matrix; where the results apply to the SEPD class they enlarge on the current literature. Also we give a convenient stochastic representation of the distribution; our Monte Carlo study illustrates the theoretical results. We also provide some empirical evidence for the usefulness of employing AEPD errors in GARCH type models for predicting downside market risk of financial assets.

1. Introduction

Observed characteristics of many financial data series have motivated exploration of classes of distributions that can accommodate properties such as fat-tailedness and skewness while nesting distributions typically used in estimation such as the normal (and skew-normal). An important desired property of any such class is that it permits maximum likelihood estimation of all parameters. Obtaining closed-form expressions for the moments of interest, such as the mean, variance, skewness and kurtosis, as well as components of the information matrix provides useful interpretable features of the distributions in the class. For applications in risk management one may in addition be interested in closed-form expressions for value-at-risk and expected shortfall of asset/portfolio returns. Classes of non-symmetric distributions that nest the skew-normal were constructed by Azzalini (1986). Other classes of distributions with the desired properties of accommodating heavy tails and skewness, the Skewed Exponential Power Distribution (SEPD) classes, were proposed in Fernandez et al. (1995), Theodossiou (2000) and Komunjer (2007); they all generalize the generalized error distribution (GED) class.[1] Many financial applications of the GED as well as its skew extensions have been considered in Hsieh (1989), Nelson (1991), Duan (1999), Rachev and Mittinik (2000), Theodossiou (2000), Ayebo and Kozubowski (2004), Komunjer (2007), Christoffersen et al. (2005) and others. Especially in applications to option pricing, the GED and its skew extensions are preferred to Student-t distributions because it is found that Student-t distributions are not suited to model continuously compounded returns (see Duan (1999) and Theodossiou (2000)). Since all moments of the GED exist, the moments of exponential transformations of GED random variables, needed to price options, can be evaluated.

Ayebo and Kozubowski (2004) presented basic properties of the SEPD of Fernandez et al. (1995), derived maximum likelihood estimators of scale and skewness parameters given other parameters, and discussed its applications in finance. Komunjer (2007) explored moments (also see Theodossiou (2000)) as well as measures such as value at risk and expected shortfall useful in financial applications. DiCiccio and Monti (2004) studied properties of MLEs of the Azzalini's (1986) SEPD, and obtained results for the information matrix (not in closed form) and for inferential properties of MLE.

* Corresponding address: Department of Economics, McGill University, 855 Sherbrooke Street West, Montreal, PQ, H3A 2T7, Canada. Tel.: +1 514 398 4834.
E-mail address: victoria.zinde-walsh@mcgill.ca (V. Zinde-Walsh).

[1] The GED class was proposed first by Subbotin (1923) and Box and Tiao (1973) called such a distribution the Exponential Power Distribution (EPD). It is also called the Generalized Power Distribution or the Generalized Laplace Distribution.

0304-4076/$ – see front matter © 2008 Elsevier B.V. All rights reserved.
doi:10.1016/j.jeconom.2008.09.038

However, for some applications in finance and risk management, the skew extensions may not be rich enough to capture all the asymmetry of distributions of asset returns, particularly asymmetry in the tails. For example, it is found especially for portfolios such as *S&P500* and *NASDAQ* that *ex post* innovations from estimated *GARCH* models (even with a leverage effect) are not normally distributed—the QQ plot of *ex post* innovations typically shows that the fit in the upper tail is good but the lower tail is heavier than that of the normal distribution (see Figure 6 of Bradley and Taqqu (2003) for *NASDAQ*, Figure 4.2 of Christoffersen (2003) for *S&P500*). To capture the asymmetry in the tails, this paper extends the SEPD to a fully asymmetric exponential power distribution (AEPD) where heavy-tailedness itself may be asymmetric with different tail exponents on different sides of the distribution.

We demonstrate that the AEPD class has desired properties: interpretable parameters to represent location, scale, and shape, closed-form expressions for the moments as well as for value at risk and expected shortfall. A maximum entropy property is shown to hold and a stochastic representation of the AEPD is given. We develop asymptotic properties of the MLE (consistency and asymptotic normality) and obtain fully closed-form expressions for the information matrix for all parameters. Thus we also provide new theoretical results such as closed-form expressions for the asymptotic covariance matrix and consistency and asymptotic normality of MLE for some SEPD classes, expanding on results currently available in the literature. Comparing the AEPD with Azzalini's (1986) SEPD class, both classes have continuous but non-differentiable densities; the latter density however involves an integral (normal cdf). Also, the AEPD has more flexible tail behavior and analytical expressions for mode and moments; while for Azzalini's (1986) SEPD, the left tail is always thinner than the right one, its odd moments involve infinite series expansions, and it is not possible to find an analytic expression for the mode. In addition, note that DiCiccio and Monti (2004) were not able to provide closed form expressions for the information matrix (nor complete proofs of asymptotics for MLE) for Azzalini's (1986) SEPD. In addition, we provide some empirical evidence for the usefulness of employing AEPD errors in GARCH type models for predicting the value at risk of financial assets.

The paper is organized as follows. Section 2 explains the relation between EPD, SEPD and AEPD classes highlighting the main features of the new AEPD class. The interpretation of parameters is provided in Section 3. Section 4 gives basic properties of the AEPD such as analytical expressions of cdf, quantiles, moments and expected shortfall. In Section 5 we establish consistency and asymptotic normality of the MLE and Section 6 provides some finite sample Monte Carlo results. We also provide an application of the AEPD in Section 7. Technical results and proofs are collected in the Appendices A–C; more detailed proofs are in Zhu and Zinde-Walsh (2007). Appendix D shows graphs of AEPD densities for different parameter combinations.

2. The relation between EPD, SEPD and AEPD

The density function of the EPD (or GED) is usually defined as:

$$f_{EP}(x \mid p, \mu, \sigma) = \frac{1}{\sigma} K_{EP}(p) \exp\left(-\frac{1}{p}\left|\frac{x-\mu}{\sigma}\right|^{p}\right), \tag{1}$$

where $\mu \in R$ and $\sigma > 0$ are the location and scale parameters respectively, $p > 0$ is the shape parameter, and $K_{EP}(p)$ is the normalizing constant, $K_{EP}(p) \equiv 1/[2p^{1/p}\Gamma(1+1/p)]$. If X is a random variable with the EPD density, then the location parameter $\mu = E(X) = med(X)$, the median of X; the scale parameter $\sigma = (E|X-\mu|^{p})^{1/p}$, which is the L_p-norm deviation, has an interpretation similar to that of the standard deviation of the normal distribution. When the shape parameter p gets smaller and smaller, the EPD becomes more and more heavy-tailed and leptokurtic. With $p = 2$, $p = 1$, and $p \to +\infty$, the EPD reduces to the normal, Laplace and uniform distributions, respectively.

So far, there are two different methods to extend the EPD to a skewed exponential power distribution (SEPD). Azzalini (1986) first proposed a family of SEPD based on the fact that if $f(\cdot)$ is a density symmetric about 0 and $\Pi(\cdot)$ an absolutely continuous distribution function such that its pdf $\Pi'(\cdot)$ is symmetric about 0, then $2\Pi(\lambda x)f(x)$ is a density for any real λ. Taking $f = f_{EP}$ and $\Pi =$ normal cdf or EPD's cdf, we get Azzalini's SEPD class. Fernandez et al. (1995) extended the EPD class to *another* family of SEPD by using a two-piece method, in which an additional skew parameter γ is introduced (also see Kotz et al. (2001), p 271).

By a method similar to that of Fernandez et al. (1995), Theodossiou (2000) and Komunjer (2007), respectively, constructed seemingly different classes of SEPD, which are actually reparametrizations of that of Fernandez et al. (1995).[2] However, Komunjer's (2007) asymmetry parameter α is interestingly interpreted as the probability such that the location parameter is exactly the α-quantile of the SEPD. Noting the interpretable nature of the parameters this paper follows a similar method to construct the AEPD.

The AEPD density has the following form:

$$f_{AEP}(x \mid \beta) = \begin{cases} \left(\frac{\alpha}{\alpha^*}\right)\frac{1}{\sigma}K_{EP}(p_1)\exp\left(-\frac{1}{p_1}\left|\frac{x-\mu}{2\alpha^*\sigma}\right|^{p_1}\right), & \text{if } x \le \mu; \\ \left(\frac{1-\alpha}{1-\alpha^*}\right)\frac{1}{\sigma}K_{EP}(p_2)\exp\left(-\frac{1}{p_2}\left|\frac{x-\mu}{2(1-\alpha^*)\sigma}\right|^{p_2}\right), & \text{if } x > \mu, \end{cases} \tag{2}$$

where $\beta = (\alpha, p_1, p_2, \mu, \sigma)^{\mathrm{T}}$ is the parameter vector, $\mu \in R$ and $\sigma > 0$ still represent location and scale, respectively, $\alpha \in (0, 1)$ is the skewness parameter, $p_1 > 0$ and $p_2 > 0$ are the left and right tail parameters, respectively, $K_{EP}(p)$ is the same as in (1), and α^* is defined as

$$\alpha^* = \alpha K_{EP}(p_1)/[\alpha K_{EP}(p_1) + (1-\alpha)K_{EP}(p_2)]. \tag{3}$$

Note that

$$\frac{\alpha}{\alpha^*}K_{EP}(p_1) = \frac{1-\alpha}{1-\alpha^*}K_{EP}(p_2) = \alpha K_{EP}(p_1) + (1-\alpha)K_{EP}(p_2) \equiv B. \tag{4}$$

The AEPD density function is still continuous at every point and unimodal with mode at μ. The parameter α^* in the AEPD density provides scale adjustments respectively to the left and right parts of the density so as to ensure continuity of the density under changes of shape parameters (α, p_1, p_2). If $p_1 = p_2 = p$, implying $\alpha^* = \alpha$, the AEPD reduce to a new version of SEPD:

$$f_{SEP}(x \mid \beta) = \begin{cases} \frac{1}{\sigma}K_{EP}(p)\exp\left(-\frac{1}{p}\left|\frac{x-\mu}{2\alpha\sigma}\right|^{p}\right), & \text{if } x \le \mu; \\ \frac{1}{\sigma}K_{EP}(p)\exp\left(-\frac{1}{p}\left|\frac{x-\mu}{2(1-\alpha)\sigma}\right|^{p}\right), & \text{if } x > \mu, \end{cases} \tag{5}$$

which is equivalent to those of Fernandez et al. (1995), Theodossiou (2000) and Komunjer (2007). This new version of SEPD provides new interesting interpretations for scale and skewness in terms of L_p distances. The skewness parameter $\alpha \in (0, 1)$ plays the same role as the parameter γ of Fernandez et al. (1995). By reparametrization, $\alpha = \gamma^2/(1+\gamma^2)$ and $\sigma = (2/p)^{1/p}(\gamma + 1/\gamma)\sigma'/2$, the SEPD (5) will become that of Fernandez et al. (1995);

[2] A referee pointed out various other references (e.g. Arellano-Valle et al. (2005) and Salinas et al. (2007)) that generalize the class of asymmetric models beyond power distribution classes, however, none of those classes considers asymmetry in the tails of the distribution.

a re-scaling of the density leads to Komunjer's (2007); letting $\alpha = (1+\lambda)/2$, $\sigma = \theta\sigma' p^{-1/p}$ and $\mu = \mu' - \delta\sigma'$, the density will be that (i.e., $f(y \mid \mu', \sigma', p, \lambda)$ in Equation 10) of Theodossiou (2000), where θ and δ are given in Equations 12 and 13 of Theodossiou (2000). With $\alpha = 1/2$, the SEPD (5) reduces to the EPD (1). The skewed Laplace distribution and skewed normal distribution are special cases of the SEPD, respectively, with $p = 1$ and $p = 2$. According to the skewness measure of Arnold and Groeneveld (1995), 1-2CDF(mode), the SEPD density is skewed to the right for $\alpha < 1/2$ and to the left for $\alpha > 1/2$.

A convenient reparametrization of (2) is obtained by rescaling,

$$f_{AEP}(x \mid \theta) = \begin{cases} \dfrac{1}{\sigma}\exp\left(-\dfrac{1}{p_1}\left|\dfrac{x-\mu}{2\alpha\sigma K_{EP}(p_1)}\right|^{p_1}\right), & \text{if } x \le \mu; \\ \dfrac{1}{\sigma}\exp\left(-\dfrac{1}{p_2}\left|\dfrac{x-\mu}{2(1-\alpha)\sigma K_{EP}(p_2)}\right|^{p_2}\right), & \text{if } x > \mu, \end{cases} \quad (6)$$

where $\theta = (\alpha, p_1, p_2, \mu, \sigma)^{\mathrm{T}}$. From the rescaled AEPD density (6), we can clearly observe the effects of the shape parameters on the distribution. The density in the form (6) is used in deriving a closed form expression for the information matrix of the maximum likelihood estimator (MLE).

3. Interpretation of parameters of AEPD

The main tools that are used for interpretation are various L_r space related distance measures. Define for $r > 0$,

$$d_L(r) \equiv [E(|X-\mu|^r \mid X \le \mu)]^{1/r},$$
$$d_R(r) \equiv [E(|X-\mu|^r \mid X > \mu)]^{1/r},$$

respectively called the L_r-*norm* deviation (or distance) conditional on $X \le \mu$ and the L_r-*norm* deviation conditional on $X > \mu$. The total conditional deviation (or distance) is $d(r) \equiv d_L(r) + d_R(r)$; the L_r-norm deviation $\|X-\mu\|_r = (E\,|X-\mu|^r)^{1/r}$.

Suppose now that random variable X has the AEPD density defined in (2) with shape parameters (α, p_1, p_2), location μ and scale σ.

Proposition 1. *The following relations hold:*

(a) $P(X \le \mu) = \alpha$; *also* $d_L(p_1) = 2\alpha^*\sigma$, *and* $d_R(p_2) = 2(1-\alpha^*)\sigma$, *where* α^* *is defined in* (3)*;*

(b) $\sigma = \frac{1}{2}[d_L(p_1) + d_R(p_2)]$;

(c) *there is a positive function* $r^*(c \mid p)$ *depending on parameter* p *and increasing in its argument,* c*, such that*

$$\alpha = \frac{d_L(r^*(c \mid p_1))}{d_L(r^*(c \mid p_1)) + d_R(r^*(c \mid p_2))}; \quad \forall c > \max\{lb(p_1), lb(p_2)\}$$

where $lb(p) \equiv [2\Gamma(1+1/p)]^{-1}\exp\{\frac{1}{p}\Psi(1/p)\}$ *and* $\Psi(x) \equiv \Gamma'(x)/\Gamma(x)$ *is a digamma function.*

(d) $d_L(r) = 2\alpha^*\sigma M(p_1, r) = 2\alpha\sigma\xi(p_1, r)/B$ *and* $d_R(r) = 2(1-\alpha^*)\sigma M(p_2, r) = 2(1-\alpha)\sigma\xi(p_2, r)/B$, *where* $M(p, r) \equiv p^{1/p}\{\Gamma((r+1)/p)/\Gamma(1/p)\}^{1/r}$; $\xi(p, r) \equiv K_{EP}(p)M(p, r)$ *is strictly increasing in* r *and decreasing in* p, B *and* $K_{EP}(p)$ *are defined above.*

Proof. See Appendix A. ■

From part (a) the location μ is the α-quantile of the r.v. X and the scale σ is related to the left and right L_p conditional deviations by the parameter α^* (for the SEPD $\alpha^* = \alpha$). Part (b) represents the scale σ via an average of the left and right conditional deviations. It follows from (a) that the ratio of the left conditional deviation to the total is α^* (for SEPD just α). Part (c) gives an interpretation of α with two adjusted order functions $r^*(c \mid p_i)$, $i = 1, 2$; in the SEPD case ($p_1 = p_2$) the left and right conditional deviations enter with a same order thus then $\alpha = d_L(r)/d(r)$ for any $r > 0$.

Part (d) allows us to investigate the effect of shape parameters a, p_1, p_2. These shape parameters have a common effect on both $d_L(r)$ and $d_R(r)$ through $B = \alpha K_{EP}(p_1) + (1-\alpha)K_{EP}(p_2)$, which represents a scale adjustment effect. Ignoring the common effect, α has the same effect on the AEPD as it does on the SEPD, but the left and right tail parameters, $p_1$and p_2, respectively control the left and right L_r-deviation, $d_L(r)$ and $d_R(r)$. Since $\xi(p, r)$ is a strictly decreasing function of p for any given r, a smaller p_1 (or p_2) leads to a larger left (or right) L_r-deviation, thus AEPD with a smaller p_1 (or p_2) has a heavier left (or right) tail.

The effect of p_1 (or p_2) on the left (or right) tail can be measured by a generalized kurtosis index $kur_L(r)$ (or $kur_R(r)$) for $r > 0$, called the left (or right) generalized kurtosis (similar to Mineo (1989) who defined generalized kurtosis as $\frac{E|X-\mu|^{2p}}{(E|X-\mu|^p)^2}$ and showed that for EPD it is $p+1$). The left and right (generalized) kurtoses are defined as

$$kur_L(r) \equiv [d_L(2r)/d_L(r)]^{2r}, \qquad kur_R(r) \equiv [d_R(2r)/d_R(r)]^{2r}.$$

With $r = 2$ we get the usual definition of kurtosis.

Proposition 2. *For the AEPD the left and right (generalized) kurtosis can be expressed as follows:*

$$kur_L(r) = \Gamma\left(\frac{1}{p_1}\right)\Gamma\left(\frac{2r+1}{p_1}\right)\bigg/\Gamma^2\left(\frac{r+1}{p_1}\right), \quad (7)$$

$$kur_R(r) = \Gamma\left(\frac{1}{p_2}\right)\Gamma\left(\frac{2r+1}{p_2}\right)\bigg/\Gamma^2\left(\frac{r+1}{p_2}\right); \quad (8)$$

they are strictly decreasing respectively in p_1 *and* p_2 *for any given* $r > 0$*, and strictly increasing in* r *given* $p_1, p_2 > 0$.

Proof. See Appendix A. ■

From the expressions for $kur_L(r)$ and $kur_R(r)$ of the AEPD, the heaviness of the left (or right) tail is controlled by only p_1 (or p_2). If $p_1 < p_2$, then $kur_L(r) > kur_R(r)$, implying that the left tail is heavier than the right. When $p_i < 2$ $(i = 1, 2)$, the AEPD is more heavy-tailed than the normal distribution. The left (or right) tail parameter p_1 (or p_2) is directly related to the left (or right) generalized kurtosis by the relation: $kur_L(p_1) = p_1 + 1$ (or $kur_R(p_2) = p_2 + 1$). Further results about kurtosis via moments are in the next section.

Fig. 1 in Appendix D plots the AEPD densities of the form (6) with $\mu = 0$ and $\sigma = 1$ for combinations of shape parameters (α, p_1, p_2). The first plot shows that for given p_1 and p_2 the density curve shifts to the right with α decreasing but its mode does not change; the second plot shows how p_2 controls only the right tail — heavier and heavier for smaller and smaller p_2. The effect of skewness parameter and tail parameters on tails is compared in the last plot. Although a smaller α leads to a fatter right tail, this influence eventually is dominated by the effect of a smaller p_2.

4. Basic properties of the AEPD

4.1. Cumulative distribution, quantile function and moments

All the formulae in this section follow straightforwardly from results for the classical EPD (summarized in (III) in Appendix A).

Suppose that X is a random variable with the standard AEPD density ($\mu = 0$, $\sigma = 1$). Denote $a \wedge b \equiv \min\{a, b\}$, $a \vee b \equiv \max\{a, b\}$, by $G(x; \gamma)$ the gamma cdf:

$$G(x; \gamma) \equiv (\Gamma(\gamma))^{-1}\int_0^x z^{\gamma-1}\exp(-z)\mathrm{d}z, \quad (9)$$

and by $G^{-1}(x;\gamma)$ the inverse function of $G(x;\gamma)$. Then for the standard AEPD, the cdf can be expressed via $G(\cdot;\cdot)$:

$$F_{AEP}(x \mid \alpha, p_1, p_2) = \begin{cases} \alpha\left[1 - G\left(\frac{1}{p_1}\left(\frac{|x|}{2\alpha^*}\right)^{p_1}; \frac{1}{p_1}\right)\right], & \text{if } x \le 0 \\ \alpha + (1-\alpha)G\left(\frac{1}{p_2}\left(\frac{|x|}{2(1-\alpha^*)}\right)^{p_2}; \frac{1}{p_2}\right), & \text{if } x > 0 \end{cases} \quad (10)$$

and the quantile function is expressed via $G^{-1}(\cdot;\cdot)$

$$F_{AEP}^{-1}(\upsilon \mid \alpha, p_1, p_2) = \begin{cases} -2\alpha^*\left[p_1 G^{-1}\left(1 - \frac{\upsilon}{\alpha}; \frac{1}{p_1}\right)\right]^{1/p_1}, & \text{if } \upsilon \le \alpha \\ 2(1-\alpha^*)\left[p_2 G^{-1}\left(1 - \frac{1-\upsilon}{1-\alpha}; \frac{1}{p_2}\right)\right]^{1/p_2}, & \text{if } \upsilon > \alpha \end{cases} \quad (11)$$

where $\upsilon \in [0, 1]$.

Note that, for any measurable function $h(X)$ of the standard AEPD random variable X, we have

$$E[h(X)] = \alpha E[h(X) \mid X \le 0] + (1-\alpha)E[h(X) \mid X > 0],$$

implying that all unconditional moments can be expressed as a weighted sum of two conditional moments. Therefore, we first give the conditional moments of the standard AEPD r.v. X. From expression for the absolute moment of EPD (42), we get for any real $r > -1$,

$$E(|X|^r \mid X < 0) = [2\alpha^*]^r E\left|Z_{p_1}\right|^r = B^{-r}\alpha^r H_r(p_1), \quad (12)$$

$$E(|X|^r \mid X > 0) = [2(1-\alpha^*)]^r E\left|Z_{p_2}\right|^r = B^{-r}(1-\alpha)^r H_r(p_2) \quad (13)$$

where Z_p is a random variable that has the standard EPD density ($\mu = 0, \sigma_p = 1$ in (1)) with power index p, B is defined in (4), $H_r(p) \equiv p^r\Gamma(\frac{1+r}{p})/\Gamma^{1+r}(\frac{1}{p})$. For any non-negative integer k, the kth right-conditional moment, $E(X^k \mid X > 0)$, has the same expression as in (13), while the kth left-conditional moment, $E(X^k \mid X < 0)$, has an expression slightly different from (12)[3]:

$$E(X^k \mid X < 0) = [-2\alpha^*]^k E\left|Z_{p_1}\right|^k = B^{-k}(-\alpha)^k H_k(p_1).$$

Thus, the kth moment of the standard AEPD r.v. X equals

$$E(X^k) = B^{-k}[(-1)^k\alpha^{1+k}H_k(p_1) + (1-\alpha)^{1+k}H_k(p_2)], \quad k = 1, 2, 3, \ldots, \quad (14)$$

and its r-absolute moment is expressed as

$$E(|X|^r) = B^{-r}[\alpha^{1+r}H_r(p_1) + (1-\alpha)^{1+r}H_r(p_2)], \quad r > -1. \quad (15)$$

In particular, the mean and variance of the standard AEPD r.v. X are given as follows:

$$E(X) = \frac{1}{B}\left[(1-\alpha)^2\frac{p_2\Gamma(2/p_2)}{\Gamma^2(1/p_2)} - \alpha^2\frac{p_1\Gamma(2/p_1)}{\Gamma^2(1/p_1)}\right], \quad (16)$$

$$\mathrm{Var}(X) = \frac{1}{B^2}\left\{(1-\alpha)^3\frac{p_2^2\Gamma(3/p_2)}{\Gamma^3(1/p_2)} + \alpha^3\frac{p_1^2\Gamma(3/p_1)}{\Gamma^3(1/p_1)} - \left[(1-\alpha)^2\frac{p_2\Gamma(2/p_2)}{\Gamma^2(1/p_2)} - \alpha^2\frac{p_1\Gamma(2/p_1)}{\Gamma^2(1/p_1)}\right]^2\right\}. \quad (17)$$

[3] Note that these moments represent lower partial moments which are used in finance literature as risk measures, see Bawa (1975).

We see that all moments can be expressed simply and conveniently in terms of gamma function. In the case of the SEPD $p_1 = p_2 = p$ and we get simplified expressions for moments:

$$E(X^k) = (2p^{1/p})^k[(-1)^k\alpha^{1+k} + (1-\alpha)^{1+k}] \times \Gamma((1+k)/p)/\Gamma(1/p), \quad (18)$$

$$E(|X|^r) = (2p^{1/p})^r[\alpha^{1+r} + (1-\alpha)^{1+r}] \times \Gamma((1+r)/p)/\Gamma(1/p), \quad (19)$$

where $k = 1, 2, \ldots,$ and $r > -1$. These provide an advantage over Azzalini's (1986) SEPD class where the expressions for the odd moments involve infinite series expansions; (18) is a reparametrization of formulae of Fernandez et al. (1995) and Komunjer (2007).

4.2. Value at risk and expected shortfall

Value at risk (VaR) for return on a portfolio or an asset is defined as the υ-quantile of the distribution of returns with a negative value corresponding to a loss. Here the quantile function $F_{AEP}^{-1}(\upsilon \mid \alpha, p_1, p_2)$ of (11) provides VaR at υ for the historical distribution of returns in the AEPD class, i.e., $VaR_{AEP}(\upsilon) \equiv F_{AEP}^{-1}(\upsilon \mid \alpha, p_1, p_2)$. The Expected Shortfall (ES) of a standard AEPD random variable X,

$$ES_{AEP}(q) \equiv E(-X \mid X < q),$$

also called Conditional Value at Risk (CVaR) represents the negative expected return (or loss) conditional on it being below the threshold q. It can be expressed in terms of the gamma CDFs with parameters $1/p_1$, $2/p_1$, $1/p_2$, and $2/p_2$:

$$ES_{AEP}(q) = \begin{cases} 2\alpha^* C(p_1)\left[\dfrac{1 - G\left(\frac{1}{p_1}\left|\frac{q}{2\alpha^*}\right|^{p_1}; \ 2/p_1\right)}{1 - G\left(\frac{1}{p_1}\left|\frac{q}{2\alpha^*}\right|^{p_1}; \ 1/p_1\right)}\right], & q \le 0; \\ \dfrac{2\alpha\alpha^* C(p_1) - 2(1-\alpha)(1-\alpha^*)C(p_2)G\left(\frac{1}{p_2}(\frac{|q|}{2(1-\alpha^*)})^{p_2}; \ 2/p_2\right)}{\alpha + (1-\alpha)G\left(\frac{1}{p_2}(\frac{|q|}{2(1-\alpha^*)})^{p_2}; 1/p_2\right)}, & q > 0, \end{cases} \quad (20)$$

where $C(p) \equiv p^{1/p}\Gamma(2/p)/\Gamma(1/p)$, $G(x;\gamma)$ is the gamma cdf given in (9). Recall that $G^{-1}(x;\gamma)$ is the inverse function of $G(x;\gamma)$. For $q = VaR_{AEP}(\upsilon)$, the ES as a function of confidence level υ, denoted by $ES_{AEP}^*(\upsilon)$, can be expressed as follows:

$$ES_{AEP}^*(\upsilon) = \begin{cases} \frac{2}{\upsilon}\alpha\alpha^* C(p_1)\left\{1 - G\left[G^{-1}\left(\frac{\alpha-\upsilon}{\alpha}; \frac{1}{p_1}\right); 2/p_1\right]\right\}, & \upsilon \le \alpha, \\ \frac{2}{\upsilon}\left\{\alpha\alpha^* C(p_1) - (1-\alpha)(1-\alpha^*)C(p_2) \times G\left[G^{-1}\left(\frac{\upsilon-\alpha}{1-\alpha}; \frac{1}{p_2}\right); 2/p_2\right]\right\}, & \upsilon > \alpha. \end{cases}$$

In practice ES is often used in the following form:

$$E(q - X \mid X < q) = q + E(-X \mid X < q), \quad (21)$$

which is the average loss when an asset return falls below q; the expression follows from $ES_{AEP}(q)$ or $ES_{AEP}^*(\upsilon)$.

4.3. Maximum entropy property

In a distribution class maximum entropy is achieved by a distribution that encodes information in the least biased way without giving any preferential measure weight to any part of the distribution (other than what is required by the distribution class itself). Here we consider a class of absolutely continuous distributions with specific shape (moment) constraints on the left and right deviations and show that the AEPD as defined in (6) has the maximum entropy property in that class.

Specifically consider for parameters $\theta = (\alpha, p_1, p_2, \mu, \sigma)$ an absolute deviation function of $x \in R$ scaled differently on two sides of $\mu : y(x) = L(x;\theta) + R(x;\theta)$, with

$$L(x;\theta) = \frac{\Gamma(1+1/p_1)\,|x-\mu|}{\alpha\sigma} 1(x<\mu),$$
$$R(x;\theta) = \frac{\Gamma(1+1/p_2)\,|x-\mu|}{(1-\alpha)\sigma} 1(x>\mu).$$

Define a class $\Omega(\alpha, p_1, p_2, \mu, \sigma)$ of absolutely continuous distributions having densities $p(x)$ with support $(-\infty, +\infty)$ that satisfy the following moment constraints on the left and right deviations of $y(x)$:

$$\|L(x;\theta)\|_{p_1} = \left(\int y(x)^{p_1} 1(x<\mu) p(x) \mathrm{d}x\right)^{1/p_1} = \left(\frac{\alpha}{p_1}\right)^{1/p_1};$$
$$\|R(x;\theta)\|_{p_2} = \left(\int y(x)^{p_2} 1(x>\mu) p(x) \mathrm{d}x\right)^{1/p_2} = \left(\frac{1-\alpha}{p_2}\right)^{1/p_2}.$$

This class allows for the location μ, scale σ and three shape parameters α, p_1, p_2 that produce different effects: when $p_1 = p_2$ parameter α alone governs which of the sides gets a larger weight, when p_1, p_2 differ the smaller imparts a heavier tail to its side regardless of a. Thus such a class for fixed values of the parameters gives rise to distributions that could fit required properties for shape in terms of the left/right conditional deviations.

Proposition 3. *The AEPD distribution in (6) has maximum entropy in the class $\Omega(\alpha, p_1, p_2, \mu, \sigma)$.*

Proof. See Appendix A. ■

5. Asymptotic properties of the maximum likelihood estimator

Since AEPD generalizes the EPD and SEPD classes, we note the asymptotic results available for the latter two classes. The MLE for the EPD parameters and its properties are investigated in Agrò (1995) where the information matrix $I(\beta)$ and the covariance matrix are derived; for $p > 2$ consistency, asymptotic normality and efficiency of MLE are proved; other theoretical results for the MLE are available when p is known. Ayebo and Kozubowski (2004) focused on estimators of scale σ and skewness parameter α in the SEPD by assuming that location μ and tail parameter p are known; they gave the expressions for the MLEs of σ and α, showed that they are consistent, asymptotically normal and efficient and provided the asymptotic covariance matrix for this subset of parameters. DiCiccio and Monti (2004) investigated properties of the MLE of all parameters for the Azzalini's SEPD class, but they did not give a closed-form expression for information matrix and did not provide a rigorous proof of asymptotics for the MLEs which is needed due to the non-smoothness of the log-likelihood function. Here we establish consistency, asymptotic normality and efficiency for MLE of all parameters in the AEPD class (which nests EPD and SEPD) with $p_1 > 1$ and $p_2 > 1$, and provide a closed-form asymptotic covariance matrix of the MLE.

Suppose that the true density $f(y \mid \theta_0)$ with $\theta_0 = (\alpha_0, p_{01}, p_{02}, \mu_0, \sigma_0)$ belongs to the AEPD class (given in (6)) with parameter vector θ in a parameter space $\Theta \subset \Xi \equiv \{\theta \mid \theta = (\alpha, p_1, p_2, \mu, \sigma), \sigma, p_1, p_2 > 0, \alpha \in (0,1), \mu \in R\}$, where Θ is assumed to be a compact set and θ_0 to be an interior point of Θ. For a random sample $y = (y_1, y_2, \ldots, y_T)$, the log-likelihood function $l_T(\theta \mid y) \equiv \sum_{t=1}^{T} \ln f(y_t \mid \theta)$ is given as follows:

$$l_T(\theta \mid y) = -T\ln\sigma - \sum_{t=1}^{T}\left(\frac{\Gamma(1+1/p_1)(\mu-y_t)}{\alpha\sigma}\right)^{p_1} 1(y_t \le \mu) - \sum_{t=1}^{T}\left(\frac{\Gamma(1+1/p_2)(y_t-\mu)}{(1-\alpha)\sigma}\right)^{p_2} 1(y_t > \mu).$$

Note that the AEPD does not satisfy the regularity conditions under which the ML estimator has the usual $\sqrt{T}$-asymptotics, because of a non-differentiable likelihood function. However, we nonetheless establish consistency of the MLE by using Theorem 2.5 in Newey and McFadden (1994) and under certain parameter restrictions establish the usual asymptotic normality for the AEPD's MLE by using Theorem 3 as well as its corollary in Huber (1967).

Proposition 4. *(Consistency of MLE). The MLE $\widehat{\theta}$ of θ_0 is consistent, i.e., $\widehat{\theta} \to^p \theta_0$.*

Proof. See Appendix C. ■

Proposition 5. *The information matrix equality $I(\theta_0) = -H(\theta_0)$ holds for $p_{01} > 1/2$, $p_{02} > 1/2$. The elements of the Fisher information matrix, ϕ_{ij},*

$$\phi_{ij} \equiv E[\partial \ln f(y_t;\theta_0)/\partial\theta_i]\cdot[\partial \ln f(y_t;\theta_0)/\partial\theta_j], \tag{22}$$

with $\phi_{ij} = \phi_{ji}$ and θ_j the jth element of parameter vector $\theta = (\alpha, p_1, p_2, \mu, \sigma)^{\mathrm{T}}$, are as follows[4]*:*

$$\begin{aligned}
&\phi_{11} = \frac{p_1+1}{\alpha} + \frac{p_2+1}{1-\alpha}, \quad \phi_{12} = -\frac{1}{p_1}, \quad \phi_{13} = \frac{1}{p_2},\\
&\phi_{14} = -\frac{1}{\sigma}\left(\frac{p_1}{\alpha} + \frac{p_2}{1-\alpha}\right),\\
&\phi_{15} = \frac{p_1-p_2}{\sigma}, \quad \phi_{22} = \frac{\alpha}{p_1^3}(1+1/p_1)\Psi'(1+1/p_1),\\
&\phi_{23} = 0, \quad \phi_{25} = -\frac{\alpha}{\sigma p_1},\\
&\phi_{24} = \frac{1}{\sigma p_1}\left[\Psi(2) - \Psi(1+1/p_1)\right],\\
&\phi_{35} = -\frac{1-\alpha}{\sigma p_2}, \quad \phi_{55} = \frac{\alpha p_1 + (1-\alpha)p_2}{\sigma^2},\\
&\phi_{33} = \frac{1-\alpha}{p_2^3}(1+1/p_2)\Psi'(1+1/p_2),\\
&\phi_{34} = -\frac{1}{\sigma p_2}\left[\Psi(2) - \Psi(1+1/p_2)\right],\\
&\phi_{44} = \frac{\Gamma(1/p_1)\Gamma(2-1/p_1)}{\alpha\sigma^2} + \frac{\Gamma(1/p_2)\Gamma(2-1/p_2)}{(1-\alpha)\sigma^2},\\
&\phi_{45} = \frac{1}{\sigma^2}(p_2-p_1),
\end{aligned} \tag{23}$$

where all the ϕ_{ij} are evaluated at the true values $(\alpha_0, p_{01}, p_{02}, \mu_0, \sigma_0)$.

Proof. See Appendix B. ■

The information matrix for the MLE of the SEPD is given below; to our knowledge these results were not available in the literature so far.

[4] By using $\Gamma(x)\Gamma(1-x) = \pi/\sin(\pi x)$ (see Artin (1964, p 26), or Farrell and Ross (1963, p 39)), the element of ϕ_{44} can also be expressed as

$$\phi_{44} = \frac{\pi}{\sigma^2}\left[\frac{1-1/p_1}{\alpha\sin(\pi/p_1)} + \frac{1-1/p_2}{(1-\alpha)\sin(\pi/p_2)}\right].$$

Corollary 6. *For the SEPD ($p_1 = p_2 = p$), $\frac{\partial \ln f}{\partial p} = \frac{\partial \ln f}{\partial p_1} + \frac{\partial \ln f}{\partial p_2}$ from (48); the terms ϕ_{ij} involving p_1 and p_2, become $\phi_{12}+\phi_{13}$, $\phi_{22}+\phi_{33}$, $\phi_{24}+\phi_{34}$ and $\phi_{25}+\phi_{35}$. The information matrix for the MLE of the SEPD parameters (α, p, μ, σ) is:*

$$I(\theta_0) = -H(\theta_0)$$
$$= \begin{pmatrix} \frac{p+1}{\alpha(1-\alpha)} & 0 & -\frac{p}{\sigma\alpha(1-\alpha)} & 0 \\ 0 & \frac{p+1}{p^4}\Psi'(\frac{p+1}{p}) & 0 & -\frac{1}{\sigma p} \\ -\frac{p}{\sigma\alpha(1-\alpha)} & 0 & \frac{\Gamma(1/p)\Gamma(2-1/p)}{\sigma^2\alpha(1-\alpha)} & 0 \\ 0 & -\frac{1}{\sigma p} & 0 & \frac{p}{\sigma^2} \end{pmatrix}.$$

Proposition 7. *(Asymptotic Normality of MLE) Suppose that $p_{01} > 1$ and $p_{02} > 1$. Then the MLE $\widehat{\theta}$ of θ_0 is asymptotically normal, i.e.,*

$$\sqrt{T}(\widehat{\theta} - \theta_0) \to^d N(0, I^{-1}(\theta_0)),$$

where $I(\theta_0)$ is the Fisher information matrix:

$$I(\theta_0) \equiv E[(\nabla_\theta \ln f(Y_t \mid \theta_0))(\nabla_\theta \ln f(Y_t \mid \theta_0))']$$

provided by (23); it can be consistently estimated by $I(\widehat{\theta})$.

Proof. See Appendix C. ■

The information matrix equality $I(\theta_0) = -H(\theta_0)$ holds only for $p_{01} > 1/2$ and $p_{02} > 1/2$, because $E[\frac{\partial \ln f}{\partial \mu}]^2$ as an element of $I(\theta_0)$ may not exist or may be negative for some points of $p_{01} \le 1/2$ and (or) $p_{02} \le 1/2$. Since $I(\theta)$ is continuous for all $\theta \in \Xi$ satisfying $p_1 > 1/2$ and $p_2 > 1/2$, it follows from the consistency of $\widehat{\theta}$ that $I(\widehat{\theta})$ is a consistent estimator of $I(\theta_0)$. The restriction $p_{01} > 1$ and $p_{02} > 1$ ensures that the expected score vector converges uniformly in the neighborhood of the location parameter and is required for the estimation of the location parameter. The restriction is not an impediment in most applications. Even for the GARCH option pricing model with GED conditional distribution in Duan (1999), this restriction is imposed in order to ensure the existence of the expected simple return. If μ_0 is known, then the usual $\sqrt{T}$-asymptotics hold for the MLEs of other parameters ($\alpha_0, p_{01}, p_{02}, \sigma_0$) without any restrictions. When location parameter μ_0 can be consistently estimated by a nonparametric method (see Andrews et al. (1972) and Bickel (2002)), the MLEs of other parameters are still consistent, asymptotically normal but may not be efficient.

6. Performance of MLE in simulation

A stochastic representation of a distribution is important to simulation studies. For given values of parameters, p_1, p_2 and α ($0 < \alpha < 1$, $p_i > 0$, $i = 1, 2$), we can generate standard AEPD random numbers by the following method: first, generate three random numbers U, W_1 and W_2, where U is drawn from standard uniform distribution $U(0, 1)$ and W_i ($i = 1, 2$) is from the gamma distribution with shape parameter $1/p_i$ and pdf $f_{W_i}(w) = \Gamma(1/p_i)^{-1} w^{1/p_i - 1} \exp(-w)$; second, define a random variable Y :

$$Y = \alpha W_1^{1/p_1}\left[\frac{\text{sign}(U-\alpha)-1}{2\Gamma(1+1/p_1)}\right] + (1-\alpha)\,W_2^{1/p_2}\left[\frac{\text{sign}(U-\alpha)+1}{2\Gamma(1+1/p_2)}\right], \quad (24)$$

where $\text{sign}(x) = +1$ if $x > 0$, -1 if $x < 0$, and 0 if $x = 0$. It is straightforward to show that random variable Y has the density (6) of standard AEPD (location $\mu = 0$, scale $\sigma = 1$). An alternative method is the inverse method, i.e., using $Y = F_{AEP}^{-1}(U)$ to generate standard AEPD random numbers, where U is a standard uniform random variable and F_{AEP} is the standard AEPD cdf. However, this method is very time-consuming, while the method given in (24) allows us to generate AEPD random numbers more quickly in Matlab.

To assess the asymptotic properties of the MLE in finite samples, following Agrò (1995), a numerical investigation of bias and variance of MLEs was made using sample sizes of T = 500, 1000, 2000, 4000, 8000. We choose $\mu_0 = 0$, $\sigma_0 = 1$ and various different true values of (α, p_1, p_2): α = 0.3, 0.5 and p_i = 0.7, 1, 1.5, 2.5 (i = 1, 2). To save space, here we only report the cases of $\alpha = 0.3$ and $p_2 = 1, 1.5$. For each set of true values of parameters and every sample size, $N = 2000$ replications are drawn from the AEPD with the set of parameter values, and then N = 2000 ML estimates $\widehat{\theta}^i$ ($i = 1, 2, \ldots, N$) are obtained using these samples. So, we can estimate the means and standard deviations of the MLEs of parameters, denoted respectively by $M(\widehat{\theta})$ and $STD(\widehat{\theta})$,

$$M(\widehat{\theta}) = \frac{1}{N}\sum_{i=1}^{N}\widehat{\theta}^i,$$

$$STD(\widehat{\theta}) = \left(\frac{1}{N}\sum_{i=1}^{N}\left[\widehat{\theta}^i - M(\widehat{\theta})\right]^2\right)^{1/2},$$

and compare these estimated standard deviations with their theoretical values which are taken from the square root of the diagonal elements of Cramer–Rao bound (i.e., $I^{-1}(\widehat{\theta})/T$). Simulation results are presented in Table 1. All entries labeled "Mean of MLEs" report $M(\widehat{\theta})$, and those in "STD Ratio" rows are the ratios of simulated standard deviations $STD(\widehat{\theta})$ to the theoretical ones from $I^{-1}(\widehat{\theta})/T$.

From our simulation studies, we can see that the estimates $\widehat{\theta}$ of all parameters seem asymptotically unbiased for all given true values, and that their variance seems to be approaching the Cramer–Rao bound for the cases of $p_{01} > 1$ and $p_{02} > 1$ (see the cases in which $(p_{01}, p_{02}) = (1.5, 1.5)$ and $(2.5, 1.5)$). However, for the cases of $p_{01} \le 1$ or $p_{02} \le 1$, the behavior of the variance appears to be problematic. More specifically, although the estimates of scale σ appear always efficient in all cases, there are significantly large ratios of standard deviation for the other parameters, especially for μ and α; but the larger the values of the tail parameters, p_1 and p_2, the more efficient the estimates of α and μ appear to be. Other observed phenomena are that (1), because of fewer observations on the left side, estimates of the left tail parameter p_1 have slower convergence and lower efficiency than those of the right tail p_2 (see the cases in which $p_{01} = p_{02} =$ 1 or 1.5); (2), in general, the MLE is more efficient in the cases with larger tail parameter p_1 or p_2 than for those with smaller tail parameters. Finally, we want to point out that for a small sample, say a size less than 500, the likelihood function may not have any maximum point. This problem still exists for the GED and is discussed in detail in Agrò (1995).

7. Forecasting value at risk: An empirical examination

In this section, we examine forecast for value at risk (VaR) with GARCH type model and compare performance for error distributions given by GED, SEPD and our AEPD.

7.1. Model and data

GARCH type models have been widely and successfully used to model financial asset returns. In general, a return process $r = \{r_t\}$

Table 1
Simulation results for the MLE of the AEPD.

	P2 = 1					P2 = 1.5					
	alpha = 0.3	p1 = 0.7	p2 = 1	sigma = 1	mu = 0	alpha = 0.3	p1 = 0.7	p2 = 1.5	sigma = 1	mu = 0	T
Mean of MLEs	0.3042	0.7125	1.0049	1.0008	0.0056	0.3041	0.7100	1.5090	0.9966	0.0053	500
	0.3021	0.7050	1.0004	0.9986	0.0033	0.3030	0.7068	1.5027	0.9985	0.0038	1000
	0.3013	0.7032	1.0008	0.9999	0.0017	0.3013	0.7032	1.5012	0.9992	0.0016	2000
	0.3005	0.7013	0.9999	0.9991	0.0007	0.3004	0.7012	1.5006	0.9995	0.0003	4000
	0.3003	0.7004	0.9993	0.9991	0.0004	0.3002	0.7007	1.4998	0.9995	0.0003	8000
	alpha = 0.3	p1 = 0.7	p2 = 1	sigma = 1	mu = 0	alpha = 0.3	p1 = 0.7	p2 = 1.5	sigma = 1	mu = 0	T
STD ratio	1.172	1.137	1.066	1.038	1.437	1.140	1.107	1.086	1.026	1.343	500
	1.342	1.126	1.115	1.016	1.612	1.331	1.172	1.128	1.032	1.557	1000
	1.334	1.120	1.077	1.011	1.565	1.318	1.124	1.138	1.031	1.545	2000
	1.280	1.104	1.069	1.022	1.479	1.230	1.071	1.126	1.039	1.418	4000
	1.224	1.097	1.048	1.002	1.375	1.196	1.065	1.110	1.041	1.350	8000
	alpha = 0.3	p1 = 1	p2 = 1	sigma = 1	mu = 0	alpha = 0.3	p1 = 1	p2 = 1.5	sigma = 1	mu = 0	T
Mean of MLEs	0.3045	1.0272	1.0065	1.0052	0.0062	0.3082	1.0318	1.5104	1.0012	0.0095	500
	0.3019	1.0141	1.0050	1.0045	0.0024	0.3023	1.0121	1.5104	1.0019	0.0022	1000
	0.3005	1.0058	1.0036	1.0024	0.0009	0.3008	1.0048	1.5047	1.0007	0.0007	2000
	0.2999	1.0019	1.0017	1.0007	0.0002	0.3005	1.0026	1.5009	1.0000	0.0004	4000
	0.2999	1.0008	1.0010	1.0004	0.0000	0.3004	1.0021	1.5005	1.0000	0.0004	8000
	alpha = 0.3	p1 = 1	p2 = 1	sigma = 1	mu = 0	alpha = 0.3	p1 = 1	p2 = 1.5	sigma = 1	mu = 0	T
STD ratio	1.178	1.185	1.092	1.032	1.264	1.175	1.184	1.094	1.021	1.215	500
	1.250	1.186	1.076	1.019	1.321	1.201	1.167	1.087	1.023	1.250	1000
	1.183	1.124	1.060	1.020	1.243	1.165	1.110	1.059	0.987	1.196	2000
	1.114	1.072	1.042	0.999	1.146	1.077	1.026	1.028	0.982	1.101	4000
	1.099	1.035	1.034	0.992	1.127	1.043	1.015	1.013	0.982	1.054	8000
	alpha = 0.3	p1 = 1.5	p2 = 1	sigma = 1	mu = 0	alpha = 0.3	p1 = 1.5	p2 = 1.5	sigma = 1	mu = 0	T
Mean of MLEs	0.3081	1.5894	1.0014	1.0079	0.0097	0.3126	1.5894	1.5049	1.0064	0.0142	500
	0.3060	1.5568	0.9989	1.0054	0.0074	0.3067	1.5499	1.5052	1.0047	0.0070	1000
	0.3030	1.5293	1.0001	1.0027	0.0039	0.3029	1.5233	1.5035	1.0021	0.0032	2000
	0.3015	1.5158	1.0004	1.0020	0.0019	0.3012	1.5108	1.5028	1.0017	0.0013	4000
	0.3010	1.5092	1.0002	1.0011	0.0012	0.3005	1.5060	1.5025	1.0013	0.0006	8000
	alpha = 0.3	p1 = 1.5	p2 = 1	sigma = 1	mu = 0	alpha = 0.3	p1 = 1.5	p2 = 1.5	sigma = 1	mu = 0	T
STD ratio	1.085	1.116	1.111	1.026	1.116	1.048	1.142	1.078	1.027	1.073	500
	1.187	1.184	1.090	1.032	1.206	1.125	1.152	1.089	1.031	1.135	1000
	1.137	1.129	1.041	1.023	1.159	1.089	1.102	1.072	1.013	1.092	2000
	1.104	1.085	1.041	1.011	1.111	1.069	1.069	1.046	0.998	1.064	4000
	1.068	1.053	1.034	0.996	1.068	1.036	1.046	1.025	1.001	1.036	8000
	alpha = 0.3	p1 = 2.5	p2 = 1	sigma = 1	mu = 0	alpha = 0.3	p1 = 2.5	p2 = 1.5	sigma = 1	mu = 0	T
Mean of MLEs	0.3013	2.6811	1.0059	1.0029	0.0037	0.3188	2.8367	1.4925	1.0106	0.0231	500
	0.3016	2.5976	1.0013	1.0016	0.0034	0.3121	2.6939	1.4951	1.0078	0.0146	1000
	0.3018	2.5616	1.0003	1.0025	0.0026	0.3057	2.5887	1.4972	1.0030	0.0068	2000
	0.3019	2.5405	0.9998	1.0020	0.0024	0.3027	2.5405	1.4973	1.0006	0.0032	4000
	0.3007	2.5149	0.9998	1.0004	0.0010	0.3017	2.5265	1.4992	1.0007	0.0019	8000
	alpha = 0.3	p1 = 2.5	p2 = 1	sigma = 1	mu = 0	alpha = 0.3	p1 = 2.5	p2 = 1.5	sigma = 1	mu = 0	T
STD ratio	0.994	1.089	1.088	1.007	1.010	0.869	1.007	0.997	1.047	0.893	500
	1.109	1.142	1.070	1.043	1.135	0.974	1.042	1.025	1.035	0.991	1000
	1.126	1.125	1.080	1.046	1.129	1.008	1.048	1.030	1.021	1.019	2000
	1.116	1.114	1.058	1.068	1.122	0.993	1.024	1.004	1.025	0.997	4000
	1.063	1.079	1.033	1.036	1.072	0.987	1.012	1.000	1.016	0.991	8000

is modeled as[5]

$$r_t = m_t + \sigma_t z_t, \tag{25}$$

where, following tradition, m_t and σ_t^2 are the conditional mean and variance of r_t given the information set available at time $t-1$ (i.e., $m_t = E_{t-1}(r_t)$ and $\sigma_t^2 = E_{t-1}(r_t - m_t)^2$), z_t are the i.i.d. innovations with zero mean and unit variance.

To capture the leverage effect, we adopt the non-linear asymmetric GARCH (NGARCH) structure of Engle and Ng (1993). The conditional distribution of the return process is modelled as the AEPD type distribution. For simplicity, we assume $m_t = m$, for any t; the return series r_t is an AEPD-NGARCH(1, 1) process,

$$\begin{aligned} r_t &= m + \sigma_t z_t, \quad z_t \sim i.i.d. AEPD(0, 1), \\ \sigma_t^2 &= b_0 + b_1 \sigma_{t-1}^2 + b_2 \sigma_{t-1}^2 (z_{t-1} - c)^2 \\ &= b_0 + b_1 \sigma_{t-1}^2 + b_2 (r_{t-1} - m - c\sigma_{t-1})^2. \end{aligned} \tag{26}$$

The parameter c in the NGARCH equation (26) captures the leverage effect; that is, a positive value of c gives rise to a negative correlation between the innovations in the asset return and its conditional volatility. The j-step-ahead forecast of σ_{t+j}^2, denoted by $\sigma_{t+j|t}^2$, is defined as $\sigma_{t+j|t}^2 \equiv E_t\left(\sigma_{t+j}^2\right)$; it is

$$\sigma_{t+1|t}^2 = b_0 + b_1 \sigma_t^2 + b_2 (r_t - m - c\sigma_t)^2, \tag{27}$$

$$\sigma_{t+j|t}^2 = b_0 + \left[b_1 + b_2(1 + c^2)\right] \sigma_{t+j-1|t}^2, \quad j \geq 2. \tag{28}$$

5 As noted by Andersen et al. (2005), this representation is not entirely general as there could be higher-order conditional dependence in the innovations.

Table 2
Parameter estimates for the AEPD-NGARCH(1, 1) models.

	m	b_0	b_1	b_2	c	α	p_1	p_2
AEPD	.0254	.0089	.8918	.0583	.8802	.400	1.182	1.820
	(.014)	(.002)	(.014)	(.008)	(.119)	(.01)	(.049)	(.084)
AEPD, $\alpha = 0.5$	.019	.010	.8852	.0603	.9071		1.384	1.539
	(.013)	(.003)	(.015)	(.008)	(.115)		(.05)	(.065)
SEPD	.0218	.010	.884	.0604	.914	.522		1.449
	(.014)	(.003)	(.015)	(.009)	(.119)	(.011)		(.049)
GED	.0296	.0095	.8853	.0597	.9111			1.437
	(.013)	(.003)	(.015)	(.009)	(.119)			(.048)

We consider daily returns[6] on the S&P500 composite index. Empirical evidence has indicated that high frequency data continue to exhibit conditional tail-fatness even after allowing for the GARCH effect (see Bollerslev et al. (1992)). Our sample covers the period from January 2, 1990 to December 31, 2002, and the sample size is $T = 3280$. The data set is from CRSP (Center for Research in Security Prices, University of Chicago).

7.2. Estimation and goodness of fit

The maximum likelihood estimate of the parameter vector ϕ, where $\phi = (m, b_0, b_1, b_2, c, \alpha, p_1, p_2)$, is obtained by maximizing the log-likelihood function

$$L(\phi; r) = \sum_{t=1}^{T} \left\{ \log \delta - \log \sigma_t + \log f_Y \left(\omega + \delta \frac{r_t - m}{\sigma_t} \mid \beta \right) \right\}, \quad (29)$$

where $f_Y(\cdot \mid \beta)$ is the standard density function of the AEPD with the distributional parameters $\beta = (\alpha, p_1, p_2)'$, $\omega \equiv \omega(\beta)$ and $\delta \equiv \delta(\beta)$ denote the mean and standard deviation of $f_Y(\cdot \mid \beta)$ respectively and as functions of β both are given in (16) and (17).[7]

To show the significance of asymmetric behavior in the tails, we consider the AEPD and nested distribution classes: the AEPD with $\alpha = 1/2$ to represent asymmetry arising only from different tail behavior, the SEPD (i.e., AEPD with $p_1 = p_2$), the GED (i.e., AEPD with $\alpha = 1/2$ and $p_1 = p_2$). The ML estimates of the parameters and their standard deviations are displayed in Table 2.

Following Mittnik and Paolella (2003), we employ four criteria for comparing the goodness of fit of the candidate models. The first is the maximum log-likelihood value (L), which can be viewed as an overall measure of goodness of fit. The second and the third are the *AICC* (Hurvich and Tsai, 1989) and the *SBC* or *SIC* (Schwarz, 1978), which modify the *AIC*, and are given by

$$AICC = -2L + \frac{2T(k+1)}{T-k-2}, \qquad SBC = -2L + \frac{k \log(T)}{T}; \quad (30)$$

k denotes the number of estimated parameters and T the number of observations. The fourth is the Anderson–Darling statistic (Anderson and Darling, 1952), defined as

$$AD = \sqrt{T} \sup_{x \in R} \frac{\left| F_T(x) - \widehat{F}(x) \right|}{\sqrt{\widehat{F}(x)(1 - \widehat{F}(x))}}, \quad (31)$$

where $\widehat{F}(x)$ denotes the estimated (parametric) cdf of innovation, and $F_T(x)$ is the empirical cdf of (*ex post*) innovations, i.e., $F_T(x) = n/T$ if there are only n *ex post* innovations $\widehat{z}_t = (r_t - \widehat{m})/\widehat{\sigma}_t$ less or equal to x for any given x.

[6] The return r_t in period t is defined as $r_t = 100 \times (P_t - P_{t-1})/P_{t-1}$, where P_t is the Level on S&P Composite Index at time t.

[7] The ML estimation is implemented in Matlab 6.1 with the command 'fmincon' and initial value $\phi_0 = (\text{mean}(r), b_0, 0.9, 0.05, 0, 0.5, 1.5, 1.5)$, where b_0 is given by the variance of returns data multiplied by $1 - b_1 - b_2 = 0.05$.

Table 3
Goodness-of-fit measures for the AEPD-NGARCH(1, 1) models.

	L	AICC	SBC	AD
AEPD	−4255.7	8529.4	8511.4	8.35
AEPD, $\alpha = 0.5$	−4260.0	8536.0	8520.0	15.53
SEPD	−4262.3	8540.6	8524.6	21.65
GED	−4264.1	8542.2	8528.2	27.54

The *AD* statistic is a reasonable measure of the discrepancy or "distance" between the two distributions, say, the empirical cdf $F_T(x)$ and the hypothetical distribution $F(x)$; this statistic gives appropriate weight to the tails of the distribution so that it can be used to measure goodness of fit in the tails. In our applications, since the innovations are assumed to have zero mean and unit variance, the estimated cdf of the innovations, $\widehat{F}(x)$ in (31), can be expressed as $\widehat{F}(x) = F_Y(\widehat{\omega} + \widehat{\delta} x \mid \widehat{\beta})$, where $F_Y(\cdot \mid \beta)$ is the cdf of the standard AEPD with $\beta = (\alpha, p_1, p_2)^{\mathrm{T}}$, $\widehat{\beta}$ is the ML estimate of β, and $\widehat{\omega}$ and $\widehat{\delta}$ are given by $\widehat{\omega} = \omega(\widehat{\beta})$ and $\widehat{\delta} = \delta(\widehat{\beta})$, which are, respectively, the estimated mean and standard deviation of $F_Y(\cdot \mid \beta)$. For simplicity we compute the *AD* statistic as follows:

$$AD = \max_j AD_j, \quad AD_j = \sqrt{T} \frac{\left| F_T(\widehat{z}_{j,T}) - \widehat{F}(\widehat{z}_{j,T}) \right|}{\sqrt{\widehat{F}(\widehat{z}_{j,T})(1 - \widehat{F}(\widehat{z}_{j,T}))}}, \quad (32)$$

where $\left\{ \widehat{z}_{j,T} \right\}_{j=1}^{\mathrm{T}}$ are the sorted (in ascending order) *ex post* innovations, thus $F_T(\widehat{z}_{j,T}) = j/T$.

Table 3 displays the four measures of goodness-of-fit for the estimated AEPD-NGARCH(1,1) models. All measures rank the distribution with full asymmetry as the best, followed by asymmetry in tails only, skewness only, and then symmetry in a descending order.

7.3. Prediction performance for downside risk

To predict the downside risk (VaR) in the period $t + j$ ($j = 1, 2, 3, \ldots$) using the information available in period t, we must give a j-step-ahead forecast of the conditional distribution of returns r_{t+j}, $\widehat{F}_{t+j|t}(r_{t+j})$. Based on the above models specified in (25) and (26), the conditional distribution is time-varying only due to the time-varying conditional mean and variance. Therefore forecasting the conditional distribution boils down to estimating the parameters of the model using the data available at time t, and then forecasting the conditional mean (m_{t+j}) and variance (σ_{t+j}^2) of r_{t+j}. Denote the time-t ML estimates of these parameters by $(\widehat{m}_t, \widehat{b}_{0t}, \widehat{b}_{1t}, \widehat{b}_{2t}, \widehat{c}_t, \widehat{\beta}_t')$ and the estimates of the j-step-ahead forecasts of m_{t+j} and σ_{t+j}^2 by $\widehat{m}_{t+j|t}$ and $\widehat{\sigma}_{t+j|t}^2$, respectively. Then, $\widehat{m}_{t+j|t} = \widehat{m}_t$ for any j, and $\widehat{\sigma}_{t+j|t}^2$ is obtained by substituting the estimated parameters into (27) and (28),

$$\widehat{\sigma}_{t+1|t}^2 = \widehat{b}_{0t} + \widehat{b}_{1t} \widehat{\sigma}_t^2 + \widehat{b}_{2t} (r_t - \widehat{m}_t - \widehat{c}_t \widehat{\sigma}_t)^2,$$

$$\widehat{\sigma}_{t+j|t}^2 = \widehat{b}_{0t} + \left[\widehat{b}_{1t} + \widehat{b}_{2t} (1 + \widehat{c}_t^2) \right] \widehat{\sigma}_{t+j-1|t}^2, \quad j \geq 2.$$

Now we consider estimation of the j-step-ahead forecast of the conditional VaR. Note that $z_{t+j} = (r_{t+j} - m)/\sigma_{t+j}$ is assumed to be an AEPD random variable with zero mean and unit variance. Then, based on the predicted values $\widehat{m}_{t+j|t}$, $\widehat{\sigma}_{t+j|t}$, $\widehat{\omega}_t = \omega(\widehat{\beta}_t)$ and $\widehat{\delta}_t = \delta(\widehat{\beta}_t)$, where $\omega(\cdot)$ and $\delta(\cdot)$ are defined in (29), we expect at time t that, conditional on the information available in period t,

$$\widehat{Y}_{AEP} \equiv \widehat{\omega}_t + \widehat{\delta}_t (r_{t+j} - \widehat{m}_{t+j|t}) / \widehat{\sigma}_{t+j|t}$$

should approximately have a standard AEPD density with parameter $\widehat{\beta}_t$ if the model specification is correct. Therefore, the j-step-ahead forecast of the conditional VaR can be estimated as follows:

$$VaR_{t+j|t}(p) \equiv \widehat{F}_{t+j|t}^{-1}(p) = \widehat{m}_{t+j|t} + \widehat{\sigma}_{t+j|t} \left[\frac{F_{AEP}^{-1}(p \mid \widehat{\beta}_t) - \widehat{\omega}_t}{\widehat{\delta}_t} \right]. \quad (33)$$

Table 4
Predictive performance for the value at risk (VaR).

p	0.01	0.025	0.05	0.10
$\widehat{\mathbf{f}}_j(p)$	$\widehat{\mathbf{f}}_1(p), \widehat{\mathbf{f}}_5(p)$	$\widehat{\mathbf{f}}_1(p), \widehat{\mathbf{f}}_5(p)$	$\widehat{\mathbf{f}}_1(p), \widehat{\mathbf{f}}_5(p)$	$\widehat{\mathbf{f}}_1(p), \widehat{\mathbf{f}}_5(p)$
AEPD	*.0110, .0122*	*.0250, .0263*	.0537, **.0538**	.1232, **.1222**
AEPD, $\alpha = 0.5$	.0116, **.0134**	.0250, **.0275**	*.0506, .0526*	*.1152, .1143*
SEPD	.0128, **.0153**	.0268, **.0287**	.0543, **.0544**	.1165, **.1161**
GED	.0146, **.0165**	.0299, **.0293**	.0598, **.0575**	.1195, **.1204**

Note that, in the calculation of the conditional VaR, we also need to give the expressions for $F_{AEP}^{-1}(p \mid \beta)$, i.e., the quantile function (see (11)). Therefore, the downside risk is determined not only by the specification of the conditional mean and NGARCH equations, but also by the distributional choice for the innovations. We can express the predictive downside risk as follows: after j periods, the return would be less than $VaR_{t+j|t}(p)$ with probability p.

To see predictive performance out-of-sample, we split the sample in two: $N = T/2 = 1640$. Then we evaluate $VaR_{t+j|t}(p)$, $N = 1640 \le t \le T - j$, for one and five steps ahead: $j = 1, 5$. We set the shortfall probabilities $p = 0.01, 0.025, 0.05, 0.1$. For each of (j, p), if the model is correctly specified we expect $100p$% of the observed r_{t+j}-values $(r_{N+j}, \ldots, r_T)$ to be less than the $VaR_{t+j|t}(p)$ implied by the model. If the observed frequency

$$\widehat{\mathbf{f}}_j(p) = \frac{1}{T - N - j + 1} \sum_{t=N}^{T-j} 1\{r_{t+j} < VaR_{t+j|t}(p)\} \tag{34}$$

is lower (higher) than p, then the model tends to overestimate (underestimate) the risk.

Table 4 shows the predictive performance of VaR; the entries in the table are the observed frequency $\widehat{f}_j(p)$ given in (34) for one and five steps ahead: $j = 1, 5$, and shortfall probabilities $p = 0.01, 0.025, 0.05, 0.1$. All the models tend to underestimate the value at risk, but the models with AEPD errors perform noticeably better than those with a single tail parameter. For small shortfall probabilities ($p = 0.01, 0.025$), the unrestricted AEPD is the best; for the $p = 0.05$ and 0.1, the AEPD with $\alpha = .5$ dominates.

Thus both by measures of fit and by forecasting performance the AEPD class provides a useful model for the error in GARCH type model of returns on assets.

Acknowledgements

The support of the Social Sciences and Humanities Research Council of Canada (SSHRC) and the *Fonds qué becois de la recherche sur la société et la culture* (FRQSC) is gratefully acknowledged.

Appendix A

The proofs make extensive use of several results.

(I) Integral (see Gradshteyn and Ryzhik (1994). #3.478)

$$\int_0^\infty x^{\upsilon-1} \exp(-\mu x^p) dx = \frac{1}{p} \mu^{-\upsilon/p} \Gamma\left(\frac{\upsilon}{p}\right),$$

for $\mu > 0, \upsilon > 0, p > 0$. (35)

(II) For the gamma function $\Gamma(x)$ and *digamma function* $\Psi(x) \equiv \Gamma'(x)/\Gamma(x)$:

$$\Psi(x) = -C - \frac{1}{x} + \sum_{i=1}^\infty \left(\frac{1}{i} - \frac{1}{x+i}\right),$$

$$\Psi^{(k)}(x) = \sum_{i=0}^\infty \frac{k(-1)^{k+1}}{(x+i)^{k+1}}, \tag{36}$$

$$\Gamma(x+1) = x\Gamma(x),$$

$$\Gamma'(x+1)/\Gamma(x) = 1 + x\Psi(x), \tag{37}$$

$$\Gamma''(x)/\Gamma(x) = \Psi'(x) + \Psi^2(x), \tag{38}$$

$$\Gamma''(x+1)/\Gamma(x) = 2\Psi(x) + x\Psi^2(x) + x\Psi'(x), \tag{39}$$

where C is Euler's constant, k any positive integer. More properties and details are in Abramowitz and Stegun (1970, p 255–263), Artin (1964, p 16–26) and Farrell and Ross (1963).

(III) Properties of EPD (based on Box and Tiao (1973) and Kotz et al. (2001)).

For Z_p with standard EPD density ($\mu = 0, \sigma_p = 1$) in (1) the cdf and quantile function are

$$F_{EP}(x \mid p) = \frac{1}{2}\left[1 + \text{sign}(x) G\left(\frac{1}{p}|x|^p; \frac{1}{p}\right)\right], \tag{40}$$

$$F_{EP}^{-1}(\upsilon \mid p) = \text{sign}(2\upsilon - 1)\left[pG^{-1}\left(|2\upsilon - 1|; \frac{1}{p}\right)\right]^{1/p}, \tag{41}$$

$G(x; \gamma)$ is the gamma cdf, and $G^{-1}(x; \gamma)$ is the inverse function of $G(x; \gamma)$. By change of variable, (35), for $M(p, r)$ (see Proposition 1(d)) the absolute moment is

$$E(|Z_p|^r) = p^{r/p}\Gamma\left(\frac{r+1}{p}\right) \Big/ \Gamma(1/p) \equiv [M(p, r)]^r,$$

$r > -1$. (42)

The expected shortfall of Z_p, $ES_{EP}(x \mid p) \equiv E(-Z_p \mid Z_p < x)$, is:

$$ES_{EP}(x \mid p) = p^{1/p} \frac{\Gamma(2/p)}{\Gamma(1/p)} \left[\frac{1 - G(\frac{1}{p}|x|^p; 2/p)}{1 + \text{sign}(x) G(\frac{1}{p}|x|^p; 1/p)}\right]. \tag{43}$$

For $x = VaR_{EP}(\upsilon) \equiv F_{EP}^{-1}(\upsilon \mid p)$, ES as a function of confidence level υ, $ES_{EP}^*(\upsilon \mid p)$, is

$$ES_{EP}^*(\upsilon \mid p) = p^{1/p} \frac{\Gamma(2/p)}{\Gamma(1/p)} \frac{1}{2\upsilon} \left\{1 - G\left[G^{-1}\left(|2\upsilon - 1|; \frac{1}{p}\right); 2/p\right]\right\}. \tag{44}$$

Proof of Proposition 1. The result $P(X \le \mu) = \alpha$ follows directly from (10). Proofs of other equalities in (a), (b), (d) of Proposition 1 boil down to calculations of $d_L(r)$ and $d_R(r)$. For $d_R(r)$ of the standard AEPD ($\mu = 0, \sigma = 1$), by change of variable $z = x/[2(1 - \alpha^*)]$ and (42) or (35), we have

$$\begin{aligned} d_R(r) &= \left\{E[|X|^r \mid X > 0]\right\}^{1/r} \\ &= \left\{\int_0^\infty x^r f_{AEP}(x; \alpha, p_1, p_2) \frac{1}{1-\alpha} dx\right\}^{1/r} \\ &= 2(1 - \alpha^*)\left\{2\int_0^\infty z^r K_{EP}(p_2) \exp\left(-\frac{1}{p_2} z^{p_2}\right) dz\right\}^{1/r} \\ &= 2(1 - \alpha^*)\{E(|Z_{p_2}|^r)\}^{1/r} = 2(1 - \alpha^*) M(p_2, r). \end{aligned}$$

To prove that $\xi(p,r) \equiv K_{EP}(p)M(p,r)$ is strictly increasing in r, we evaluate the derivative of $\ln M(p,r)$ with respect to r and show $\frac{\partial \ln M(p,r)}{\partial r} > 0$. Note that

$$\frac{\partial \ln M(p,r)}{\partial r} = \frac{1}{pr}\Psi\left(\frac{r+1}{p}\right) - \frac{1}{r^2}[\log\Gamma((r+1)/p) - \log\Gamma(1/p)].$$

By the mean value theorem, we have

$$\frac{\partial \ln M(p,r)}{\partial r} = \frac{1}{pr}\left[\Psi\left(\frac{r+1}{p}\right) - \Psi\left(\frac{\varepsilon r+1}{p}\right)\right],$$

where $0 < \varepsilon < 1$.

Since $\Psi'(x)$ is positive for any $x > 0$ (see Abramowitz and Stegun (1970), 6.4.10), it follows that $\frac{\partial \ln M(p,r)}{\partial r} > 0$ for any $r > 0$ and $p > 0$.

To show that $\xi(p,r)$ is strictly decreasing in p for any given $r > 0$, write

$$\frac{\partial \ln \xi(p,r)}{\partial p} = \frac{1}{p^2}\Psi(1+1/p) + \frac{1}{pr}\left[\frac{1}{p}\Psi\left(\frac{1}{p}\right) - \frac{r+1}{p}\Psi\left(\frac{r+1}{p}\right)\right],$$

and note that the second part of the above expression, denoted by $h(p,r)$,

$$h(p,r) = \frac{C}{p^2} + \sum_{i=1}^{\infty}\frac{1}{pr}\left[g_i\left(\frac{1}{p}\right) - g_i\left(\frac{r+1}{p}\right)\right], \tag{45}$$

where $g_i(x) \equiv x/i - x/(i+x)$, C is Euler's constant and (45) for $h(p,r)$ is based on (36). From the mean value theorem and $g_i'(x) > 0$, $g_i''(x) > 0$ for any $x > 0$ and $i \geq 1$, it follows that $h_i(p,r) \equiv \frac{1}{pr}\left[g_i(\frac{1}{p}) - g_i(\frac{r+1}{p})\right]$ is strictly decreasing in r for every $i \geq 1$; so $h(p,r)$ is a decreasing function of r. Therefore, for $r > 0$

$$\frac{\partial \ln \xi(p,r)}{\partial p} < \frac{1}{p^2}\Psi(1+1/p) + \lim_{r\to 0^+}\frac{1}{pr}\left[\frac{1}{p}\Psi\left(\frac{1}{p}\right) - \frac{r+1}{p}\Psi\left(\frac{r+1}{p}\right)\right] = \frac{1}{p^2}\left[p - \frac{1}{p}\Psi'\left(\frac{1}{p}\right)\right] = -\left(\frac{1}{p}\right)^3\sum_{i=1}^{\infty}\frac{1}{(i+1/p)^2} < 0,$$

since $\Psi(1+x) = \Psi(x) + 1/x$ by (37); the last equality uses (36) for $\Psi'(x)$.

To prove Proposition 1(c), define an increasing function $r = r^*(c \mid p) \equiv \xi^{-1}(c \mid p)$ for a given p. Note that $\xi(p,r) \downarrow lb(p)$ as $r \downarrow 0$ and $\xi(p,r) \uparrow +\infty$ as $r \uparrow +\infty$ (by using Equality 6.1.20 in Abramowitz and Stegun (1970)). When $c > lb(p)$, $r = \xi^{-1}(c \mid p) > 0$ and thus $r^*(c \mid p_1) > 0$ and $r^*(c \mid p_2) > 0$ for any $c > \max\{lb(p_1), lb(p_2)\}$. Using definition of $r^*(c \mid p)$ and equalities in Proposition 1(d), we get Proposition 1(c). ■

Proof of Proposition 2. The expressions for $kur_L(r)$ and $kur_R(r)$ in Proposition 2 are easily obtained using equalities in Proposition 1(d). Here we prove only that $k(r,p) \equiv \Gamma(\frac{1}{p})\Gamma(\frac{2r+1}{p})/\Gamma^2(\frac{r+1}{p})$ is strictly decreasing in p and increasing in r. The second point follows from

$$\frac{\partial \ln k(r,p)}{\partial r} = \frac{2}{p}\left[\Psi\left(\frac{2r+1}{p}\right) - \Psi\left(\frac{r+1}{p}\right)\right], \quad p > 0, r > 0$$

and $\Psi'(x) > 0$ (see (36)), implying $\frac{\partial \ln k(r,p)}{\partial r} > 0$ for any $r > 0$, $p > 0$. The first point follows from

$$\frac{\partial \ln k(r,p)}{\partial p} = \frac{1}{p}\left[-\rho\left(\frac{2r+1}{p}\right) - \rho\left(\frac{1}{p}\right) + 2\rho\left(\frac{r+1}{p}\right)\right],$$

where $\rho(x) = x\Psi(x)$ is strictly convex in $(0,+\infty)$ (from (36) $\rho''(x) = 2\sum_{i=0}^{\infty} i/(i+x)^3 > 0$ for any $x > 0$). Then $\frac{\partial \ln k(r,p)}{\partial p} < 0$ for any $p > 0$ and $r > 0$. ■

Proof of Proposition 3. The entropy of a distribution with density $f(x;\theta)$ is by definition

$$H(f) \equiv -\int_{-\infty}^{+\infty} f(x;\theta)\ln f(x;\theta)\mathrm{d}x.$$

A straightforward calculation shows that for AEPD

$$H(f) = \ln\sigma + \frac{\alpha}{p_1} + \frac{1-\alpha}{p_2}.$$

By Theorem 13.2.1 of Kagan et al. (1973) (also see Proposition 2.4.6 in Kotz et al. (2001)) for all densities $p(x)$ supported on $(-\infty,+\infty)$ that satisfy:

$$\int_{-\infty}^{+\infty}[L(x;\theta)]^{p_1}p(x)\mathrm{d}x = \frac{\alpha}{p_1}; \quad \int_{-\infty}^{+\infty}[R(x;\theta)]^{p_2}p(x)\mathrm{d}x = \frac{1-\alpha}{p_2}, \tag{46}$$

the maximum entropy is attained by distributions with the densities of the form

$$p_{ME}(x) = \exp\{-\lambda_0 - \lambda_1[L(x;\theta)]^{p_1} - \lambda_2[R(x;\theta)]^{p_2}\}$$

(and only by them), if constants λ_0, λ_1 and λ_2, such that $p_{ME}(x) > 0$ for all $x \in (-\infty,+\infty)$ and satisfies (46), exist. We find a unique set $\{\lambda_0,\lambda_1,\lambda_2\}$ such that $p_{ME}(x)$ is the AEPD density (6). From the conditions in (46) and $\int_{-\infty}^{+\infty} p_{ME}(x)\mathrm{d}x = 1$ straightforward calculations show that

$$\frac{\alpha\sigma}{\lambda_1^{1/p_1}} + \frac{(1-\alpha)\sigma}{\lambda_2^{1/p_2}} = \frac{\sigma}{\lambda_1^{1+1/p_1}} = \frac{\sigma}{\lambda_2^{1+1/p_2}} = \exp(\lambda_0),$$

implying

$$\alpha\lambda_1 + (1-\alpha)\lambda_2 = 1, \quad \lambda_1^{1+1/p_1} = \lambda_2^{1+1/p_2}. \tag{47}$$

This uniquely determines $(\lambda_1,\lambda_2) = (1,1)$ because λ_2 as a function of λ_1 is strictly decreasing by the first equation in (47) and increasing by the second, and thus $\lambda_0 = \ln\sigma$. ■

Appendix B

Appendix B is devoted to the derivation of the information matrix and to verifying the information matrix equality. Expectations are always taken with respect to the true underlying distribution $f(y;\theta_0)$, where $\theta_0 = (\alpha_0, p_{01}, p_{02}, \mu_0, \sigma_0)$.

Suppose that y_t $(t = 1,2,\ldots,T)$ are i.i.d. observations from the AEPD whose density $f(y_t;\theta)$ with $\theta \in \Theta$ is defined in (6). Let

$$L \equiv L(y_t;\theta) \equiv \frac{\Gamma(1+1/p_1)\,|\mu - y_t|}{\alpha\sigma}1(y_t < \mu),$$

$$R \equiv R(y_t;\theta) \equiv \frac{\Gamma(1+1/p_2)\,|y_t - \mu|}{(1-\alpha)\sigma}1(y_t > \mu).$$

Then the log-density function $\ln f(y_t;\theta) = -\ln\sigma - [L(y_t;\theta)]^{p_1} - [R(y_t;\theta)]^{p_2}$, and the score (vector) for observation t, $\frac{\partial}{\partial\theta}\ln f(y_t;\theta)$, is given by

$$\frac{\partial \ln f}{\partial \alpha} = \frac{p_1}{\alpha}[L(y_t;\theta)]^{p_1} - \frac{p_2}{1-\alpha}[R(y_t;\theta)]^{p_2}, \tag{48}$$
$$\frac{\partial \ln f}{\partial p_1} = \left[\frac{1}{p_1}\Psi(1+1/p_1) - \ln L(y_t;\theta)\right][L(y_t;\theta)]^{p_1},$$
$$\frac{\partial \ln f}{\partial p_2} = \left[\frac{1}{p_2}\Psi(1+1/p_2) - \ln R(y_t;\theta)\right][R(y_t;\theta)]^{p_2},$$
$$\frac{\partial \ln f}{\partial \mu} = -\frac{\Gamma(1/p_1)}{\alpha\sigma}[L(y_t;\theta)]^{p_1-1} + \frac{\Gamma(1/p_2)}{(1-\alpha)\sigma}[R(y_t;\theta)]^{p_2-1},$$
$$\frac{\partial \ln f}{\partial \sigma} = \frac{p_1}{\sigma}[L(y_t;\theta)]^{p_1} + \frac{p_2}{\sigma}[R(y_t;\theta)]^{p_2} - \frac{1}{\sigma},$$

where for $x = 0$ and $p > 0$ set $x^p \ln x = 0$.

To derive the information matrix $I(\theta_0) \equiv E[\frac{\partial}{\partial\theta}\ln f(y_t,\theta_0)\frac{\partial}{\partial\theta'}\ln f(y_t;\theta_0)]$ and the Hessian $H(\theta_0) \equiv E[\frac{\partial^2}{\partial\theta\partial\theta'}\ln f(y_t;\theta_0)]$ and to verify the information matrix equality $I(\theta_0) = -H(\theta_0)$, we first give the following Lemmas.

Lemma 8. *For any real number $r > -1$ and integer $m = 0, 1, 2$, we have*

$$E[L(y_t;\theta_0)]^r[\ln L(y_t;\theta_0)]^m 1(y_t < \mu_0)$$
$$= \frac{\alpha_0}{p_{01}^{m+1}}\frac{\Gamma^{(m)}((1+r)/p_{01})}{\Gamma(1+1/p_{01})}, \tag{49}$$
$$E[R(y_t;\theta_0)]^r[\ln R(y_t;\theta_0)]^m 1(y_t > \mu_0)$$
$$= \frac{1-\alpha_0}{p_{02}^{m+1}}\frac{\Gamma^{(m)}((1+r)/p_{02})}{\Gamma(1+1/p_{02})}, \tag{50}$$

where $\Gamma^{(m)}(\cdot)$ is the mth order derivative of $\Gamma(\cdot)$ and $\Gamma^{(0)}(\cdot)$ means $\Gamma(\cdot)$.

Proof. [8]Both equalities (49) and (50) are similarly. Here we only show (49). Denote by EL the expectation of the left hand side of (49), then

$$EL = \int_{-\infty}^{\mu}[L(y;\theta)]^r[\ln L(y;\theta)]^m f(y;\theta)\mathrm{d}y$$
$$= \int_{-\infty}^{\mu}[L(y;\theta)]^r[\ln L(y;\theta)]^m \frac{1}{\sigma}\exp\{-[L(y;\theta)]^{p_1}\}\mathrm{d}y.$$

Then a change of variable $x = [L(y;\theta)]^{p_1}$ results in

$$EL = \frac{\alpha}{p_1^{m+1}\Gamma(1+1/p_1)}\int_0^{+\infty} x^{(1+r)/p_1-1}(\ln x)^m \exp(-x)\mathrm{d}x$$
$$= \frac{\alpha}{p_1^{m+1}\Gamma(1+1/p_1)}\Gamma^{(m)}((1+r)/p_1),$$

(for derivatives of gamma function see Farrell and Ross (1963, p 22)). ■

Lemma 9. *The score vector for observation t, $\frac{\partial}{\partial\theta}\ln f(y_t;\theta)$, satisfies*

$$E\left[\frac{\partial}{\partial\theta}\ln f(y_t;\theta_0)\right] = 0. \tag{51}$$

[8] For simplicity, we omit the subscript "0"on the true parameters in all the following proofs.

Proof. To verify this we use (49), (50) and (37)–(39) in Appendix A-(II). Here we show $E[\frac{\partial \ln f}{\partial p_1}] = 0$; other calculations are similar. In fact,

$$E\left[\frac{\partial \ln f}{\partial p_1}\right]$$
$$= \frac{1}{p_1}\Psi(1+1/p_1)E[L(y_t;\theta)]^{p_1} - E[L(y_t;\theta)]^{p_1}\ln L(y_t;\theta)$$
$$= \frac{1}{p_1}\Psi(1+1/p_1)\frac{\alpha\Gamma(1+1/p_1)}{p_1\Gamma(1+1/p_1)} - \frac{\alpha\Gamma'(1+1/p_1)}{p_1^2\Gamma(1+1/p_1)}$$
$$= \frac{\alpha}{p_1^2}\Psi(1+1/p_1) - \frac{\alpha}{p_1^2}\Psi(1+1/p_1) = 0. \quad ■$$

Proof of Proposition 5. We derive expressions for $E[\partial^2\ln f(y_t;\theta)/\partial\theta_i\partial\theta_j]$ and $E[\partial \ln f(y_t;\theta)/\partial\theta_i]\cdot[\partial\ln f(y_t;\theta)/\partial\theta_j]$ separately and then verify

$$E\left[\frac{\partial \ln f(y_t;\theta)}{\partial\theta_i}\cdot\frac{\partial \ln f(y_t;\theta)}{\partial\theta_j}\right] = -E\left[\frac{\partial^2 \ln f(y_t;\theta)}{\partial\theta_i\partial\theta_j}\right],$$
$$i, j = 1, 2, \ldots, 5.$$

In the proof we use $1(y_t < \mu)1(y_t > \mu) = 0$ and make use of (49)–(51) and (37)–(39) in Appendix A-(II). Here we show only the equality associated with ϕ_{44}; the others are proved similarly. In fact,

$$E\left[\frac{\partial \ln f}{\partial\mu}\right]^2$$
$$= \left[\frac{\Gamma(1/p_1)}{\alpha\sigma}\right]^2 E[L(y_t;\theta)]^{2(p_1-1)}$$
$$+ \left[\frac{\Gamma(1/p_2)}{(1-\alpha)\sigma}\right]^2 E[R(y_t;\theta)]^{2(p_2-1)}$$
$$= \left[\frac{\Gamma(1/p_1)}{\alpha\sigma}\right]^2 \frac{\alpha\Gamma(2-1/p_1)}{p_1\Gamma(1+1/p_1)}$$
$$+ \left[\frac{\Gamma(1/p_2)}{(1-\alpha)\sigma}\right]^2 \frac{(1-\alpha)\Gamma(2-1/p_2)}{p_2\Gamma(1+1/p_2)}$$
$$= \frac{\Gamma(1/p_1)\Gamma(2-1/p_1)}{\alpha\sigma^2} + \frac{\Gamma(1/p_2)\Gamma(2-1/p_2)}{(1-\alpha)\,\sigma^2};$$

also

$$-E\left[\frac{\partial^2 \ln f}{\partial\mu^2}\right]$$
$$= \Gamma\left(\frac{1}{p_1}\right)\frac{(p_1-1)\,\Gamma(1+1/p_1)}{(\alpha\sigma)^2}E[L^{p_1-2}]$$
$$+ \Gamma\left(\frac{1}{p_2}\right)\frac{(p_2-1)\Gamma(1+1/p_2)}{[(1-\alpha)\sigma]^2}E[R^{p_2-2}]$$
$$= \left[\frac{\Gamma(1/p_1)}{\alpha\sigma}\right]^2\frac{p_1-1}{p_1}\frac{\alpha\Gamma(1-1/p_1)}{p_1\Gamma(1+1/p_1)}$$
$$+ \left[\frac{\Gamma(1/p_2)}{(1-\alpha)\sigma}\right]^2\frac{p_2-1}{p_2}\frac{(1-\alpha)\Gamma(1-1/p_2)}{p_2\Gamma(1+1/p_2)}$$
$$= \frac{1}{\sigma^2}\left[\frac{\Gamma(1/p_1)(1-1/p_1)\Gamma(1-1/p_1)}{\alpha} + \frac{\Gamma(1/p_2)(1-1/p_2)\Gamma(1-1/p_2)}{1-\alpha}\right]$$
$$= \frac{1}{\sigma^2}\left[\frac{\Gamma(1/p_1)\Gamma(2-1/p_1)}{\alpha} + \frac{\Gamma(1/p_2)\Gamma(2-1/p_2)}{1-\alpha}\right]. \quad ■$$

Remark 10. Note that calculation of $E[\partial^2 \ln f/\partial\mu^2]$ requires $p_i > 1$ to ensure $1 - 1/p_i > 0$ in the domain $(0, +\infty)$ of definition of $\Gamma(x)$. If considering the Gamma function $\Gamma(z)$ defined on the complex plane except $z = 0, -1, -2, \ldots$, however, the information matrix equality may hold for all $p_{01} > 0$ and $p_{02} > 0$ except for points $1/n$ $(n = 2, 3, 4, \ldots)$ of p_{01} and p_{02}. We restrict p_{01} and p_{02} to satisfy $p_{01} > 1/2$ and $p_{02} > 1/2$ since (1) $I(\theta_0)$ and $H(\theta_0)$ are undefined and thus discontinuous at points $1/n$ $(n = 2, 3, 4, \ldots)$; and (2) the information matrix equality has no significance for p_{01} and p_{02} in intervals $(\frac{1}{2n+1}, \frac{1}{2n})$, $n = 1, 2, 3, \ldots$ as $E[\frac{\partial \ln f}{\partial \mu}]^2$ is negative when both p_{01} and p_{02} are in these intervals (see the expression for $E[\frac{\partial \ln f}{\partial \mu}]^2$ and properties of the gamma function). The existence of $E[\partial^2 \ln f/\partial\mu^2]$ at $p_{01} = 1$ and (or) $p_{02} = 1$ is due to the fact that $x\Gamma(x) \to 1$ or $\sin(x)/x \to 1$ as $x \to 0$.

Appendix C

Appendix C establishes consistency and asymptotic normality of MLE of all parameters in AEPD. The following lemma is used in proof of Proposition 7.

Lemma 11. (a) *For any* $\varepsilon > 0$ *there exists a positive constant* M_0, *that may depend on* ε, *such that*

$$|\ln x| \le M_0\left(1 + x^{-\varepsilon} + x^{\varepsilon}\right), \quad \text{for any } x > 0. \tag{52}$$

(b) *For any* (μ^*, q^*) *such that* $|q^* - q| \le d$ *and* $|\mu^* - \mu| \le d$, *the following inequalities hold:*

$$\left(\mu^* - y\right)^{q^*} \le 2 + (\mu + d - y)^{q+d} + (\mu - d - y)^{q-d},$$
$$\text{if } y < \mu - d; \tag{53}$$

$$(\mu^* - y)^{q^*} \le 1 + (\mu + d - y)^{q+d}, \quad \text{if } q^* > 0, y < \mu - d; \tag{54}$$

$$\left(y - \mu^*\right)^{q^*} \le 2 + (y - \mu + d)^{q+d} + (y - \mu - d)^{q-d},$$
$$\text{if } y > \mu + d; \tag{55}$$

$$\left(y - \mu^*\right)^{q^*} \le 1 + (y - \mu + d)^{q+d}, \quad \text{if } q^* > 0, y > \mu + d. \tag{56}$$

(c) *Suppose that* Y *is an AEPD r.v. with density* $f(y \mid \theta_0)$ *defined in* (6), *where* $\theta_0 = (\alpha_0, p_{01}, p_{02}, \mu_0, \sigma_0)$. *Then, for any* $\mu \in R$ *and* $r > -1$, *the following inequality holds:*

$$E\,|Y - \mu|^r \le M_1(\mu, r; \theta_0)\Gamma\left(\frac{1+r}{p_{01}}\right) + M_2(\mu, r; \theta_0)\Gamma\left(\frac{1+r}{p_{02}}\right), \tag{57}$$

where $M_1(\cdot, \cdot; \theta_0)$ *and* $M_2(\cdot, \cdot; \theta_0)$ *are two positive continuous functions.*

Proof. Part (a) is immediate since for any $\varepsilon > 0$, $x^{\varepsilon}|\ln x| \to 0$ as $x \to 0^+$, and $|\ln x|/x^{\varepsilon} \to 0$, as $x \to +\infty$. Part (b) is easy for the cases of $q^* > 0$ and $q^* < 0$; the cases of $|\mu \pm d - y| > 1$ and $|\mu \pm d - y| < 1$ are considered subsequently. Part (c) is proved by using the c_r-inequality (see Loève (1977, p 157)), $|y - \mu_0|^p \ge 2^{1-p}|y - \mu|^p - |\mu_0 - \mu|^p$ for $p \ge 1$, and then change of variable. For $r > -1$,

$$E\,|Y - \mu|^r \le D_1(\mu, \theta_0)\int_{-\infty}^{\mu} |y - \mu|^r \exp\left[-2^{1-p_{01}}C_1(\theta_0)\,|y - \mu|^{p_{01}}\right]dy + D_2(\mu, \theta_0)\int_{\mu}^{+\infty} |y - \mu|^r \exp\left[-2^{1-p_{02}}C_2(\theta_0)\,|y - \mu|^{p_{02}}\right]dy$$

$$= \sum_{i=1}^{2} M_i(\mu, r; \theta_0)\int_0^{+\infty} x^{\frac{1+r}{p_{0i}}-1}e^{-x}dx$$

$$= \sum_{i=1}^{2} M_i(\mu, r; \theta_0)\Gamma\left(\frac{1+r}{p_{0i}}\right),$$

where $C_i(\theta_0)$, $D_i(\mu, \theta_0)$ and $M_i(\mu, r; \theta_0)$ $(i = 1, 2)$ are certain functions depending on parameters in brackets. ■

Proof of Proposition 4 (*Consistency*). The consistency of the MLE $\widehat{\theta}_T$ can be shown by verifying the conditions of Theorem 2.5 in Newey and McFadden (1994), which holds under conditions that are primitive and also quite weak. Condition (ii) of Theorem 2.5, compactness of the parameter set, is ensured by considering compact Θ. Condition (iii) requires that the log-likelihood $\ln f(y \mid \theta)$ be continuous at each $\theta \in \Theta$ with probability one. This holds by inspection. We only need to check the identification condition and dominance condition (conditions (i) and (iv) of Theorem 2.5).

The identification condition implies that if $\theta \ne \theta_0$ then $\Pr\{f(y \mid \theta) \ne f(y \mid \theta_0)\} > 0$. It is sufficient to show that for any given $\theta \in \Theta; \theta \ne \theta_0$ there exists a set of positive probability, $S(\theta)$, such that

$$\ln f(y \mid \theta) \ne \ln f(y \mid \theta_0), \ \text{a.e. } y \in S(\theta). \tag{58}$$

The proof uses the fact that any AEPD random variable Y has positive probability on any interval. If $\mu \ne \mu_0$, say, $\mu > \mu_0$, then for $y \in (\mu_0, \mu]$ the function $\ln f(y \mid \theta)$ is strictly increasing but $\ln f(y \mid \theta_0)$ strictly decreases, so (58) always holds on $(\mu_0, \mu]$. Now suppose $\mu = \mu_0$. We show that (58) is true almost everywhere in $(-\infty, \mu_0]$ (or $(\mu_0, +\infty)$) if $p_1 \ne p_{01}$ (or $p_2 \ne p_{02}$). Suppose $p_2 \ne p_{02}$. Then, for $y \in (\mu_0, +\infty)$, $\ln f(y \mid \theta) = -\ln\sigma - C_2(\theta)(y - \mu_0)^{p_2}$ (since $\mu = \mu_0$) and $\ln f(y \mid \theta_0) = -\ln\sigma_0 - C_2(\theta_0)(y - \mu_0)^{p_{02}}$, where $C_2(\theta) = (\Gamma(1 + 1/p_2)/((1 - \alpha)\sigma))^{p_2}$. Since both functions on $(\mu_0, +\infty)$ are power functions, they intersect at no more than two points, implying that (58) holds for $S(\theta) = (\mu_0, +\infty)$. Similarly, for $\mu = \mu_0$, $p_1 = p_{01}$ and $p_2 = p_{02}$, it is easy to show that (58) holds if $\alpha \ne \alpha_0$ or $\sigma \ne \sigma_0$ (see Newey and McFadden (1994, p. 2126)).

The dominance condition of Theorem 2.5, $E[\sup_{\theta\in\Theta}|\ln f(Y \mid \theta)|] < \infty$, can be verified here by the compactness of parameter set Θ and the boundedness of the $\overline{p}$th absolute moment of a standard AEPD r.v., where $\overline{p}$ is the supremum of p_1 and p_2 in Θ. Since the parameter set Θ is compact, any continuous function of θ is bounded on Θ. By using the c_r-inequality (see Loève (1977), p. 157), i.e., $|a + b|^r \le c_r|a|^r + c_r|b|^r$, with $c_r = 1$ or 2^{r-1} as $0 < r \le 1$ or $r \ge 1$, we have $|\ln f(Y \mid \theta)| \le K_1 + K_2|X|^{\overline{p}}$ for all $\theta \in \Theta$, for K_1 and K_2 positive constants and $X = \sigma_0(Y - \mu_0)$, a standard AEPD r.v. with parameters $(\alpha_0, p_{01}, p_{02})$. Since $E[|X|^{\overline{p}}] < \infty$ by (15) the dominance condition is satisfied. ■

Proof of Proposition 7 (*Asymptotic Normality*). The proof of the asymptotic normality result proceeds by verifying the conditions of Theorem 3 as well as its corollary in Huber (1967). Following the notation of Huber (1967), let $\psi(y, \theta) = \frac{\partial \ln f(y,\theta)}{\partial\theta}$, the score vector, and set

$$\lambda(\theta) = E\psi(y, \theta),\ u(y, \theta, d) \equiv \sup_{\theta^*\in D^*}\left|\psi(y, \theta^*) - \psi(y, \theta)\right|, \tag{59}$$

where $D^* \equiv \{\theta^* \mid |\theta^* - \theta| \le d\}$ and all expectations are with respect to the true underlying distribution $f(y; \theta_0)$ with $\theta_0 = (\alpha_0, p_{01}, p_{02}, \mu_0, \sigma_0)$. The condition N-1 (i.e., for each fixed θ, $\psi(y, \theta)$ is measurable and separable) in Assumption A-1 of Huber (1967) is immediate; conditions (N-2) and (N-4): $\lambda(\theta_0) = 0$ and $E[|\psi(y, \theta_0)|^2] < \infty$, hold by (51) and the fact that ϕ_{ii} in (22) are finite. For the MLE $\widehat{\theta}$, we have $\sum_{t=1}^{T}\psi(y_t, \widehat{\theta}) = 0$; then Equation (27) of Huber (1967) holds. Since consistency has been proved, the only remaining condition is the condition (N-3): there are strictly positive numbers a, b, c, d_0 such that

$$|\lambda(\theta)| \geq a\,|\theta-\theta_0|, \quad \text{for } |\theta-\theta_0| \leq d_0, \tag{60}$$

$$Eu(y,\theta,d) \leq bd, \quad \text{for } |\theta-\theta_0|+d \leq d_0, d \geq 0, \tag{61}$$

$$E[u(y,\theta,d)^2] \leq cd, \quad \text{for } |\theta-\theta_0|+d \leq d_0, d \geq 0, \tag{62}$$

where $|\theta|$ denotes any norm equivalent to the Euclidean norm.

We check condition (61). Separate the location parameter from the other parameters, $\tau = (\alpha, p_1, p_2, \sigma)$, i.e. $\theta = (\tau, \mu)$ and $\theta^* = (\tau^*, \mu^*)$. Then

$$u(y,\theta,d) \leq \sup_{\theta^*\in D^*} \left|\psi(y,\tau^*,\mu^*) - \psi(y,\tau^*,\mu)\right| + \sup_{|\tau^*-\tau|\leq d} \left|\psi(y,\tau^*,\mu) - \psi(y,\tau,\mu)\right|. \tag{63}$$

The bound in (61) is easily verified for the second part in (63), because the location μ is fixed and $\psi(y,\tau,\mu)$ as a function of τ is smooth enough. For the first part in (63), express each element of $\psi(y,\tau,\mu)$ using (48) as:

$$C(\tau) + \left[C_{11}(\tau)\,|\mu-y|^{q_1} + C_{12}(\tau)\,|\mu-y|^{q_1} \times \ln|\mu-y|\right] 1(y<\mu) + \left[C_{21}(\tau)\,|y-\mu|^{q_2} + C_{22}(\tau)\,|y-\mu|^{q_2} \ln|y-\mu|\right] 1(y>\mu), \tag{64}$$

where $(q_1, q_2) = (p_1, p_2)$ or $(q_1, q_2) = (p_1 - 1, p_2 - 1)$; $C(\cdot)$, $C_{ij}(\cdot)$ are continuous functions of $\tau = (\alpha, p_1, p_2, \sigma)$, bounded on compact Θ. We need to show

$$E\left[\sup_{\theta^*\in D^*}\left|(\mu^*-y)^{q_1^*} 1(y<\mu^*) - (\mu-y)^{q_1^*} 1(y<\mu)\right|\right] \leq bd, \tag{65}$$

$$E\left[\sup_{\theta^*\in D^*}\left|(\mu^*-y)^{q_1^*} \ln|\mu^*-y|\, 1(y<\mu^*) - (\mu-y)^{q_1^*} \ln|\mu-y|\, 1(y<\mu)\right|\right] \leq bd. \tag{66}$$

Here we show (66); condition (65) is verified similarly; the counterparts with "$1(y > \mu)$" are similar. Denoting by $\underline{p}$ the infimum of p_i in $\theta \in \Theta$, by the assumption $p_i > 1$ $(i = 1, 2)$, we get $\underline{p} > 1$, $q_1^* \geq \underline{p} - 1 \equiv \underline{q} > 0$. Taking $d_0 < \min\{\underline{q}/2, \frac{1}{3}\}$ and noting that $\left|x^{\underline{q}} \ln x\right|$ is bounded in $(0, 1)$, (66) reduces to

$$E\left[\sup_{\theta^*\in D^*}\left|(\mu^*-y)^{q_1^*} \ln(\mu^*-y) - (\mu-y)^{q_1^*} \ln(\mu-y)\right| \times 1(y<\mu-2d)\right] \leq bd. \tag{67}$$

By the mean-value theorem, (52) and (53), for any (μ^*, q_1^*) with $\left|q_1^* - q_1\right| \leq d$ and $|\mu^* - \mu| \leq d$, for $y < \mu - 2d$ we get

$$\begin{aligned}
&\left|(\mu^*-y)^{q_1^*} \ln(\mu^*-y) - (\mu-y)^{q_1^*} \ln(\mu-y)\right| \\
&= \left|(\widetilde{\mu}-y)^{q_1^*-1}\left\{q_1^* \ln(\widetilde{\mu}-y) + 1\right\}\right| \left|\mu^*-\mu\right| \\
&\leq d\left[(\widetilde{\mu}-y)^{q_1^*-1}\left|\ln(\widetilde{\mu}-y)\right| + (\widetilde{\mu}-y)^{q_1^*-1}\right] \\
&\leq dM_0(\varepsilon)\left[(\widetilde{\mu}-y)^{q_1^*-1} + (\widetilde{\mu}-y)^{q_1^*-1-\varepsilon} + (\widetilde{\mu}-y)^{q_1^*-1+\varepsilon}\right] \\
&\leq dM_0 \sum_{i=1}^{3}\left[2 + (\mu+d-y)^{q_1+d-\delta_i} + (\mu-d-y)^{q_1-d-\delta_i}\right],
\end{aligned} \tag{68}$$

where $\widetilde{\mu}$ is a real number between μ and μ^*, $\delta_1 = 1$, $\delta_2 = 1 + \varepsilon$ and $\delta_3 = 1 - \varepsilon$. Note that $q_1 \pm d - \delta_i > -1$ as long as $\varepsilon < \underline{q}/2$, say $\varepsilon = \underline{q}/4$, because $d \leq d_0$ and $q_1 \geq \underline{q}$. Applying (57) we get the bound in (67) since Θ is compact.

To verify the condition (62), it is sufficient to show that

$$E\left[\sup_{\theta^*\in D^*}\left|(\mu^*-y)^{q_1^*} \ln(\mu^*-y) - (\mu-y)^{q_1^*} \ln(\mu-y)\right| \times 1(y<\mu-2d)\right]^2 \leq cd. \tag{69}$$

For any (μ^*, q_1^*) such that $\left|q_1^* - q_1\right| \leq d$ and $|\mu^* - \mu| \leq d$, we have

$$\begin{aligned}
&\left|(\mu^*-y)^{q_1^*} \ln(\mu^*-y) - (\mu-y)^{q_1^*} \ln(\mu-y)\right| \\
&\leq M_0(\varepsilon) \sum_{i=1}^{3}\left[1 + (\mu+d-y)^{q_1+d-1+\delta_i}\right] 1(y<\mu-2d),
\end{aligned} \tag{70}$$

using (54), δ_i, $\varepsilon < \underline{q}$ in (68); then $q_1 + d - 1 + \delta_i > 0$. Combining (70) with (68) and using the c_r-inequality (see Loève (1977, p 157)) yields

$$\begin{aligned}
&\left[\sup_{\theta^*\in D^*}\left|(\mu^*-y)^{q_1^*} \ln(\mu^*-y) - (\mu-y)^{q_1^*} \ln(\mu-y)\right|\right]^2 \times 1(y<\mu-2d) \\
&\leq dK_0\left\{1 + \sum_{i=1}^{K}\left[(\mu+d-y)^{\xi_i} + (\mu-d-y)^{\eta_i}\right] \times 1\,(y<\mu-2d)\right\},
\end{aligned}$$

where constant $K_0 > 0$ may depend on ε, integer K is $0 < K < 28$, ξ_i and η_i are real numbers greater than -1 when ε is small enough, say $\varepsilon = \underline{q}/4$. Thus, (69) follows from (57) and the assumption of compactness of the parameter space Θ.

A sufficient condition for (60) to hold is that $\lambda(\theta)$ has continuous (partial) derivatives in some neighborhood $O(\theta_0, d_0)$ of θ_0; indeed, this condition and the fact that the Hessian $H(\theta_0)$ is negative definite implies (60). Here we show that if $\lambda_4(\theta) = E\left[\partial \ln f(y,\theta)/\partial\mu\right]$, then $\partial\lambda_4(\theta)/\partial\mu$ is continuous; the continuity of other partial derivatives is easily proved by using Lemma 3.6 of Newey and McFadden (1994), the c_r-inequality and (52)–(57). Note that $\lambda_4(\theta) =$

$$A_1(\tau)E\,|\mu-y|^{p_1-1}\,1(y<\mu) + A_2(\tau)E\,|y-\mu|^{p_2-1}\,1(y>\mu) = \int_0^{+\infty} a(x;\theta)\mathrm{d}x,$$

where $a(x,\theta) \equiv A_1(\tau)x^{p_1-1}f(\mu-x;\theta_0) + A_2(\tau)x^{p_2-1}f(x+\mu;\theta_0)$, A_1 and A_2 are continuously differentiable functions of $\tau = (\alpha, p_1, p_2, \sigma)$, bounded over the compact parameter space Θ, $f(y;\theta_0)$ is the true AEPD density. Let $d_0 > 0$ be small enough that $O(\theta_0, d_0) \equiv \{\theta : |\theta-\theta_0| < d_0\} \subset \Theta$. Then, obviously, $a(x,\theta)$ is continuously differentiable in the neighborhood $O(\theta_0, d_0)$ of θ_0, a.s.; and by the c_r-inequality and compactness of the parameter space Θ,

$$\sup_{|\theta-\theta_0|<d_0}\left|\frac{\partial a(x,\theta)}{\partial\mu}\right| \leq \begin{cases} B_0 x^{\underline{p}-1}, & 0 \leq x \leq 1 \\ B_0 x^{2(\overline{p}-1)}\exp(-B_1 x), & x > 1 \end{cases}$$

where B_0 and B_1 are positive constants that do not depend on θ, $\underline{p} > 1$ and $\overline{p} > 1$ are, respectively, the infimum and supremum of p_i in $\theta \in \Theta$. From Lemma 3.6 of Newey and McFadden (1994) it follows that $\lambda_4(\theta)$ is continuously differentiable with respect to μ in the neighborhood $O(\theta_0, d_0)$ of θ_0. ■

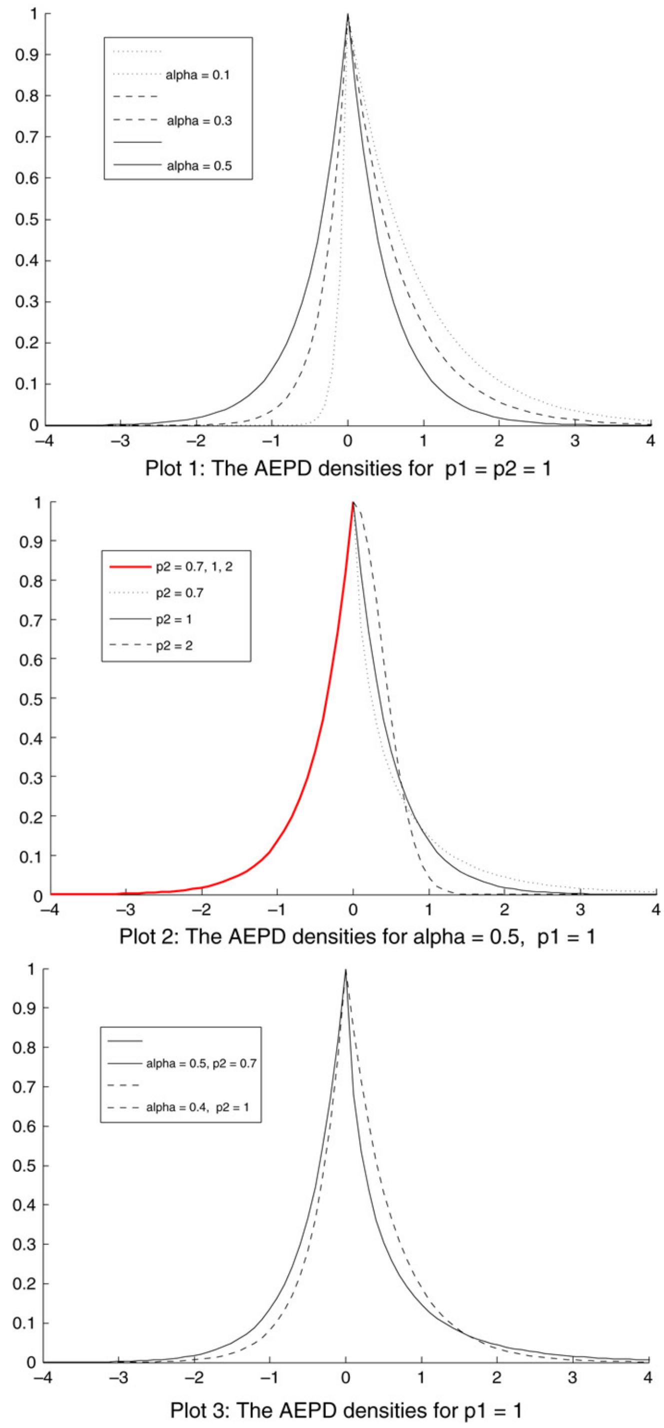

Fig. 1. The AEPD densities for combinations of (α, p_1, p_2).

Appendix D

References

Abramowitz, M., Stegun, I.A., 1970. Handbook of Mathematical Functions. In: National Bureau of Standard, Applied Math. Series, vol. 55. US Department of Commerce.

Agrò, G., 1995. Maximum likelihood estimation for the exponential power function parameters. Comm. Simulation Comput. 24 (2), 523–536.

Anderson, T., Darling, D., 1952. Asymptotic theory of certain "goodness of fit"criteria based on stochastic process. The Annals of Mathematical Statistics 23, 193–212.

Andersen, T.G., Bollerslev, T., Christoffersen, P.F., Diebold, F.X., 2005. Volatility and correlation forecasting. In: Elliott, Graham, Granger, Clive W.J., Timmermann, Allan (Eds.), Handbook of Economic Forecasting. North Holland, Amsterdam.

Andrews, D.F., Bickel, P.J., Hampel, F.R., Huber, P.J., Rogers, W.H., Tukey, J.W., 1972. Robust Estimates of Location. Princeton University Press, Princeton.

Arnold, B.C., Groeneveld, R.A., 1995. Measuring skewness with respect to the mode. Amer. Statist. 49, 34–38.

Artin, E., 1964. The Gamma Function. Holt, Rinehart and Winston, Inc.

Ayebo, A., Kozubowski, T.J., 2004. An asymmetric generalization of Gaussian and Laplace laws. J. Probab. Statist. Sci. 1 (2), 187–210.

Arellano-Valle, R.B., G'omez, H.W., Quintana, F.A., 2005. Statistical inference for a general class of asymmetric distributions. J. Statist. Plann. Inference 128, 427–443.

Azzalini, A., 1986. Further results on a class of distributions which includes the normal ones. Statistica 46, 199–208.

Bawa, V.S., 1975. Optimal rules for ordering uncertain prospects. J. Financ. Econom. 2, 95–121.

Bickel, D.R., 2002. Robust estimators of the mode and skewness of continuous data. Comput. Statist. Data Anal. 39, 153–163.

Bollerslev, T., Chou, R.Y., Kroner, K.F., 1992. ARCH modeling in finance: A selective review of the theory and empirical evidence. J. Econom. 52, 5–59.

Box, G.E.P, Tiao, G.C., 1973. Bayesian Inference in Statistical Analysis. Addison Wesley ed.

Bradley, B.O., Taqqu, M.S., 2003. Financial risk and heavy tails. In: Rachev, S.T. (Ed.), Handbook of Heavy Tailed Distributions in Finance.

Christoffersen, P.F., 2003. Elements of Financial Risk Management. Academic Press.

Christoffersen, P.F., Jacobs, K., Wang, Y., 2005. Volatility components: Affine restrictions and non-normal innovations. Working Paper. McGill University.

DiCiccio, T., Monti, A.C., 2004. Inferential aspects of the skew exponential power distribution. J. Amer. Statist. Assoc. 99 (466), 439–450.

Duan, J.-C., 1999. Conditionally fat-tailed distributions and the volatility smile in options. Working Paper. University of Toronto.

Engle, R.F., Ng, V.K., 1993. Measuring and testing the impact of news on volatility. J. Finance 48, 1749–1778.

Farrell, O.J., Ross, B., 1963. Solved Problems: Gamma and Beta Functions, Legendre Polynomials, Bessel Functions. The Macmillan Company, New York.

Fernandez, C., Osiewalski, J., Steel, M.F.J., 1995. Modeling and inference with υ-distributions. J. Amer. Statist. Assoc. 90 (432), 1331–1340.

Gradshteyn, I.S., Ryzhik, I.M., 1994. Tables of Integrals, Series and Products. Academic Press.

Hsieh, D., 1989. Modeling heteroskedasticity in daily foreign exchange rates. J. Business Econom. Statist. 7, 307–317.

Huber, P.J., 1967. The behavior of maximum likelihood estimates under nonstandard conditions. In: LeCam, L.M., Neyman, J. (Eds.), Proceedings of the Fifth Berkeley Symposium on Mathematical Statistics and Probability. University of California Press, Berkeley.

Hurvich, C., Tsai, C., 1989. Regression and time series model selection in small samples. Biometrika 76, 297–307.

Kagan, A.M., Linnik, Yu.V., Rao, C.R., 1973. Characterization Problems in Mathematical Statistics. Wiley, New York.

Komunjer, I., 2007. Asymmetric power distribution: Theory and applications to risk measurement. J. Appl. Econom. 22, 891–921.

Kotz, S., Kozubowski, T.J., Podgorski, K., 2001. The Laplace Distribution and Generalizations: A Revisit with Applications to Communications, Economics, Engineering, and Finance. Birkhauser, Boston.

Loève, M., 1977. Probability Theory I, 4th ed.

Mineo, A., 1989. The norm-p estimation of location, scale and simple linear regression parameters. In: Lecture Notes in Statistics, Statistical Modelling, vol. 57. Springer-Verlag, pp. 222–233.

Mittnik, S., Paolella, M.S., 2003. Prediction of financial downside-risk with heavy-tailed conditional distributions. In: Rachev, S.T. (Ed.), Handbook of Heavy Tailed Distributions in Finance.

Nelson, D.B., 1991. Conditional heteroskedasticity in asset returns: A new approach. Econometrica 59 (2), 347–370.

Newey, W.K., McFadden, D., 1994. Large Sample Estimation and Hypothesis Testing. In: Engle, R., McFadden, D. (Eds.), Handbook of Econometrics, vol. 4. North-Holland, Amsterdam.

Rachev, S.T., Mittinik, S., 2000. Stable Paretian Models in Finance. Wiley.

Salinas, H.S., Arellano-Valle, R.B., Gez, H.W., 2007. The extended skew exponential power distributions and its derivations. In: Multivariate Skewed Distributions. Comm. Statistics Theory Methods 36 (9), 1673–1689 (special issues).

Schwarz, G., 1978. Estimating the dimension of a model. The Annals of Statistics 6, 461–464.

Subbotin, M.T., 1923. On the law of frequency of error. Matemat. Sbornik 31, 296–301.

Theodossiou, P., 2000. Skewed generalized error distribution of financial assets and option pricing. SSRN Working Paper.

Zhu, D., Zinde-Walsh, V., 2007. Properties and estimation of asymmetric exponential power distribution. McGill and CIREQ Working Paper 13-2007.