JOURNAL OF APPLIED ECONOMETRICS
J. Appl. Econ. **22**: 891–921 (2007)
Published online 23 July 2007 in Wiley InterScience
(www.interscience.wiley.com) DOI: 10.1002/jae.961

ASYMMETRIC POWER DISTRIBUTION: THEORY AND APPLICATIONS TO RISK MEASUREMENT

IVANA KOMUNJER*

Department of Economics, University of California, San Diego, California, USA

SUMMARY

Theoretical literature in finance has shown that the risk of financial time series can be well quantified by their expected shortfall, also known as the tail value-at-risk. In this paper, I construct a parametric estimator for the expected shortfall based on a flexible family of densities, called the asymmetric power distribution (APD). The APD family extends the generalized power distribution to cases where the data exhibits asymmetry. The first contribution of the paper is to provide a detailed description of the properties of an APD random variable, such as its quantiles and expected shortfall. The second contribution of the paper is to derive the asymptotic distribution of the APD maximum likelihood estimator (MLE) and construct a consistent estimator for its asymptotic covariance matrix. The latter is based on the APD score whose analytic expression is also provided. A small Monte Carlo experiment examines the small sample properties of the MLE and the empirical coverage of its confidence intervals. An empirical application to four daily financial market series reveals that returns tend to be asymmetric, with innovations which cannot be modeled by either Laplace (double-exponential) or Gaussian distribution, even if we allow the latter to be asymmetric. In an out-of-sample exercise, I compare the performances of the expected shortfall forecasts based on the APD-GARCH, Skew-*t*-GARCH and GPD-EGARCH models. While the GPD-EGARCH 1% expected shortfall forecasts seem to outperform the competitors, all three models perform equally well at forecasting the 5% and 10% expected shortfall. Copyright © 2007 John Wiley & Sons, Ltd.

Received 26 December 2005; Revised 14 August 2006

1. INTRODUCTION

What is the tail behavior of financial time series, and in particular whether we can quantify it, is a question of fundamental importance in risk management. Ultimately, this question cannot be answered without having an appropriate measure of risk. This paper focuses on a particular risk measure—called the expected shortfall—which has gained considerable interest in the financial community. In financial terms, the expected shortfall—also known as the tail value-at-risk (tail-VaR) or (α-risk)—represents the tail-loss in the market value of a given portfolio, over a given time horizon. In mathematical terms, the expected shortfall is the expected value of (minus) the difference between the portfolio's return and its α-quantile, conditional on this difference being negative.

There exists a long-standing literature on risk assessment in economic models, some of which has emphasized the importance of the expected shortfall for the purposes of risk measurement. Examples include decision-theoretic models of choice under uncertainty (see Feldstein, 1969; Hanoch and Levy, 1969; Bawa, 1978) or models of optimal portfolio choice (see Markowitz,

* Correspondence to: Ivana Komunjer, UCSD Department of Economics, 9500 Gilman Drive, La Jolla, CA 92093-0508, USA. E-mail: komunjer@ucsd.edu

1952; Feldstein, 1969; Bawa and Lindenberg, 1977; Bawa, 1978). More recently, there has been interest in the axiomatic foundations of expected shortfall. Those have been established in papers by Artzner *et al*. (1999) and Follmer and Schied (2002). Finally, the important work by Bassett *et al*. (2004) provided the link between the Choquet expected utility theory and the expected shortfall, thus grounding it within the decision-theoretic framework of models of choice under uncertainty.

Unsurprisingly, the econometrics literature on this popular risk measure has been rapidly growing. Current work offers a variety of approaches to the expected shortfall estimation, ranging from fully parametric (see Aas and Haff, 2006), to semi-parametric (see McNeil and Frey, 2000; Bassett *et al*., 2004), and nonparametric (see Scaillet, 2004; Chen, 2005; Fermanian and Scaillet, 2005; Scaillet, 2005). Despite a strong appeal of the semi- and nonparametric methods—as they make weak assumptions about the true data-generating process—their difficulty lies in the ability to estimate the expected shortfall variance. Indeed, Bassett *et al*.'s (2004) semi-parametric estimator of the expected shortfall can be viewed as a by-product of a standard quantile regression (see Koenker and Bassett, 1978). As such, it inherits all of the difficulties related to the consistent estimation of the asymptotic covariance matrix, typically found in the quantile regression literature (see Buchinsky, 1995; Fitzenberger, 1997; Komunjer, 2005).

In this paper, I construct a fully parametric estimator of the expected shortfall. As already pointed out, the main advantage of this approach over the semi- or nonparametric approaches is that it allows the expected shortfall models not only to be estimated but also easily tested for. Its main drawback, on the other hand, is to impose strong constraints on the shape of the density of interest, and in particular on the distribution tails. For example, a double-exponential (Laplace) assumption forces the density tails to decay exponentially, while a Gaussian assumption implies exponential square decay. Hence, a desirable feature of a fully parametric approach is to be based on a flexible family of densities.[1] Such a family should generalize most commonly used benchmarks, e.g. Gaussian and Laplace, be sufficiently flexible to generate the range of shapes that are of interest in financial applications, e.g. skewness and heavy tails, be of closed form and sufficiently parsimonious to facilitate estimation and testing.

Previous literature contains many flexible distributions that satisfy most of the above features: (1) stable distributions, Pearson family or Tukey-λ family generate a broad range of skewness and kurtosis values, but do not have closed form density functions, hence cannot be estimated via maximum likelihood methods. Particularly interesting flexible densities are obtained as generalizations of Student-t density: (2) a generalized t-distribution and a skewed t-distribution (see Hansen, 1994; Fernandez and Steel, 1998; Giot and Laurent, 2004; Patton, 2004; Kuester *et al*., 2006; Paolella, 2006). They allow for skewness and excess kurtosis and are parsimonious. In general, however, those densities are not log-concave. The generalized t-distribution becomes so as its number of degrees of freedom tends to infinity, in which case it reduces to (3) a generalized power distribution (GPD).[2] This family allows for a flexible tail-decay parameter but does not allow for any asymmetry in the data, which can potentially affect the precision of the corresponding expected shortfall estimates. This drawback is particularly severe in the context of financial return time series, which are known to have nonzero skewness.

[1] The use of flexible parametric forms in order to avoid the disadvantages of nonparametric inference has already been advocated by the literature on partially adaptive estimation, for example (see McDonald and Newey, 1988; McDonald 1991, 1997).

[2] Also known as the exponential power distribution (EPD) or the generalized error distribution (GED). For the limit result on the generalized t-distribution see Johnson *et al*. (1994, p. 422).

DOI: 10.1002/jae

I address the above issues by examining a particular family of distributions that combine the flexible tail decay property of the GPD with the asymmetry. I call this family an asymmetric power distribution (APD) family of densities. While some of the properties of the APD densities—such as their moments, for example—have already been studied in the literature (see Fernandez *et al.*, 1995; Ayebo and Kozubowski, 2003) little attention has been devoted to their risk-related characteristics, such as their quantiles or expected shortfalls.

The first contribution of this paper is to complete the existing results by deriving analytic expressions of the APD quantiles and expected shortfalls. Its second contribution is to develop the asymptotic theory for maximum likelihood estimators (MLEs) of the APD parameters. To the best of my knowledge, the current literature does not provide complete estimation results for the distribution families similar to the APD. I fill this gap by showing that the APD MLE is consistent and asymptotically normal with an asymptotic covariance matrix which equals the inverse of the Fisher information matrix. I moreover construct a consistent estimator of this asymptotic covariance matrix based on the APD score, whose analytic expression I derive in the paper. In particular, this paper provides estimators for the expected shortfalls and their standard errors which are easy to compute.

The remainder of the paper is organized as follows. Section 2 gives a formal definition of an APD density and studies basic properties of random variables which are APD distributed. In Section 3, I derive an analytic expression for the expected shortfall of an APD random variable. Section 4 discusses the simulation of an APD random variable and the maximum likelihood estimation of its parameters. Finally, Section 5 gives an empirical application to several daily financial return series and concludes the paper. Appendix A contains useful lemmas whose proofs can be found in a technical appendix available on the journal's website. Appendix B contains the proofs of the propositions stated in the text.

2. DEFINITION AND BASIC PROPERTIES

The family of distributions studied in this paper combines the flexible tail decay property of the GPD family, measured by a parameter denoted λ, with the asymmetry, quantified by a parameter α, $0 < \alpha < 1$. Hence, it can be viewed as a generalization of the GPD family—which corresponds to the special symmetric case $\alpha = 1/2$—to a broader class of densities that are possibly asymmetric. I therefore call it the asymmetric power distribution (APD) family of densities. A formal definition of a probability density function (pdf) of an APD random variable is as follows.

Definition 1 (APD pdf): *Consider a function* $f : \mathbb{R} \to \mathbb{R}_+^*$, $u \mapsto f(u)$ *such that*

$$f(u) = \begin{cases} \dfrac{\delta_{\alpha,\lambda}^{1/\lambda}}{\Gamma(1+1/\lambda)} \exp\left[-\dfrac{\delta_{\alpha,\lambda}}{\alpha^{\lambda}}|u|^{\lambda}\right], & \text{if } u \leq 0, \\ \dfrac{\delta_{\alpha,\lambda}^{1/\lambda}}{\Gamma(1+1/\lambda)} \exp\left[-\dfrac{\delta_{\alpha,\lambda}}{(1-\alpha)^{\lambda}}|u|^{\lambda}\right], & \text{if } u > 0 \end{cases} \tag{1}$$

where $0 < \alpha < 1$, $\lambda > 0$ *and* $\delta_{\alpha,\lambda} = \dfrac{2\alpha^{\lambda}(1-\alpha)^{\lambda}}{\alpha^{\lambda} + (1-\alpha)^{\lambda}}$. *The function* $f(\cdot)$ *thus defined is a probability density function and any random variable* U *with density* $f(\cdot)$ *is called standard APD.*

DOI: 10.1002/jae

It is easy to verify that: (i) $\forall u \in \mathbb{R}$, $f(u) \geq 0$, and (ii) $\int_{\mathbb{R}} f(u)\mathrm{d}u = 1$, which ensure that $f(\cdot)$ is a probability density.[3] The function $f(\cdot)$ is moreover continuously differentiable on $\mathbb{R}^*$. The parameter λ controls the tail decay whereas α measures the degree of asymmetry.

Note that the probability density in equation (1) can be viewed as a reparametrization of the skewed exponential power distribution (SEPD) proposed by Fernandez *et al.* (1995) or of its generalized version—asymmetric exponential power distribution (AEPD)—proposed by Ayebo and Kozubowski (2003).[4] For example, if V has a standard SEPD density with shape parameters $\gamma = [\alpha^{-1}(1-\alpha)]^{1/2}$ and $q = \lambda$, then $[\alpha(1-\alpha)]^{1/2}(2\delta_{\alpha,\lambda})^{-1/\lambda}V$ has the standard APD density with shape parameters α and λ. While in the parametrizations of SEPD and AEPD, the parameters controlling asymmetry (γ and $\kappa = 1/\gamma$ respectively) have no immediate interpretation, in the APD representation the asymmetry parameter α corresponds to the probability that U be lower than its mode—zero. In other words, under the APD density parametrization (1), α is the portion of the probability mass under $f(\cdot)$ that is left from the mode of U. Hence departures of α from a half directly account for the extent of asymmetry in $f(\cdot)$.

When α equals one half, the APD pdf defined in equation (1) is symmetric around zero. In this important special case $f(\cdot)$ reduces to the standard GPD density.[5] The GPD family, indexed by a single parameter λ, includes distributions that change gradually from short-tailed distributions, for $\infty > \lambda \geq 2$, to fat-tailed ones, when $2 > \lambda > 0$, as the exponent λ decreases. Special cases of the GPD include: uniform ($\lambda = \infty$), Gaussian ($\lambda = 2$) and Laplace ($\lambda = 1$) distributions.

When α is different from one half, the APD pdf is asymmetric. Special cases $\lambda = 1$ and $\lambda = 2$ have already been studied in the literature. They correspond to the asymmetric Laplace distribution,[6] obtained when $\lambda = 1$, and the two-piece normal distribution,[7] obtained when $\lambda = 2$. The original motivation for introducing such distributions was mainly to generalize the simple Laplace (double exponential) and Gaussian cases to situations in which the two halves of the distribution have different averages. Figure 1 plots the standard APD density for fixed values of the tail parameter λ.

One can easily generalize the APD family in order to accommodate for different location and scale parameters, by using the location-scale property of the pdf $f(\cdot)$ in equation (1). For given values of α and λ, such that $0 < \alpha < 1$ and $\lambda > 0$, let X be an APD random variable defined as

$$X = \theta + \phi U \tag{2}$$

[3] Note that $0 < 2\alpha^\lambda(1-\alpha)^\lambda \leq \alpha^{2\lambda} + (1-\alpha)^{2\lambda} < \alpha^\lambda + (1-\alpha)^\lambda$, so $0 < \delta_{\alpha,\lambda} < 1$.

[4] The density of a standard SEPD random variable is given by

$$f(u) = \begin{cases} c\exp[-\frac{1}{2}(\gamma|u|)^q], & \text{if } u \leq 0, \\ c\exp[-\frac{1}{2}(|u|/\gamma)^q], & \text{if } u > 0 \end{cases}$$

where $\gamma, q > 0$ and $c^{-1} = 2^{1/q}\Gamma(1+1/q)(\gamma + 1/\gamma)$ (see Fernandez *et al.*, 1995, Kotz *et al.* 2001, p. 271). AEPD density—with location parameter equal to zero—is simply obtained from the expression above by letting $\kappa = 1/\gamma$ and replacing $2^{1/q}$ (respectively 2) with a scale parameter $\sigma > 0$ (respectively σ^q) (see Ayebo and Kozubowski, 2003).

[5] See Johnson, *et al.* (1994, p. 194) and Kotz *et al.* (2001, p. 219).

[6] Also known as two-piece double exponential distribution. See Govindarajulu (1966), Birnbaum and Mike (1970), Bain and Engelhardt (1973), Sheynin (1977), Jakuszenkow (1979), Lingappaiah (1988), Johnson *et al.* (1994, p. 193), Balakrishnan and Basu (1995), Balakrishnan *et al.* (1996), Kotz *et al.* (2001).

[7] See Johnson *et al.* (1994, pp. 173, 190).

DOI: 10.1002/jae

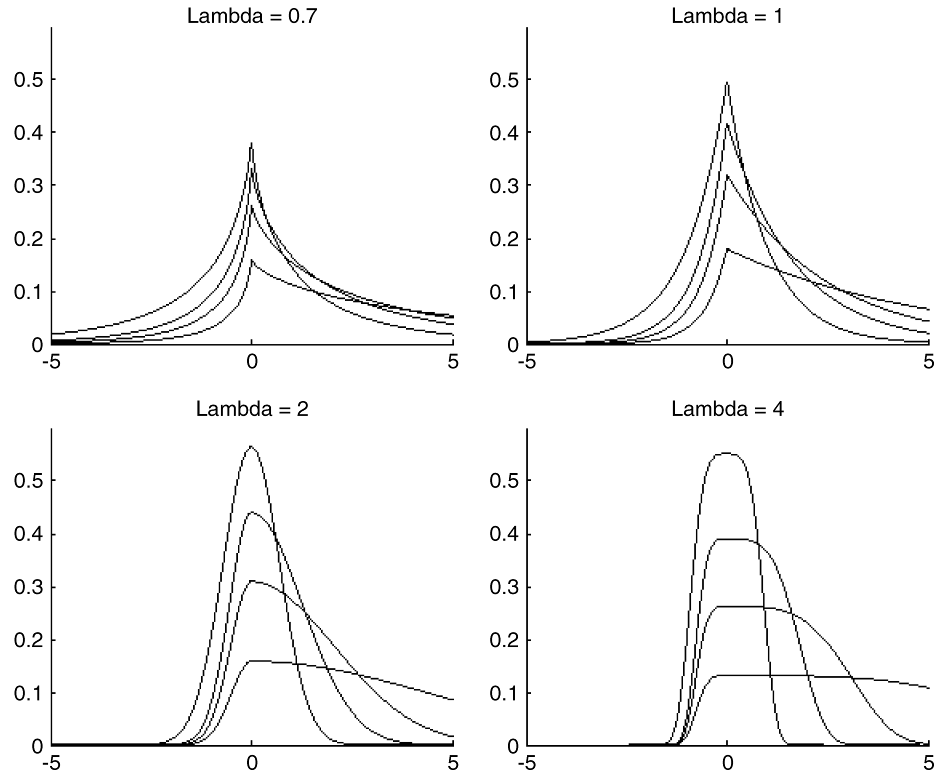

Figure 1. APD density: $X = u$ and $Y = f(u)$ for $\alpha = 0.1, 0.2, 0.3, 0.5$ and $\lambda = 0.7, 1, 2, 4$

with a location parameter $\theta, \theta \in \mathbb{R}$, and a positive scale $\phi, \phi > 0$, so that X has density $f_X(\cdot)$, $f_X(x) = \phi^{-1} f(\phi^{-1}[x - \theta])$, for any $x \in \mathbb{R}$, where $f(\cdot)$ is as defined in equation (1). In the most general case, then, the APD density $f_X(\cdot)$ depends on four parameters α, λ, θ and ϕ, with $0 < \alpha < 1, \lambda > 0, \theta \in \mathbb{R}$ and $\phi > 0$. In Appendix A, I provide the expressions for the cumulative distribution function (cdf) $F(\cdot)$ of a standard APD random variable U and its quantile function $F^{-1}(\cdot)$. The expressions for the cdf $F_X(\cdot)$ of X and its inverse $F_X^{-1}(\cdot)$ are then easily obtained from $F_X(x) = F(\phi^{-1}[x - \theta])$, for any $x \in \mathbb{R}$, and $F_X^{-1}(v) = \theta + \phi^{-1} F^{-1}(v)$, for any $v \in (0, 1)$.

An interesting property of any APD random variable X, resulting from the expressions of $F_X(\cdot)$ and its inverse, is that $\alpha = F_X(\theta)$. In other words, the probability α is such that the mode θ of the APD density $f_X(\cdot)$ corresponds exactly to the α-quantile of X. For example, in the symmetric case where θ is the median of X, the probability α equals one half.

In financial applications, the APD variable of interest ε is often standardized: $E(\varepsilon) = 0$ and $\text{var}(\varepsilon) = 1$. In this case, the pdf of ε, denoted $f_\varepsilon(\cdot)$, is given by

$$f_\varepsilon(z) = \begin{cases} \dfrac{\lambda}{\mu} \dfrac{\Gamma(2/\lambda)}{\Gamma(1/\lambda)^2} \exp\left[-\left(\dfrac{\Gamma(2/\lambda)}{\alpha \Gamma(1/\lambda)} \right)^\lambda \left| \dfrac{z}{\mu} + 1 - 2\alpha \right|^\lambda \right], & \text{if } z \leq -(1 - 2\alpha)\mu, \\ \dfrac{\lambda}{\mu} \dfrac{\Gamma(2/\lambda)}{\Gamma(1/\lambda)^2} \exp\left[-\left(\dfrac{\Gamma(2/\lambda)}{(1 - \alpha)\Gamma(1/\lambda)} \right)^\lambda \left| \dfrac{z}{\mu} + 1 - 2\alpha \right|^\lambda \right], & \text{if } z > -(1 - 2\alpha)\mu \end{cases} \quad (3)$$

where $0 < \alpha < 1, \lambda > 0$ and μ is a positive constant defined as $\mu = \Gamma(2/\lambda)\{\Gamma(3/\lambda)\Gamma(1/\lambda)[1 -$

DOI: 10.1002/jae

$3\alpha + 3\alpha^2] - \Gamma(2/\lambda)^2[1-2\alpha]^2\}^{-1/2}$. The above quantity $-(1-2\alpha)\mu$ corresponds to the α-quantile of the random variable ε, i.e. $F_\varepsilon^{-1}(-(1-2\alpha)\mu) = \alpha$, where $F_\varepsilon(\cdot)$ is the cdf of ε. Note that in the special case $\alpha = 1/2$, the density $f_\varepsilon(\cdot)$ reduces to the standardized GPD density, $f_\varepsilon(z) = [\lambda/(\Lambda\Gamma(1/\lambda)2^{1+1/\lambda})]\exp[-|z/\Lambda|^\lambda/2]$, for $z \in \mathbb{R}$, where $\Lambda = [2^{-2/\lambda}\Gamma(1/\lambda)/\Gamma(3/\lambda)]^{1/2}$.[8]

3. EXPECTED SHORTFALL

I now turn to the study of moments and moment related parameters of an APD random variable. Expressions for non-centered moments of the standard APD random variable U can easily be obtained from those of AEPD random variables.[9] Table I summarizes all the moment results.

For example, the mean and variance of U are given by

$$E(U) = \frac{\Gamma(2/\lambda)}{\Gamma(1/\lambda)}[1-2\alpha]\delta_{\alpha,\lambda}^{-1/\lambda}, \tag{4}$$

$$\mathrm{var}(U) = \frac{\Gamma(3/\lambda)\Gamma(1/\lambda)[1-3\alpha+3\alpha^2] - \Gamma(2/\lambda)^2[1-2\alpha]^2}{[\Gamma(1/\lambda)]^2}\delta_{\alpha,\lambda}^{-2/\lambda} \tag{5}$$

Table I. Moments of a standard APD random variable U

λ	Symmetric case $\alpha = 1/2$	General case $0 < \alpha < 1$
$\lambda > 0$	$E(U) = 0$ $\mathrm{var}(U) = \frac{\Gamma(3/\lambda)}{\Gamma(1/\lambda)}$ $\mathrm{sk}(U) = 0$ $\mathrm{ku}(U) = \frac{\Gamma(5/\lambda)\Gamma(1/\lambda)}{[\Gamma(3/\lambda)]^2}$	see Lemma 4
$\lambda = 1$	$E(U) = 0$	$E(U) = (1-2\alpha)\frac{1}{2\alpha(1-\alpha)}$
(Laplace)	$\mathrm{var}(U) = 2$	$\mathrm{var}(U) = \frac{(1-\alpha)^2+\alpha^2}{[2\alpha(1-\alpha)]^2}$
	$\mathrm{sk}(U) = 0$	$\mathrm{sk}(U) = (1-2\alpha)\frac{2(\alpha^2-\alpha+1)}{[(1-\alpha)^2+\alpha^2]^{3/2}}$
	$\mathrm{ku}(U) = 6$	$\mathrm{ku}(U) = 3\left\{3 - \left[\frac{2\alpha(1-\alpha)}{(1-\alpha)^2+\alpha^2}\right]^2\right\}$
$\lambda = 2$	$E(U) = 0$	$E(U) = -(1-2\alpha)\frac{\sqrt{(1-\alpha)^2+\alpha^2}}{\sqrt{2\pi}\alpha(1-\alpha)}$
(Gaussian)	$\mathrm{var}(U) = \frac{1}{2}$	$\mathrm{var}(U) = \frac{[(3\pi-8)(3\alpha^2-3\alpha+1)+2][(1-\alpha)^2+\alpha^2]}{12\pi\alpha^2(1-\alpha)^2}$
	$\mathrm{sk}(U) = 0$	$\mathrm{sk}(U) = -(1-2\alpha)\frac{\sqrt{54}[(5\pi-16)(5\alpha^2-5\alpha+1)-4]}{5[2+(3\pi-8)(3\alpha^2-3\alpha+1)]^{3/2}}$
	$\mathrm{ku}(U) = 3$	$\mathrm{ku}(U) = \frac{9[(15\pi^2+16\pi-192)(5\alpha^4-10\alpha^3+10\alpha^2-5\alpha+1)]}{5[2+(3\pi-8)(3\alpha^2-3\alpha+1)]^2} - \frac{9[5(7\pi-24)(4\alpha^2-4\alpha+1)-\pi+12]}{5[2+(3\pi-8)(3\alpha^2-3\alpha+1)]^2}$

[8] When $\alpha = 1/2$, $\mu = \Gamma(2/\lambda)/[4\Gamma(3/\lambda)\Gamma(1/\lambda)]^{1/2}$ and $\Lambda = \mu 2^{1+1/\lambda}\Gamma(2/\lambda)/\Gamma(1/\lambda)$ (see Nelson, 1991).

[9] Note that U is also an AEPD random variable with shape λ, location 0, scale $\sigma = \left[\frac{\alpha^\lambda + (1-\alpha)^\lambda}{2\alpha^{\lambda/2}(1-\alpha)^{\lambda/2}}\right]^{1/\lambda}$ and skewness parameter $\kappa = \left[\frac{1-\alpha}{\alpha}\right]^{1/2}$ (see Ayebo and Kozubowski, 2003).

When $\alpha = 1/2$ and $\lambda > 0$, the random variable U has the GPD density.[10] In the asymmetric Laplace case, obtained when $0 < \alpha < 1$ and $\lambda = 1$, the third and fourth centered moments U are bounded: $-2 \leq \text{sk}(U) \leq 2$ and $6 \leq \text{ku}(U) \leq 9$. Note that the bounded values for $\text{sk}(U)$ and $\text{ku}(U)$ make the asymmetric Laplace distribution not well suited for financial applications, in which it is often the case that the series of interest exhibit non-zero skewness and high values of kurtosis. In the symmetric case $\alpha = 1/2$, the random variable U is standard Laplace and we obtain the well-known results: $E(U) = 0$, $\text{var}(U) = 2$, $\text{sk}(U) = 0$ and $\text{ku}(U) = 6$. Figure 2 plots the first four moments of a standard APD random variable U. Expressions for different centered moments of X follow directly from equation (15).

I now turn to an important moment-related parameter of the standard APD random variable U with shape parameters $\alpha, 0 < \alpha < 1$, and $\lambda, \lambda > 0$, which is its $\overline{\alpha}$-expected shortfall, denoted $ES(\overline{\alpha})$. For any probability level $\overline{\alpha}, 0 < \overline{\alpha} < 1, ES(\overline{\alpha})$ is defined as

$$ES(\overline{\alpha}) = E[\overline{q} - U | U \leq \overline{q}] \tag{6}$$

where $\overline{q}$ corresponds to the $\overline{\alpha}$-quantile of U, i.e. $\overline{q} = F^{-1}(\overline{\alpha})$ and $F(\cdot)$ is the cdf of U. In other words, $ES(\overline{\alpha})$ is the expected value of the loss $(\overline{q} - U)$ conditional on this loss being positive,

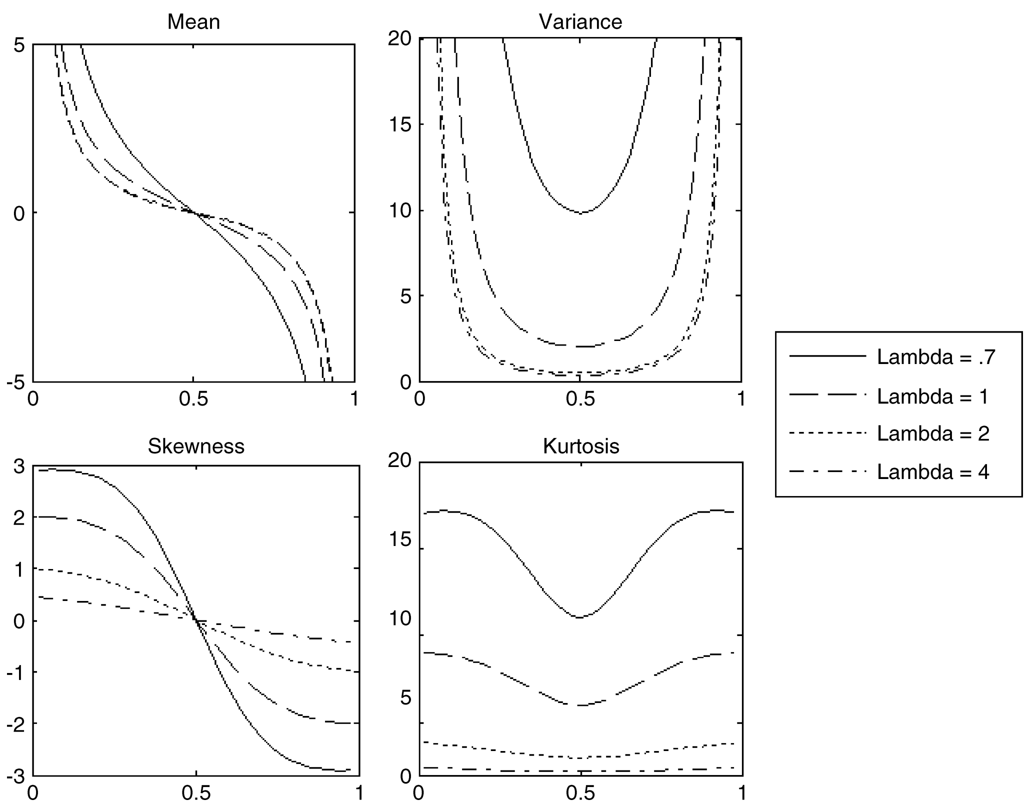

Figure 2. Moments of a standard APD random variable: $0 < \alpha < 1, \lambda = .7, 1, 2, 4$

[10] For the moments of a GPD random variable see Johnson *et al.* (1994, pp. 194–195) and Kotz *et al.* (2001, p. 219).

DOI: 10.1002/jae

i.e. conditional on U being lower than its $\overline{\alpha}$-quantile $\overline{q}$. In the following proposition, I derive the analytic expression for the $\overline{\alpha}$-expected shortfall of U.

Proposition 1 (Expected Shortfall): *For any probability $\overline{\alpha}$ smaller than α, $0 < \overline{\alpha} \leq \alpha$, the $\overline{\alpha}$-expected shortfall of the standard APD random variable U, $ES(\overline{\alpha})$, is given by*

$$ES(\overline{\alpha}) = \frac{\alpha}{\overline{\alpha}} \frac{\alpha}{\delta_{\alpha,\lambda}^{1/\lambda}} \frac{\Gamma(2/\lambda)}{\Gamma(1/\lambda)} \left[1 - I\left(\frac{I^{-1}\left(1 - \frac{\overline{\alpha}}{\alpha}, 1/\lambda\right)}{\sqrt{2}}, 2/\lambda \right) \right] + \overline{q} \tag{7}$$

where $\overline{q}$ is the $\overline{\alpha}$-quantile of U, $\overline{q} = -\left[\frac{\alpha^\lambda}{\delta_{\alpha,\lambda}\sqrt{\lambda}}\right]^{1/\lambda} [I^{-1}(1 - \overline{\alpha}/\alpha, 1/\lambda)]^{1/\lambda}$ when $\overline{\alpha} \leq \alpha$, $I(x, \gamma)$ is Pearson's incomplete gamma function, $I(x, \gamma) = [\Gamma(\gamma)]^{-1} \int_0^{x\sqrt{\gamma}} t^{\gamma-1} \exp(-t)\mathrm{d}t$, $I^{-1}(y, \gamma)$ is its inverse and the constant $\delta_{\alpha,\lambda}$ is as in Definition 1. When $\overline{\alpha} > \alpha > 0$, the $\overline{\alpha}$-expected shortfall of U equals

$$ES(\overline{\alpha}) = \frac{\alpha}{\overline{\alpha}} \frac{\alpha}{\delta_{\alpha,\lambda}^{1/\lambda}} \frac{\Gamma(2/\lambda)}{\Gamma(1/\lambda)} - \frac{(1-\alpha)}{\overline{\alpha}} \frac{(1-\alpha)}{\delta_{\alpha,\lambda}^{1/\lambda}} \frac{\Gamma(2/\lambda)}{\Gamma(1/\lambda)} I\left(\frac{I^{-1}\left(1 - \frac{1-\overline{\alpha}}{1-\alpha}, 1/\lambda\right)}{\sqrt{2}}, 2/\lambda \right) + \overline{q} \tag{8}$$

where $\overline{q}$ is the $\overline{\alpha}$-quantile of U, $\overline{q} = \left[\frac{(1-\alpha)^\lambda}{\delta_{\alpha,\lambda}\sqrt{\lambda}}\right]^{1/\lambda} [I^{-1}(1 - (1-\overline{\alpha})/(1-\alpha), 1/\lambda)]^{1/\lambda}$ when $\overline{\alpha} > \alpha$.

In Figures 3 and 4, I plot the $\overline{\alpha}$-expected shortfall of a standard APD random variable U with different shape parameters α and λ as functions of the probability $\overline{\alpha}$. It is interesting to note that, unlike the quantile function $\overline{\alpha} \mapsto \overline{q} = F^{-1}(\overline{\alpha})$, the function $\overline{\alpha} \mapsto ES(\overline{\alpha})$ is not monotone on (0, 1).

Taking into account the location-scale property of the pdf $f_X(\cdot)$, the above results are easily transposable to any APD random variable X in equation (2). For any $\overline{\alpha}$, $0 < \overline{\alpha} < 1$, the $\overline{\alpha}$-expected shortfall $ES_X(\overline{\alpha})$ of X (defined as the expected value of the loss $(\overline{q}_X - X)$ conditional on X being lower than its $\overline{\alpha}$-quantile $\overline{q}_X$) equals $ES_X(\overline{\alpha}) = \phi ES(\overline{\alpha})$. As previously, one can use equations (7) and (8) to derive the expressions for the $\overline{\alpha}$-expected shortfall $ES_\varepsilon(\overline{\alpha})$ of the standardized APD random variable ε. We have $ES_\varepsilon(\overline{\alpha}) = \phi_\varepsilon ES(\overline{\alpha})$, with $\phi_\varepsilon = \Gamma(1/\lambda)[\Gamma(3/\lambda)\Gamma(1/\lambda)[1 - 3\alpha + 3\alpha^2] - \Gamma(2/\lambda)^2[1-2\alpha]^2]^{-1/2}\delta_{\alpha,\lambda}^{1/\lambda}$.[11]

With the results of Proposition 1 in hand, I now proceed with the estimation of the $\overline{\alpha}$-expected shortfall. To this end, I first establish the asymptotic properties of a maximum likelihood estimator (MLE) for the APD parameters.

[11] Note that we can write $\varepsilon = \theta_\varepsilon + \phi_\varepsilon U$, where the location of ε equals $\theta_\varepsilon = -(1-2\alpha)\mu$ and its scale is $\phi_\varepsilon = \Gamma(1/\lambda)[\Gamma(2/\lambda)]^{-1}\delta_{\alpha,\lambda}^{1/\lambda}\mu$, with μ as in Lemma 3.

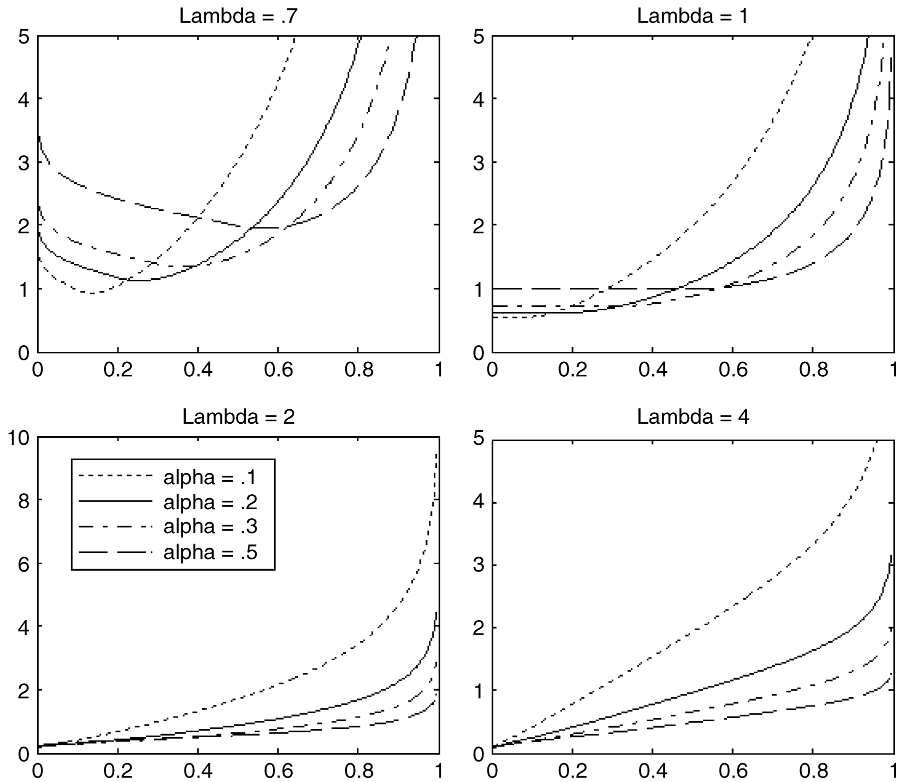

Figure 3. $\overline{\alpha}$-Expected shortfall $ES(\overline{\alpha})$ (as a function of $\overline{\alpha}$) of a standard APD random variable $U(\alpha = 0.1, 0.2, 0.3, 0.5$ and $\lambda = 0.7, 1, 2, 4)$

4. SIMULATION AND MAXIMUM LIKELIHOOD ESTIMATION

In this section I discuss two problems: (1) how to simulate a random variable which is APD distributed; and (2) how to estimate its true parameters. These problems often arise together in Monte Carlo studies, for example.

Similar to the GPD case, random variates from the APD family can be obtained by direct transformation of gamma variates (see Johnson, 1979). For given values of α and λ, $0 < \alpha < 1$ and $\lambda > 0$, the method for generating standard APD random variates is as follows: (1) generate a gamma variate W with shape parameter $1/\lambda$ and pdf $f_W(w) = \Gamma(1/\lambda)^{-1} w^{1/\lambda - 1} \exp(-w)$; (2) divide W by $\delta_{\alpha,\lambda}$ and raise to $1/\lambda$ power, thus obtaining $V = (W/\delta_{\alpha,\lambda})^{1/\lambda}$; (3) generate a random sign variable S equal to $+1$ with probability $(1-\alpha)$ and to -1 with probability α; finally (4) let $U = -\alpha V \cdot \mathbb{I}(S \leq 0) + (1-\alpha) V \cdot \mathbb{I}(S > 0)$.[12] It is straightforward to show that such random variable U has density $f(\cdot)$ as defined in equation (1) and is hence standard APD distributed.

Alternatively, having determined the expressions of the standard APD cdf $F(\cdot)$ and its inverse $F^{-1}(\cdot)$ in Lemmas 1 and 2, respectively, standard APD random variates can be generated by using

[12] The function $\mathbb{I}(\cdot)$ is the standard indicator function, i.e. for any event A, we have $\mathbb{I}(A) = 1$ if A is true and $\mathbb{I}(A) = 0$ otherwise.

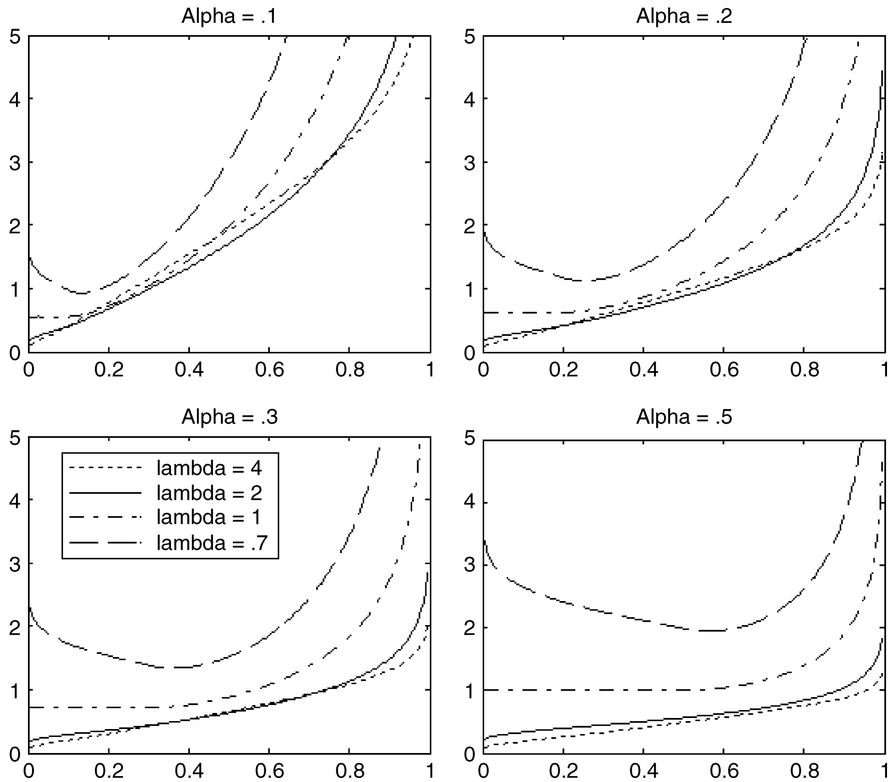

Figure 4. $\overline{\alpha}$-Expected shortfall $ES(\overline{\alpha})$ (as a function of $\overline{\alpha}$) of a standard APD random variable U ($\alpha = 0.1, 0.2, 0.3, 0.5$ and $\lambda = 0.7, 1, 2, 4$)

an inversion method. The inversion method can be summarized as follows: (1) generate a uniform variate V in (0, 1); then (2) let $U = F^{-1}\ (V)$.[13]

I now turn to the problem of estimating the parameters of the APD density $f_X(\cdot)$, which is a function of the asymmetry parameter $\alpha, 0 < \alpha < 1$, the exponent $\lambda, \lambda > 0$, the location $\theta, \theta \in \mathbb{R}$, and the scale $\phi, \phi > 0$. Let β denote the parameter vector, $\beta = (\alpha, \lambda, \theta, \phi)'$. I follow the usual convention and let β_0 be the true value of β which needs to be estimated, $\beta_0 = (\alpha_0, \lambda_0, \theta_0, \phi_0)'$. In this paper, I focus on the MLE for β_0, which I denote $\hat{\beta}_T$.

Recall that for any given α and $\lambda, 0 < \alpha < 1$ and $\lambda > 0$, the APD random variable X in equation (2) has density $f_X(\cdot)$ given by $f_X(x) = \phi^{-1} f(\phi^{-1}[x - \theta])$, for any $x \in \mathbb{R}$, where $f(\cdot)$ is as defined in equation (1). Let then $X_1, \ldots, X_T$ be a random sample from an APD distribution with density $f_X(\cdot)$ parametrized by β, and let $x_1, \ldots, x_T$ be the corresponding observations. The APD normalized log-likelihood $L_T(\beta), L_T(\beta) = T^{-1} \sum_{t=1}^{T} \ln f_X(x_t|\beta)$, takes the form

$$L_T(\beta) = -\ln \phi + \tfrac{1}{\lambda} \ln \delta_{\alpha,\lambda} - \ln \Gamma(1 + 1/\lambda)$$

[13] Note that the first simulation method requires a gamma random number generator that accepts values of the shape parameter greater than zero, while the inversion method only requires a uniform random number generator.

DOI: 10.1002/jae

$$- \frac{\delta_{\alpha,\lambda}}{\phi^{\lambda}} T^{-1} \sum_{t=1}^{T} \left[\frac{|x_t - \theta|^{\lambda}}{\alpha^{\lambda}} \mathbb{I}(x_t \leq \theta) + \frac{|x_t - \theta|^{\lambda}}{(1-\alpha)^{\lambda}} \mathbb{I}(x_t > \theta) \right], \tag{9}$$

where $\delta_{\alpha,\lambda}$ is as in Definition 1, i.e., $\delta_{\alpha,\lambda} = \dfrac{2\alpha^{\lambda}(1-\alpha)^{\lambda}}{\alpha^{\lambda} + (1-\alpha)^{\lambda}}$.

The MLE $\hat{\beta}_T$ is obtained as a solution to the problem $\max_{\beta \in B} L_T(\beta)$ where B is a compact parameter set, $B \subset (0, 1) \times (1/2, +\infty) \times \mathbb{R} \times \mathbb{R}_+^*$. The standard asymptotic normality results for MLEs require that the objective function $L_T(\beta)$ be twice continuously differentiable, which is not the case here. There exist, however, asymptotic normality results for non-smooth functions and I shall hereafter use the one proposed by Newey and McFadden (1994). The basic insight of their approach is that the smoothness condition on the objective function $L_T(\beta)$ can be replaced by the smoothness of its limit, which in the standard maximum likelihood case corresponds to the expectation $L_0(\beta) = E[\ln f_X(X_t|\beta)]$, with the requirement that certain remainder terms are small. Hence, the standard differentiability assumption is replaced by a 'stochastic differentiability' condition, which can then be used to show that the MLE $\hat{\beta}_T$ is consistent and asymptotically normal. This is the result of the following proposition.

Proposition 2 (APD MLE): *Let $X_1, \ldots, X_T$ be a random sample from an APD distribution with an unknown parameter β_0, $\beta_0 \in \mathring{B}$. Then, the MLE $\hat{\beta}_T$ of β_0 is consistent and asymptotically normal:*

$$\sqrt{T}(\hat{\beta}_T - \beta_0) \xrightarrow{d} \mathcal{N}(0, J(\beta_0)^{-1})$$

where $J(\beta)$ is the Fisher information matrix, $J(\beta) = E[(\nabla_\beta \ln f_X(X_t|\beta))(\nabla_\beta \ln f_X(X_t|\beta))']$. Moreover:

$$J_T(\hat{\beta}_T)^{-1} \xrightarrow{p} J(\beta_0)^{-1}$$

where $J_T(\hat{\beta}_T) = T^{-1} \sum_{t=1}^{T} (\nabla_\beta \ln f_X(x_t|\hat{\beta}_T))(\nabla_\beta \ln f_X(x_t|\hat{\beta}_T))'$. An analytic expression of the APD score, $\nabla_\beta \ln f_X(X_t|\beta)$, is provided in Appendix A.

I study the small sample performance of the above MLE, $\hat{\beta}_T$, and of its covariance matrix estimator, $J_T(\hat{\beta}_T)^{-1}$, by conducting a Monte Carlo experiment. In particular, I study the small sample bias of $\hat{\beta}_T$ and the empirical coverage of its 95% confidence interval obtained from $J_T(\hat{\beta}_T)^{-1}$. For a given value of the true parameter $\beta_0 = (\alpha_0, \lambda_0, \theta_0, \phi_0)'$, I generate $N = 10,000$ replications of the sequence $x_1, \ldots, x_T$ from the APD random variable X with density $f_X(\cdot)$.[14]

The parameter α_0 is taken to be equal to 0.1, 0.25 and 0.5, while λ_0 takes the values 1, 2 and 4. The parameters θ_0 and ϕ_0 are held fixed to 0 and 1, respectively, in all of the performed replications. The sample size T is chosen to be 250 and 1000.[15]

For each replicate n, $1 \leq n \leq N$, the true parameter β_0 is estimated by $\hat{\beta}_{T,n} = (\hat{\alpha}_{T,n}, \hat{\lambda}_{T,n}, \hat{\theta}_{T,n}, \hat{\phi}_{T,n})'$. The maximization of the APD log-likelihood $L_T(\beta)$ is done numerically by using Matlab `fmincon` built-in optimization routine. The parameter space B is set to $(0, 1) \times$

[14] The simulations are performed by using a Matlab gamma random number generator with default seed values, which are obtained when the state of the Matlab pseudo-random number generator is set to zero.

[15] Monte Carlo results for the sample sizes $T = 500, 5000$ are reported in Table IA available in the technical appendix.

$(1/2, 20) \times \mathbb{R} \times \mathbb{R}_+^*$.[16] For each of the components of $\hat{\beta}_{T,n}$ I compute the 95% confidence intervals, using the covariance matrix estimator $J_T(\hat{\beta}_{T,n})^{-1}$ defined in Proposition 2. Table II reports the mean value of the MLEs, $\hat{\beta}_T = (\hat{\alpha}_T, \hat{\lambda}_T, \hat{\theta}_T, \hat{\phi}_T)'$ where $\hat{\beta}_T = N^{-1}\sum_{n=1}^{N}\hat{\beta}_{T,n}$, as well as the empirical levels of the corresponding 95% confidence intervals, denoted $(p_{T,\alpha}, p_{T,\lambda}, p_{T,\theta}, p_{T,\phi})'$. As expected, the mean values of the Monte Carlo MLEs converge with the sample size T to the true value β_0 in all of the configurations studied in this experiment. Also, the empirical levels of the 95% confidence intervals for $\hat{\alpha}_{T,n}, \hat{\lambda}_{T,n}, \hat{\theta}_{T,n}$ and $\hat{\phi}_{T,n}$ converge with T to their nominal coverage.

5. EMPIRICAL APPLICATION

I study four financial time series obtained from the Center for Research in Security Prices (CRSP) during a period from 2 January 1990 to 19 May 2006. These consist of daily prices P_t of two indices: S&P500 and NASDAQ; one individual security: Microsoft; and one exchange rate: British pound (BP/USD) expressed in terms of the US dollar. For each series of prices, I construct the series of log-returns, $r_t = 100\ln(P_t/P_{t-1})$, which I adjust to take into account events such as stock splits on individual securities. The original data is split into two periods: an in-sample period ranging from 2 January 1990 to 31 December 2002 (T observations), which is used for estimation, and an out-of-sample period from 2 January 2003 to 19 May 2006 (R observations), used for forecasting.

Table II. MLE sample means and empirical levels ($T = 250$ and 1000)

α_0	λ_0	T	α_T	λ_T	θ_T	φ_T	$p_{T\alpha}$	$p_{T\lambda}$	$p_{T\theta}$	$p_{T\varphi}$
0.1	1	250	0.099	1.021	0.003	0.979	90.97%	95.52%	84.58%	91.61%
0.25	1	250	0.250	1.019	0.004	0.997	91.90%	95.69%	85.30%	93.89%
0.5	1	250	0.500	1.016	−0.001	0.999	92.04%	95.20%	83.43%	94.12%
0.1	2	250	0.094	2.073	−0.033	0.930	91.98%	95.82%	89.99%	91.68%
0.25	2	250	0.248	2.081	−0.003	0.981	93.68%	96.69%	93.48%	93.09%
0.5	2	250	0.499	2.061	−0.004	0.982	94.84%	96.96%	95.00%	95.92%
0.1	4	250	0.085	4.338	−0.111	0.841	99.34%	96.82%	83.78%	99.13%
0.25	4	250	0.240	4.320	−0.032	0.947	92.87%	97.46%	92.29%	91.39%
0.5	4	250	0.499	4.345	−0.003	0.939	94.52%	97.12%	94.32%	96.80%
0.1	1	1000	0.100	1.005	0.001	0.996	94.15%	95.16%	91.52%	93.43%
0.25	1	1000	0.250	1.003	0.000	0.998	93.32%	94.96%	92.38%	94.01%
0.5	1	1000	0.500	1.002	−0.003	0.996	93.68%	95.00%	91.92%	95.08%
0.1	2	1000	0.099	2.014	−0.007	0.986	94.24%	95.83%	94.86%	94.17%
0.25	2	1000	0.250	2.020	0.001	0.997	94.29%	95.82%	94.67%	94.63%
0.5	2	1000	0.500	2.016	0.000	0.997	94.88%	95.08%	94.40%	95.00%
0.1	4	1000	0.097	4.066	−0.022	0.969	94.29%	95.82%	94.62%	94.26%
0.25	4	1000	0.249	4.059	−0.002	0.993	94.96%	96.16%	94.88%	94.64%
0.5	4	1000	0.499	4.062	−0.002	0.986	95.52%	96.24%	95.68%	96.72%

NOTE: Monte Carlo results are obtained with $N = 10,000$ replications of the time series $\{X_t\}$ with $t = 1, \ldots, T$, where X_t's are i.i.d. APD distributed with TDGP values: $\alpha_0, \lambda_0, \theta_0 = 0$ and $\varphi_0 = 1$.

[16] Setting an upper bound on the shape parameter λ (such as 20 in this Monte Carlo experiment) seemed to improve the speed of convergence of the Matlab `fmincon` routine.

DOI: 10.1002/jae

5.1. In-Sample Analysis

For each series of returns r_t, I estimate an APD-GARCH(1,1) model:

$$r_t = \mu + \sigma_t \varepsilon_t, \tag{10}$$

$$\sigma_t^2 = \omega_0 + \omega_1 \sigma_{t-1}^2 + \omega_2 (r_{t-1} - \mu)^2 \tag{11}$$

where the innovations ε_t are assumed to be i.i.d. with APD density $f_\varepsilon(\cdot)$ as in equation (3) so that $E(\varepsilon_t) = 0$ and $E(\varepsilon_t^2) = 1$ (see Engle, 1982; Bollerslev, 1986). The parameter vector $\omega = (\omega_0, \omega_1, \omega_2)'$ satisfies $\omega_0 > 0, 0 < \omega_1, \omega_2 < 1$ and $0 < \omega_1 + \omega_2 < 1$, which are the standard stationarity and invertibility conditions. Given that the innovations ε_t are standardized, their density is parametrized by only two parameters α, $0 < \alpha < 1$, and λ, $\lambda > 0$.

Table III reports the first four unconditional moments of the returns r_t. A quick glance at Table III reveals that all series exhibit high values of kurtosis, ranging from 5.65 (BP/USD) to 138.58 (Microsoft). Skewness of the return series is generally negative.

There are two approaches to estimating the parameters of the APD-GARCH(1,1) model (10)–(11). First is a 'one-step' method which constructs maximum likelihood estimates for the parameters α and λ of the conditional APD density of the innovations along with the parameters μ and ω of the GARCH model. The MLE $\hat{\beta}_T = (\hat{\mu}_T, \hat{\omega}'_T, \hat{\alpha}_T, \hat{\lambda}_T)'$ is obtained by maximizing the APD log-likelihood $L_T(\beta) = T^{-1} \sum_{t=1}^{T} [-\ln \sigma_t + \ln f_\varepsilon(\varepsilon_t|\beta)]$, with σ_t and ε_t as defined by the GARCH(1,1) model (10)–(11), and where $f_\varepsilon(\cdot)$ is as defined in equation (3). Sufficient conditions for the consistency and asymptotic normality of the MLE in the APD-GARCH(1,1) model (10)–(11) can be found in Straumann (2005).[17]

Second is a 'two-step' method in which the estimation is performed sequentially. In the first step, ω and μ are estimated by using a Gaussian quasi-maximum likelihood estimator (QMLE) $(\tilde{\omega}'_T, \tilde{\mu}_T)'$, which, under standard regularity conditions, is consistent and asymptotically normal (see Bollerslev and Wooldridge, 1992). In the second step, estimated values $\tilde{\omega}_T$ and $\tilde{\mu}_T$ are used to construct the residuals $\tilde{\varepsilon}_t = \tilde{\sigma}_t^{-1}(r_t - \tilde{\mu}_T)$, with $\tilde{\sigma}_t^2 = \tilde{\omega}_{0T} + \tilde{\omega}_{1T}\tilde{\sigma}_{t-1}^2 + \tilde{\omega}_{2T}(r_{t-1} - \tilde{\mu}_T)^2$, whose distribution parameters (under the APD assumption) are then consistently estimated by an APD MLE $(\tilde{\alpha}_T, \tilde{\lambda}_T)'$. While both approaches yield consistent estimates for the APD-GARCH(1,1) parameters in (10)–(11), consistent standard errors are more difficult to obtain in the 'two-step' case, as they often require the use of subsampling or bootstrap methods (see Aas and Haff, 2006).

Table III. Descriptive statistics for the unconditional distribution of returns (and their residuals)

	Mean		Variance		Skewness		Kurtosis		
Return series (residuals)	r_t	ε_t	r_t	ε_t	r_t	ε_t	r_t	ε_t	T
S&P500 composite index	0.03	−0.02	1.11	1.01	−0.12	−0.40	6.81	5.22	3278
NASDAQ composite index	0.03	−0.03	2.61	1.01	−0.01	−0.50	8.55	4.61	3278
Microsoft	0.09	−0.01	7.14	1.00	−5.21	0.13	138.58	4.72	3272
BP/USD exchange rate	0.00	0.01	0.33	1.01	0.24	0.06	5.65	4.75	3263

NOTE: T is the number of observations.

[17] Similar to the i.i.d. case, studied in detail in Section 4, these will involve checking the validity of various moment conditions (for details, see Theorems 6.1.1 and 6.3.3 in Straumann, 2005).

DOI: 10.1002/jae

In this paper, I estimate the parameters μ, ω, α and λ by using the 'one-step' approach.[18] Table IV reports the parameter estimates $(\hat{\mu}_T, \hat{\omega}'_T, \hat{\alpha}_T, \hat{\lambda}_T)'$ and their consistent standard errors which are obtained from $J_T(\hat{\beta}_T)^{-1}$, where $J_T(\hat{\beta}_T) = T^{-1}\sum_{t=1}^{T}[\nabla_\beta(-\ln\hat{\sigma}_t + \ln f_\varepsilon(\varepsilon_t|\hat{\beta}_T))]$ $[\nabla_\beta(-\ln\hat{\sigma}_t + \ln f_\varepsilon(\varepsilon_t|\hat{\beta}_T))]'$, and $\hat{\sigma}_t^2 = \hat{\omega}_{0T} + \hat{\omega}_{1T}\hat{\sigma}_{t-1}^2 + \hat{\omega}_{2T}(r_{t-1} - \hat{\mu}_T)^2$ are the conditional variance estimates.

As can be seen from Table IV, estimated values $\hat{\alpha}_T$ of the asymmetry parameter α range from 0.46 (Microsoft) to 0.59 (NASDAQ). A simple Wald test of the restriction $\alpha = 1/2$ shows that in two cases out of four (NASDAQ and Microsoft), the value of α is significantly different from 1/2 (with probability 95%). In other words, the residuals for those series are found to be asymmetric. Another interesting finding is that in all four cases the estimated values $\hat{\lambda}_T$ of the exponent λ are found to be significantly different from both 1 and 2, thus invalidating the assumptions that the innovations ε_t are double-exponential (Laplace) or normally distributed. Note that this conclusion holds even if we allow for asymmetry in the density of ε_t.

I use the above estimated values to construct the residuals $\hat{\varepsilon}_t = \hat{\sigma}_t^{-1}(r_t - \hat{\mu}_T)$. Table III reports the first four unconditional moments of $\hat{\varepsilon}_t$'s (in addition to those of the returns r_t). The unconditional distribution of the residuals is skewed and leptokurtic, which tends to reject the assumption that the residuals are GPD distributed. Moreover, one can reject the assumption that the latter are Laplace distributed since in all four cases their kurtosis lies outside the interval [6,9].[19]

Finally, using the results of Proposition 1, for any $\overline{\alpha}$, $0 < \overline{\alpha} < 1$, I am able to compute an estimate $\widehat{ES}_\varepsilon(\overline{\alpha})$ for the unconditional $\overline{\alpha}$-expected shortfall for each of the four series of innovations, which I plot in Figure 5. For the purposes of risk management, the quantities of interest are the conditional $\overline{\alpha}$-expected shortfalls of returns r_t, which by using the location-scale property of conditional heteroskedasticity models such as GARCH(1,1) in (10)–(11), can be estimated at each point in

Table IV. APD MLE of the GARCH(1,1) model

Return series	ω_0	ω_1	ω_2	μ	α	λ
S&P500 composite index	0.00	0.94	0.06	0.04	0.51	1.38
	(0.00)	(0.01)	(0.01)	(0.01)	(0.01)	(0.05)
	[2.15]	[87.52]	[5.42]	[3.10]	[40.84]	[30.19]
NASDAQ composite index	0.01	0.90	0.10	0.08	0.59	1.55
	(0.00)	(0.02)	(0.02)	(0.02)	(0.01)	(0.05)
	[3.45]	[58.90]	[6.44]	[4.68]	[54.11]	[31.07]
Microsoft	0.23	0.87	0.09	0.14	0.46	1.50
	(0.07)	(0.02)	(0.02)	(0.04)	(0.01)	(0.02)
	[3.36]	[35.56]	[5.79]	[3.81]	[40.43]	[70.24]
BP/USD exchange rate	0.00	0.93	0.07	−0.01	0.52	1.31
	(0.00)	(0.04)	(0.03)	(0.01)	(0.02)	(0.08)
	[0.00]	[21.01]	[2.17]	[−0.65]	[25.11]	[17.18]

NOTE: APD MLE for the GARCH(1,1) model: $r_t = \mu + \sigma_t\varepsilon_t$, where $\sigma_t^2 = \omega_0 + \omega_1\sigma_{t-1}^2 + \omega_2(r_{t-1} - \mu)^2$ and ε_t is APD distributed with $E(\varepsilon_t) = 0$ and $\mathrm{var}(\varepsilon_t) = 1$. Consistent standard errors () and t statistics [] are in parentheses. Values of the APD log-likelihood: S&P500, 1.310; NASDAQ, 1.625; Microsoft, 2.224, BP/USD, 0.777.

[18] As a comparison, I report the estimation results from a 'two-step' approach in Tables IIA and IIIA of the technical appendix.

[19] Figure 1A in the technical appendix shows histograms of the residuals $\hat{\varepsilon}_t$ together with the fitted APD (standardized) probability density with parameters $(\hat{\alpha}_T, \hat{\lambda}_T)$ for each of the four series studied.

DOI: 10.1002/jae

time t using $\widehat{ES}_{r_t}(\overline{\alpha}) = \hat{\sigma}_t \widehat{ES}_\varepsilon(\overline{\alpha})$. Figure 6 plots $\widehat{ES}_{r_t}$ (5%) as a function of time t for each of the four series of returns r_t during the period 12 March 1993–3 March 1995 (500 observations).

Alternatively, we can represent $\widehat{ES}_{r_t}(\overline{\alpha})$ as a function of the probability $\overline{\alpha}$ but keeping a particular time τ fixed. For example, in Figure 7, I set the date $\tau = 16$ March 1993 (Tuesday) and plot $\widehat{ES}_{r_\tau}(\overline{\alpha})$ as a function of $\overline{\alpha}$ for each of the six series studied. Note that the time τ can be chosen arbitrarily within the sample.

In order to compare the riskiness of two different securities with returns r_{1t} and r_{2t} we need to compare, at each point in time t, their conditional expected shortfalls $ES^{(1)}_{r_t}(\overline{\alpha})$ and $ES^{(2)}_{r_t}(\overline{\alpha})$, for

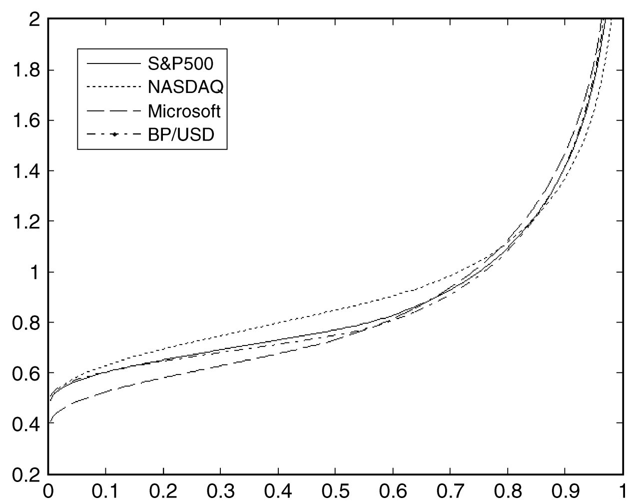

Figure 5. In-sample estimates for the unconditional $\overline{\alpha}$-expected shortfall of $\varepsilon_t : \widehat{ES}_\varepsilon(\overline{\alpha})(0 < \overline{\alpha} < 1)$

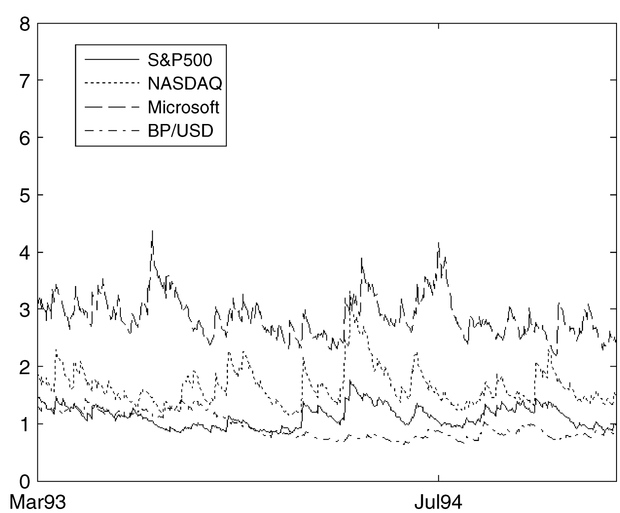

Figure 6. In-sample estimates of the conditional 5% expected shortfalls of r_t (12 March 1993–3 March 1995)

DOI: 10.1002/jae

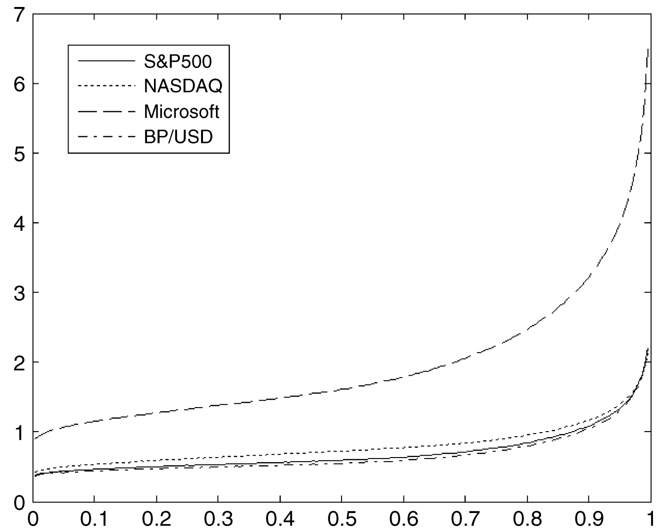

Figure 7. In-sample estimates for the conditional $\overline{\alpha}$-expected shortfalls of r_t, $0 < \overline{\alpha} < 1$ (obtained for Tuesday 16 March 1993)

all values of $\overline{\alpha}$, $0 < \overline{\alpha} < 1$. If for every $\overline{\alpha}$ the conditional $\overline{\alpha}$-expected shortfall of the first security is smaller than the conditional $\overline{\alpha}$-expected shortfall of the second security, i.e., if $ES^{(1)}_{r_t}(\overline{\alpha}) \leq ES^{(2)}_{r_t}(\overline{\alpha})$, then any risk-averse investor with distorted perception of the true probabilities (optimistic or pessimistic) prefers holding the first security to holding the second one (see Bassett *et al.*, 2004).

Figure 8 plots the 95% confidence intervals for $\widehat{ES}_{r_t}(\overline{\alpha})$, $0 < \overline{\alpha} < 1$, for each of the four series studied. Those confidence intervals were obtained by using the delta method in which $\nabla_\beta \widehat{ES}_{r_t}(\overline{\alpha})$ was replaced by a numerical gradient.[20] The ith component $\partial\widehat{ES}_{r_t}(\overline{\alpha})/\partial\beta^{(i)}$ of the gradient was computed over a grid of points $\hat{\beta}^{(i)}_T + \delta s^{(i)}_T$, in which $s^{(i)}_T$ is the standard error of $\hat{\beta}^{(i)}_T$ obtained from $J_T(\hat{\beta}_T)^{-1}$ and the step δ takes values $-0.8, -0.6, \ldots, 0.8$.

As can be seen from Figure 8, the two stock indices (S&P500 and NASDAQ) and exchange rate (BP/USD) clearly dominate the individual stock (Microsoft). However, there is no clear ranking among the securities within those two groups in terms of their riskiness as measured by their respective expected shortfalls.

5.2. Out-of-Sample Analysis

I now turn to the out-of-sample evaluation of the $\overline{\alpha}$-expected shortfall one-step-ahead forecasts obtained from the APD-GARCH(1,1) model (10)–(11). I adopt a fixed forecasting scheme, which means that all forecasts depend on the same set of parameters estimated in-sample, i.e., over the first T observations. In other words, at each out-of-sample date τ, $T+1 \leq \tau \leq T+R$, I compute

[20] An alternative method of obtaining the confidence intervals for the shortfall estimate $\widehat{ES}_{r_t}(\overline{\alpha})$ is via Monte Carlo simulations. For example, the 95% confidence interval can be computed as the empirical 95% coverage interval for $\widehat{ES}_{r_t}(\overline{\alpha})$ obtained in the Monte Carlo simulations of the parameter $(\hat{\mu}_T, \hat{\omega}'_T, \hat{\alpha}_T, \hat{\lambda}_T)'$ in which the latter is assumed to be normally distributed with mean and covariance matrix as estimated via APD-GARCH MLE (see Table IV). Table IVA in the technical appendix compares the standard errors obtained using the delta method with those using the Monte Carlo simulation approach.

DOI: 10.1002/jae

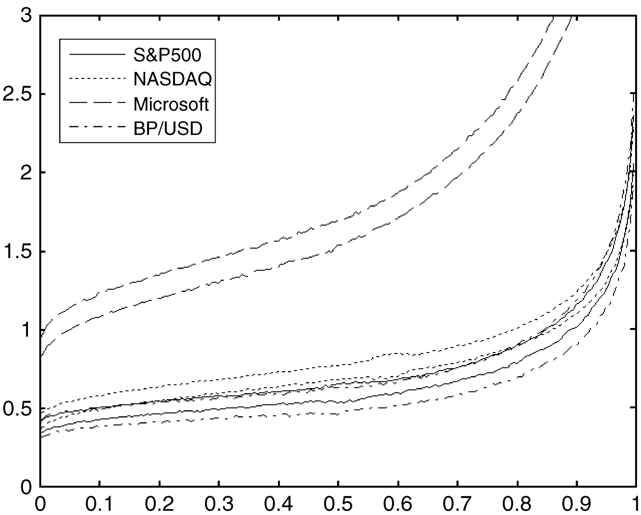

Figure 8. 95% confidence intervals for the in-sample estimates of the conditional $\overline{\alpha}$-expected shortfalls of r_t, $1 < \overline{\alpha} < 1$ (obtained for Tuesday 16 March 1993)

the one-step-ahead $ES(1\%)$, $ES(5\%)$ and $ES(10\%)$ forecasts recursively. Note that $\widehat{ES}_\varepsilon(\overline{\alpha})$ remains constant over the out-of-sample period as it only depends on the parameter estimates $\hat{\beta}_T$ obtained in-sample.

In addition to the ES (1%), ES (5%) and ES (10%) forecasts from the APD-GARCH(1,1) model (10)–(11), I also consider the ES (1%), ES (5%) and ES (10%) forecasts originated from the following two models: an asymmetric Student-t GARCH(1,1) model and a GPD-EGARCH(1,1) model.

The asymmetric (skewed) Student-t GARCH(1,1) specification is similar to the one in (10)–(11) except that the innovations ε_t are now assumed to be i.i.d. with an asymmetric (skewed) Student-t distribution whose density

$$f_\varepsilon(z) = \begin{cases} \dfrac{2}{\gamma + 1/\gamma} bc \left[1 + \dfrac{(\gamma(bz+a))^2}{\nu - 2}\right]^{-(\nu+1)/2}, & \text{if } z \le -a/b, \\ \dfrac{2}{\gamma + 1/\gamma} bc \left[1 + \dfrac{((bz+a)/\gamma)^2}{\nu - 2}\right]^{-(\nu+1)/2}, & \text{if } z \ge -a/b \end{cases} \tag{12}$$

is parametrized by ν, $2 < \nu < \infty$, and γ, $\gamma > 0$, and where the constants a, b and c are given by $a = \dfrac{\Gamma((\nu-1)/2)\sqrt{\nu-2}}{\sqrt{\pi}\Gamma(\nu/2)}(\gamma - 1/\gamma)$, $b^2 = (\gamma^2 + 1/\gamma^2 - 1) - a^2$ and $c = \dfrac{\Gamma((\nu+1)/2)}{\sqrt{\pi(\nu-2)}\Gamma(\nu/2)}$ (see Hansen, 1994; Fernandez and Steel, 1998; Giot and Laurent, 2004; Patton, 2004; Kuester *et al*., 2006; Paolella, 2006).[21] The innovations density in equation (12) is standardized so $E(\varepsilon_t) = 0$ and $E(\varepsilon_t^2) = 1$. Table V reports the in-sample parameter estimates of the skewed Student-t GARCH(1,1) model.

[21] Note that the expression of Hansen's (1994) skewed Student-t density follows from equation (12) by a simple reparametrization which sets the degrees of freedom parameter $\eta = \nu$ and the skewness parameter $\lambda = (\gamma^2 - 1)/(\gamma^2 + 1)$.

DOI: 10.1002/jae

Table V. Skewed Student-*t* MLE of the GARCH(1,1) model

Return series	ω_0	ω_1	ω_2	μ	$\ln(\gamma)$	ν
S&P500 composite index	0.00	0.94	0.06	0.05	−0.02	6.69
	(0.00)	(0.01)	(0.01)	(0.01)	(0.02)	(0.63)
	[1.94]	[106.86]	[6.25]	[3.85]	[−1.25]	[10.68]
NASDAQ composite index	0.01	0.91	0.09	0.08	−0.17	9.63
	(0.00)	(0.01)	(0.01)	(0.02)	(0.02)	(1.15)
	[2.63]	[69.42]	[6.81]	[5.21]	[−7.31]	[8.38]
Microsoft	0.23	0.87	0.09	0.15	0.07	8.18
	(0.07)	(0.02)	(0.02)	(0.04)	(0.02)	(1.00)
	[3.46]	[35.83]	[5.01]	[3.59]	[3.17]	[8.20]
BP/USD exchange rate	0.00	0.94	0.06	−0.01	0.01	4.90
	(0.00)	(0.01)	(0.01)	(0.01)	(0.02)	(0.02)
	[1.84]	[80.67]	[5.31]	[−0.94]	[0.33]	[266.67]

NOTE: Skewed Student-*t* MLE for the GARCH(1,1) model: $r_t = \mu + \sigma_t\varepsilon_t$, where $\sigma_t^2 = \omega_0 + \omega_1\sigma_{t-1}^2 + \omega_2(r_{t-1} - \mu)^2$ and ε_t is skew Student-*t* distributed with $E(\varepsilon_t) = 0$ and $\mathrm{var}(\varepsilon_t) = 1$. Consistent standard errors () and t statistics [] are in parentheses. Values of the skew Student-*t* log-likelihood: S&P500, 1.308; NASDAQ, 1.623; Microsoft, 2.220, BP/USD, 0.770.

The second competing model that I consider is a GPD-EGARCH(1,1) model:

$$r_t = \mu + \sigma_t\varepsilon_t, \tag{13}$$

$$\ln\sigma_t^2 = \omega_0 + \omega_1\ln\sigma_{t-1}^2 + \omega_2(r_{t-1} - \mu) + \omega_3|r_{t-1} - \mu| \tag{14}$$

where ε_t is assumed to be i.i.d. with GPD density that is standardized (see Nelson, 1991). Note that a GPD density with shape parameter λ is a special case of an APD density with same shape λ but whose asymmetry parameter α has been set to $\alpha = 1/2$. Though the distribution of the innovations ε_t in the GPD-EGARCH(1,1) model is symmetric, the term $\omega_2(r_{t-1} - \mu) + \omega_3|r_{t-1} - \mu|$ in the EGARCH equation (14) allows the conditional variance σ_t^2 to respond asymmetrically to increases and decreases in returns r_{t-1}. Table VI reports the in-sample parameter estimates of the GPD-EGARCH(1,1) model.

In order to evaluate the accuracy of the expected shortfall forecasts, I use a test similar to that proposed by McNeil and Frey (2000). Let $q_t(\overline{\alpha})$ denote the conditional $\overline{\alpha}$-quantile of the returns r_t, which in the case of conditional heteroskedasticity models such as the one in (10)–(11) can also be written as $q_t(\overline{\alpha}) = \mu + \sigma_t q_\varepsilon(\overline{\alpha})$, where $q_\varepsilon(\overline{\alpha})$ is the unconditional $\overline{\alpha}$-quantile of the innovation term ε_t. The test for accuracy of forecasts $\widehat{ES}_{r_t}(\overline{\alpha})$ is then based on the variable

$$R_t(\overline{\alpha}) = \frac{[q_t(\overline{\alpha}) - r_t] - ES_{r_t}(\overline{\alpha})}{\sigma_t}$$

which in the models considered here reduces to

$$R_t(\overline{\alpha}) = (q_\varepsilon(\overline{\alpha}) - \varepsilon_t) - ES_\varepsilon(\overline{\alpha})$$

Under correct specification, the variables $R_t(\overline{\alpha})$ are i.i.d. and such that $E[R_t(\overline{\alpha})|\varepsilon_t \le q_\varepsilon(\overline{\alpha})] = 0$.

For each of the models studied, I construct the $\overline{\alpha}$-quantile and $\overline{\alpha}$-expected shortfall forecasts, denoted $\hat{q}_\varepsilon(\overline{\alpha})$ and $\widehat{ES}_\varepsilon(\overline{\alpha})$, respectively, using the analytic expressions for $q_\varepsilon(\overline{\alpha})$ and $ES_\varepsilon(\overline{\alpha})$, in

DOI: 10.1002/jae

Table VI. GPD MLE of the EGARCH(1,1) model

Return series	ω_0	ω_1	ω_2	ω_3	μ	λ
S&P500 composite index	−0.09	0.94	−0.10	0.11	0.03	1.40
	(0.01)	(0.01)	(0.01)	(0.02)	(0.01)	(0.03)
	[−6.61]	[109.80]	[−7.02]	[6.93]	[2.78]	[53.48]
NASDAQ composite index	−0.09	0.92	−0.04	0.11	0.10	1.50
	(0.01)	(0.01)	(0.01)	(0.02)	(0.02)	(0.05)
	[−6.42]	[76.14]	[−4.76]	[6.68]	[5.53]	[29.30]
Microsoft	0.04	0.90	−0.03	0.07	0.07	1.52
	(0.02)	(0.02)	(0.01)	(0.01)	(0.02)	(0.04)
	[2.38]	[40.05]	[−4.52]	[4.71]	[3.76]	[42.72]
BP/USD exchange rate	−0.16	0.94	0.00	0.20	0.00	1.31
	(0.01)	(0.01)	(0.02)	(0.02)	(0.01)	(0.04)
	[−10.75]	[150.01]	[0.14]	[11.33]	[−0.48]	[32.10]

NOTE: GPD MLE for the EGARCH(1,1) model: $r_t = \mu + \sigma_t \varepsilon_t$, where $\ln \sigma_t^2 = \omega_0 + \omega_1 \ln \sigma_{t-1}^2 + \omega_2 (r_{t-1} - \mu) + \omega_3 |r_{t-1} - \mu|$, and ε_t is GPD distributed with $E(\varepsilon_t) = 0$, and $\text{var}(\varepsilon_t) = 1$. Consistent standard errors () and t statistics [] are in parentheses. Values of the GPD log-likelihood: S&P500, 1.304; NASDAQ, 1.640; Microsoft, 2.222, BP/USD, 0.774.

which the density parameters are set to be equal to the previously obtained MLEs (see Tables IV, V and VI).[22] In the case of the APD-GARCH(1,1) and GPD-EGARCH(1,1) models $q_\varepsilon(\overline{\alpha})$ and $ES_\varepsilon(\overline{\alpha})$ are computed directly using the results of Lemma 2 and Proposition 1, respectively. In the case of the asymmetric (skewed) Student-t GARCH(1,1) model, the analytic expression for $q_\varepsilon(\overline{\alpha})$ can be found in Patton (2004). Using straightforward algebra, I further show that, for quantiles $q_\varepsilon(\overline{\alpha}) \leq -a/b$, the $\overline{\alpha}$-expected shortfall of a (standardized) asymmetric (skewed) Student-t random variable with density defined in equation (12), equals

$$ES_\varepsilon(\overline{\alpha}) = q_\varepsilon(\overline{\alpha}) + \frac{2}{\overline{\alpha}(\gamma^2+1)} \left\{ \frac{c\gamma - 2}{\gamma\gamma - 1} \left[1 + \frac{(\gamma(bz+a))^2}{\nu - 2} \right]^{-(\nu-1)/2} + aF_\nu(\gamma(bz+a)) \right\}$$

where $F_\nu(\cdot)$ is the cdf of a (standardized) Student-t random variable with ν degrees of freedom. With $\hat{q}_\varepsilon(\overline{\alpha})$ and $\widehat{ES}_\varepsilon(\overline{\alpha})$ in hand, I then construct the out-of-sample residuals $\hat{R}_t(\overline{\alpha}) = (\hat{q}_\varepsilon(\overline{\alpha}) - \hat{\varepsilon}_t) - \widehat{ES}_\varepsilon(\overline{\alpha})$, and their out-of-sample conditional average $\overline{R}_t(\overline{\alpha}) = R^{-1} \sum_{\tau=T+1}^{T+R} \hat{R}_t(\overline{\alpha}) \mathbb{I}(\hat{\varepsilon}_t \leq \hat{q}_\varepsilon(\overline{\alpha}))$.

In Table VII I report the values of $\overline{R}_t(\overline{\alpha}), \widehat{ES}_\varepsilon(\overline{\alpha})$ and of the empirical coverages $\hat{P}(\overline{\alpha}) = R^{-1} \sum_{\tau=T+1}^{T+R} \mathbb{I}(\hat{\varepsilon}_t \leq \hat{q}_\varepsilon(\overline{\alpha}))$ for each of the three forecasting models employed and each of the four series studied. As can be seen from Table VII, the out-of-sample empirical findings are mixed. The standard t-test rejects (at the 5% level) the null hypothesis $H_0 : E[R_t(\overline{\alpha}) | \varepsilon_t \leq q_\varepsilon(\overline{\alpha})] = 0$ in favor of the alternative $H_1 : E[R_t(\overline{\alpha}) | \varepsilon_t \leq q_\varepsilon(\overline{\alpha})] \neq 0$ twice in the case of APD-GARCH(1,1) *ES* (1%) forecasts (NASDAQ and BP/USD) and three times in the case of skewed Student-t GARCH(1,1) *ES* (1%) forecasts (S&P500, NASDAQ and BP/USD). In all other cases, the null hypothesis of the correct specification of the tail residuals cannot be rejected.

It is worth pointing out that this expected shortfall evaluation procedure à la McNeil and Frey (2000) is developed along similar lines to the binomial test of conditional quantile forecasts.

[22] Figure 2A in the technical appendix plots the out-of-sample *ES* (5%) forecasts for S&P500 daily returns obtained from the three competing models.

DOI: 10.1002/jae

Table VII. Out-of-Sample ES forecast evaluation

Return series	R(1%)	R(5%)	R(10%)	$ES_\varepsilon(1\%)$	$ES_\varepsilon(5\%)$	$ES_\varepsilon(10\%)$	P(1%)	P(5%)	P(10%)
APD-GARCH(1, 1)									
S&P500	−0.26	−0.13	−0.16	0.51	0.57	0.60	0.50%	4.70%	12.30%
	[−1.18]	[−0.39]	[−0.39]						
NASDAQ	−0.29	−0.22	−0.20	0.51	0.58	0.63	0.30%	4.50%	12.00%
	[−4.20]	[−0.70]	[−0.53]						
Microsoft	1.88	0.38	0.19	0.43	0.49	0.52	0.60%	3.00%	6.00%
	[0.97]	[0.26]	[0.17]						
BP/USD	−0.45	−0.13	−0.11	0.52	0.57	0.60	0.40%	4.50%	10.10%
	[−6.85]	[−0.46]	[−0.27]						
Skew-t-GARCH(1.1)									
S&P500	−0.44	−0.19	−0.16	0.69	0.62	0.61	0.50%	5.40%	13.20%
	[−1.96]	[−0.52]	[−0.38]						
NASDAQ	−0.44	−0.29	−0.24	0.71	0.67	0.68	0.30%	4.70%	12.30%
	[−5.08]	[−0.91]	[−0.62]						
Microsoft	1.71	0.36	0.13	0.56	0.53	0.54	0.60%	3.00%	6.50%
	[0.90]	[0.25]	[0.12]						
BP/USD	−0.83	−0.23	−0.16	0.85	0.68	0.64	0.20%	4.60%	10.60%
	[−40.13]	[−0.81]	[−0.41]						
GPD-EGARCH(1, 1)									
S&P500	−0.21	−0.21	−0.18	0.49	0.55	0.58	0.50%	4.40%	10.70%
	[−0.95]	[−0.61]	[−0.47]						
NASDAQ	−0.17	−0.10	−0.08	0.46	0.52	0.56	0.80%	5.80%	13.50%
	[−0.88]	[−0.27]	[−0.19]						
Microsoft	1.30	0.38	0.10	0.45	0.52	0.56	0.70%	2.50%	5.70%
	[0.69]	[0.26]	[0.09]						
BP/USD	−0.24	−0.11	−0.06	0.53	0.58	0.61	1.00%	7.90%	16.30%
	[−1.34]	[−0.28]	[−0.13]						

NOTE: *t* statistics [] are in parentheses.

Hence, it is subject to the same drawbacks as the binomial test: (1) it assumes that under the null hypothesis the events $\mathbb{I}(\hat{\varepsilon}_t \leq \hat{q}_\varepsilon(\overline{\alpha}))$ are i.i.d. while empirical evidence shows they exhibit clustering; and (2) perhaps even more importantly, it does not take into account the estimation error in conditional quantile and expected shortfall forecasts. In the case of conditional quantile forecasts, this results in a test which generally has a limited ability to distinguish among alternative hypotheses and thus has low power, even in moderately large samples, as shown by Kupiec (1995) and Lopez (1997).

It is interesting to note, however, that the rejections of the null appear to occur when the unconditional expected shortfall estimates $ES_\varepsilon(\overline{\alpha})$ are high, and hence potentially overestimated. For example, in the case of ES (1%) forecasts for NASDAQ daily returns, both the APD and asymmetric (skewed) Student-t assumption lead to values of $ES_\varepsilon(1\%)$ which are higher (0.51 and 0.71, respectively) than under the GPD assumption (0.46). An overestimation in $ES_\varepsilon(\overline{\alpha})$ automatically leads to negative values of $\overline{R}_t(\overline{\alpha})$ and the changes in estimated conditional variances $\hat{\sigma}_t^2$ do not seem to be able to counter this effect. Another interesting finding is that for low values of probabilities $\overline{\alpha}$ (such as 1%, 5% and 10% studied here), low and potentially underestimated values of $ES_\varepsilon(\overline{\alpha})$ apparently lead to fewer rejections of the null hypothesis. For example, in the case of ES (1%) forecasts for Microsoft daily returns, all three specifications produce values of $\overline{R}_t(\overline{\alpha})$ that are positive (1.88, 1.71 and 1.30, respectively), yet none is found to be significantly different from zero at the 5% level.

DOI: 10.1002/jae

6. CONCLUSION

This paper introduces a new family of probability distributions—called the asymmetric power distribution (APD)—that generalizes the GPD distributions to cases where the density function is asymmetric. I derive analytic expressions for the quantiles and expected shortfalls of APD distributed random variables, show how to simulate APD variates, and finally how to estimate—via maximum likelihood—its four parameters α, λ, θ and ϕ. Moreover, I propose a consistent estimator of the covariance matrix of the MLE based on the APD score, whose analytic expression is derived in the paper. The small sample properties of the MLE and its covariance matrix estimator are studied through a Monte Carlo experiment.

I further apply my theoretical result to the conditional expected shortfall estimation and forecasting. Using daily returns on four financial market series, I found their innovations to be asymmetric—with asymmetry parameter α significantly different from one half—and with the exponent parameter λ comprised within [1.31, 1.55] (and significantly different from 1 and 2, which invalidates Laplace and Gaussian assumptions). Moreover, by computing the 95% confidence intervals for the conditional 5% expected shortfalls of the daily returns, I find that the two stock indices (S&P500 and NASDAQ) and the BP/USD exchange rate clearly dominate (in-sample) the individual stock (Microsoft) in terms of their riskiness.

In an out-of-sample forecasting exercise, I compare the performance of three competing models: APD-GARCH (1,1), skewed Student-*t* GARCH (1,1) and GPD-EGARCH (1,1). The out-of-sample empirical coverage results suggest that the APD-GARCH (1,1) model outperforms its competitors at forecasting the conditional 5% and 10% value-at-risk. While all three models perform well at forecasting the conditional 5% and 10% expected shortfalls, GPD-EGARCH (1,1) dominates the other two models in the case of the conditional 1% expected shortfall forecasts. This interesting finding seems to suggest that—at least in the case of (extreme) tails—modeling the asymmetry in the conditional variance is perhaps even more important than in the unconditional distribution of the innovations, when forecasting the expected shortfall out-of-sample.

APPENDIX A: PROPERTIES OF AN APD RANDOM VARIABLE

Lemma 1 (APD cdf) *For given values of α and λ, $0 < \alpha < 1$ and $\lambda > 0$, let U be a standard APD random variable with pdf $f(\cdot)$ as defined in equation (1). For any $u \in \mathbb{R}$, the cumulative distribution function $F(\cdot)$ of U then equals*

$$F(u) = \begin{cases} \alpha\left[1 - I\left(\dfrac{\delta_{\alpha,\lambda}}{\alpha^{\lambda}}\sqrt{\lambda}|u|^{\lambda},\, 1/\lambda\right)\right], & if\ u \leq 0, \\ 1-(1-\alpha)\left[1 - I\left(\dfrac{\delta_{\alpha,\lambda}}{(1-\alpha)^{\lambda}}\sqrt{\lambda}|u|^{\lambda},\, 1/\lambda\right)\right], & if\ u > 0 \end{cases}$$

where $\delta_{\alpha,\lambda}$ is as in Definition 1 and $I(x,\gamma)$ is Pearson's incomplete gamma function, $I(x,\gamma) = [\Gamma(\gamma)]^{-1}\int_0^{x\sqrt{\gamma}} t^{\gamma-1}\exp(-t)\,dt$.[23]

[23] Note that in the special case of an asymmetric Laplace distribution ($\lambda = 1$), the cdf F above simplifies to $F(u) = \alpha\exp[2(1-\alpha)u]$, if $u \leq 0$, and $1-(1-\alpha)\exp(-2\alpha u)$, if $u > 0$ (see, e.g., Johnson *et al.*, 1994, p. 193).

Lemma 2 (APD quantiles): *For any probability v, $v \in (0, 1)$, the v-quantile of the standard APD random variable U, $F^{-1}(v)$, equals*

$$F^{-1}(v) = \begin{cases} -\left[\dfrac{\alpha^{\lambda}}{\delta_{\alpha,\lambda}\sqrt{\lambda}}\right]^{1/\lambda} \cdot \left[I^{-1}\left(1 - \frac{v}{\alpha}, 1/\lambda\right)\right]^{1/\lambda}, & if\ v \leq \alpha, \\ \left[\dfrac{(1-\alpha)^{\lambda}}{\delta_{\alpha,\lambda}\sqrt{\lambda}}\right]^{1/\lambda} \cdot \left[I^{-1}\left(1 - \dfrac{1-v}{1-\alpha}, 1/\lambda\right)\right]^{1/\lambda}, & if\ v > \alpha \end{cases}$$

where $\delta_{\alpha,\lambda}$ is as in Definition 1 and $I^{-1}(y, \gamma)$ is the inverse of the Pearson's incomplete gamma function, i.e., $x = I^{-1}(y, \gamma)$ is equivalent to $y = I(x, \gamma)$. In particular, $F^{-1}(\alpha) = 0$.[24]

Lemma 3 (Standardized a-Quantile density): *Let ε be an APD random variable which is standardized, i.e., $E[\varepsilon] = 0$ and $var(\varepsilon) = 1$. Then the density of ε, denoted $f_{\varepsilon}(\cdot)$, is given by*

$$f_{\varepsilon}(z) = \begin{cases} \dfrac{\lambda}{\mu}\dfrac{\Gamma(2/\lambda)}{\Gamma(1/\lambda)^2} \exp\left[-\left(\dfrac{\Gamma(2/\lambda)}{\alpha\Gamma(1/\lambda)}\right)^{\lambda} |\frac{z}{\mu} + 1 - 2\alpha|^{\lambda}\right], & if\ z \leq -(1-2\alpha)\mu, \\ \dfrac{\lambda}{\mu}\dfrac{\Gamma(2/\lambda)}{\Gamma(1/\lambda)^2} \exp\left[-\left(\dfrac{\Gamma(2/\lambda)}{(1-\alpha)\Gamma(1/\lambda)}\right)^{\lambda} |\frac{z}{\mu} + 1 - 2\alpha|^{\lambda}\right], & if\ z > -(1-2\alpha)\mu \end{cases}$$

where $0 < \alpha < 1, \lambda > 0$ and μ is a positive constant defined as $\mu = \Gamma(2/\lambda)\{\Gamma(3/\lambda)\Gamma(1/\lambda)[1 - 3\alpha + 3\alpha^2] - \Gamma(2/\lambda)^2[1-2\alpha]^2\}^{-1/2}$.

Lemma 4 (APD moments): *For any $r \in \mathbb{N}$ we have*

$$E(U^r) = \frac{\Gamma((1+r)/\lambda)}{\Gamma(1/\lambda)} \frac{(1-\alpha)^{1+r} + (-1)^r \alpha^{1+r}}{\delta_{\alpha,\lambda}^{r/\lambda}} \tag{15}$$

Lemma 5 (APD score): *If X is a four-parameter APD random variable in (2) with density $f_X(x) = \phi^{-1} f(\phi^{-1}[x - \theta])$, where $f(\cdot)$ is as defined in equation (1), then its score, $s_t = \nabla \ln f_X(X_t)$, is given by*

$$s_{t,\alpha} = \frac{\partial}{\partial\alpha} \ln f_X(X_t) = \frac{\partial\delta_{\alpha,\lambda}}{\partial\alpha}\left\{\frac{1}{\lambda\delta_{\alpha,\lambda}} - \frac{|X_t - \theta|^{\lambda}}{\phi^{\lambda}}\left[\frac{1}{\alpha^{\lambda}}\mathbb{I}(X_t \leq \theta) + \frac{1}{(1-\alpha)^{\lambda}}\mathbb{I}(X_t > \theta)\right]\right\}$$
$$+ \lambda\delta_{\alpha,\lambda}\frac{|X_t - \theta|^{\lambda}}{\phi^{\lambda}}\left[\frac{1}{\alpha^{\lambda+1}}\mathbb{I}(X_t \leq \theta) - \frac{1}{(1-\alpha)^{\lambda+1}}\mathbb{I}(X_t > \theta)\right]$$

[24] As previously, in the asymmetric Laplace case, the quantile function F^{-1} simplifies to $F^{-1}(v) = -[2(1-\alpha)]^{-1}\ln(\alpha/v)$, if $v \leq \alpha$, and $(2\alpha)^{-1}\ln((1-\alpha)/(1-v))$, if $v > \alpha$.

$$s_{t,\lambda} = \frac{\partial}{\partial\lambda}\ln f_X(X_t) = \frac{1}{\lambda^2}\left[-\ln\delta_{\alpha,\lambda} + \frac{\lambda}{\delta_{\alpha,\lambda}}\frac{\partial\delta_{\alpha,\lambda}}{\partial\lambda} + \Psi(1+1/\lambda)\right]$$
$$- \left[\frac{\partial\delta_{\alpha,\lambda}}{\partial\lambda} + \delta_{\alpha,\lambda}\ln\frac{|X_t-\theta|}{\phi}\right]\frac{|X_t-\theta|^\lambda}{\phi^\lambda}\left[\frac{1}{\alpha^\lambda}\mathbb{I}(X_t \le \theta) + \frac{1}{(1-\alpha)^\lambda}\mathbb{I}(X_t > \theta)\right]$$
$$+ \delta_{\alpha,\lambda}\frac{|X_t-\theta|^\lambda}{\phi^\lambda}\left[\frac{\ln\alpha}{\alpha^\lambda}\mathbb{I}(X_t \le \theta) + \frac{\ln(1-\alpha)}{(1-\alpha)^\lambda}\mathbb{I}(X_t > \theta)\right]$$
$$s_{t,\theta} = \frac{\partial}{\partial\theta}\ln f_X(X_t) = \lambda\frac{|X_t-\theta|^{\lambda-1}}{\phi^\lambda}\left[\frac{\delta_{\alpha,\lambda}}{(1-\alpha)^\lambda} - 2\cdot\mathbb{I}(X_t \le \theta)\right]$$
$$s_{t,\phi} = \frac{\partial}{\partial\phi}\ln f_X(X_t) = -\tfrac{1}{\phi}\left\{1 - \lambda\delta_{\alpha,\lambda}\frac{|X_t-\theta|^\lambda}{\phi^\lambda}\left[\frac{1}{\alpha^\lambda}\mathbb{I}(X_t \le \theta) + \frac{1}{(1-\alpha)^\lambda}\mathbb{I}(X_t > \theta)\right]\right\},$$

where $\delta_{\alpha,\lambda} = \dfrac{2\alpha^\lambda(1-\alpha)^\lambda}{\alpha^\lambda + (1-\alpha)^\lambda}$, $\dfrac{\partial\delta_{\alpha,\lambda}}{\partial\alpha} = \lambda\delta_{\alpha,\lambda}\left[\dfrac{(1-2\alpha)}{\alpha(1-\alpha)} - \dfrac{\alpha^{\lambda-1} - (1-\alpha)^{\lambda-1}}{\alpha^\lambda + (1-\alpha)^\lambda}\right]$, $\dfrac{\partial\delta_{\alpha,\lambda}}{\partial\lambda} = \delta_{\alpha,\lambda}\left\{\ln[\alpha(1-\alpha)] - \dfrac{\alpha^\lambda\ln\alpha + (1-\alpha)^\lambda\ln(1-\alpha)}{\alpha^\lambda + (1-\alpha)^\lambda}\right\}$ *and* $\Psi(1+1/\lambda) = \dfrac{\Gamma'(1+1/\lambda)}{\Gamma(1+1/\lambda)}$ *is a digamma function.*

APPENDIX B: PROOFS

Proof of Proposition 1: For a given probability $\overline{\alpha}$, $0 < \overline{\alpha} \le \alpha$, let $\overline{q}$ denote the $\overline{\alpha}$-quantile of U, i.e., $\overline{q} = F^{-1}(\overline{\alpha})$. We then have, for any $u \in \mathbb{R}$, $\Pr(U \le u | U \le \overline{q}) = \Pr(U \le u, U \le \overline{q})/\Pr(U \le \overline{q})$, so that $\Pr(U \le u | U \le \overline{q}) = 1$, if $u \ge \overline{q}$ and $\Pr(U \le u | U \le \overline{q}) = F(u)/\overline{\alpha}$, otherwise. Hence, the $\overline{\alpha}$-expected shortfall of U equals

$$E(\overline{q} - U | U \le \overline{q}) = \tfrac{1}{\overline{\alpha}}\int_{-\infty}^{\overline{q}}(\overline{q} - u)f(u)\mathrm{d}u \tag{16}$$

Recall that $\overline{\alpha} \le \alpha$ so that $\overline{q} = F^{-1}(\overline{\alpha}) \le F^{-1}(\alpha) = 0$. In that case equation (16) becomes

$$E(\overline{q} - U | U \le \overline{q}) = \frac{\overline{\alpha}^{-1}\delta_{\alpha,\lambda}^{1/\lambda}}{\Gamma(1+1/\lambda)}\int_{-\infty}^{\overline{q}}(\overline{q} - u)\exp\left[-\frac{\delta_{\alpha,\lambda}}{\alpha^\lambda}(-u)^\lambda\right]\mathrm{d}u$$
$$= \frac{\overline{\alpha}^{-1}\delta_{\alpha,\lambda}^{1/\lambda}}{\Gamma(1+1/\lambda)}\int_0^{+\infty} v\exp\left[-\frac{\delta_{\alpha,\lambda}}{\alpha^\lambda}(v - \overline{q})^\lambda\right]\mathrm{d}v$$

where I have set $v = \overline{q} - u$. Note that we have $u \le \overline{q} \le 0$ so that $v \ge 0 \ge \overline{q}$. Now let $s = \frac{\delta_{\alpha,\lambda}}{\alpha^\lambda}(v - \overline{q})^\lambda$ so that $v = \left[\frac{\alpha^\lambda}{\delta_{\alpha,\lambda}}s\right]^{1/\lambda} + \overline{q}$ and $dv = \frac{1}{\lambda}\left[\frac{\alpha^\lambda}{\delta_{\alpha,\lambda}}\right]^{1/\lambda}s^{1/\lambda-1}ds$. The integral above becomes $E(\overline{q} - U | U \le \overline{q}) = \frac{1}{\overline{\alpha}}\frac{\alpha}{\lambda\Gamma(1+1/\lambda)}\int_b^{+\infty}\left(\frac{\alpha}{\delta_{\alpha,\lambda}^{1/\lambda}}s^{1/\lambda} + \overline{q}\right)s^{1/\lambda-1}\exp(-s)\mathrm{d}s$, where $b =$

DOI: 10.1002/jae

$\frac{\delta_{\alpha,\lambda}}{\alpha^{\lambda}}(-\overline{q})^{\lambda}$. Hence

$$E(\overline{q}-U|U\leq\overline{q})=\frac{\overline{\alpha}^{-1}\alpha}{\lambda\Gamma(1+1/\lambda)}\left[\frac{\alpha}{\delta_{\alpha,\lambda}^{1/\lambda}}\int_{b}^{+\infty}s^{2/\lambda-1}\exp(-s)\mathrm{d}s+\overline{q}\int_{b}^{+\infty}s^{1/\lambda-1}\exp(-s)\mathrm{d}s\right]$$
$$=\frac{\alpha}{\overline{\alpha}}\left\{\frac{\alpha}{\delta_{\alpha,\lambda}^{1/\lambda}}\frac{\Gamma(2/\lambda)}{\Gamma(1/\lambda)}\left[1-I\left(\frac{\delta_{\alpha,\lambda}}{\alpha^{\lambda}}(-\overline{q})^{\lambda}\sqrt{\lambda/2},2/\lambda\right)\right]+\overline{q}\left[1-I\left(\frac{\delta_{\alpha,\lambda}}{\alpha^{\lambda}}(-\overline{q})^{\lambda}\sqrt{\lambda},1/\lambda\right)\right]\right\}$$

Recall from Lemma 2 that $\overline{q}=-\left[\frac{\alpha^{\lambda}}{\delta_{\alpha,\lambda}\sqrt{\lambda}}\right]^{1/\lambda}\left[I^{-1}\left(1-\frac{\overline{\alpha}}{\alpha},1/\lambda\right)\right]^{1/\lambda}$ so that $E(\overline{q}-U|U\leq\overline{q})=\frac{\alpha}{\overline{\alpha}}\frac{\alpha}{\delta_{\alpha,\lambda}^{1/\lambda}}\frac{\Gamma(2/\lambda)}{\Gamma(1/\lambda)}\left[1-I\left(I^{-1}\left(1-\frac{\overline{\alpha}}{\alpha},1/\lambda\right)/\sqrt{2},2/\lambda\right)\right]+\overline{q}$, which shows that (7) holds. In the case where $\overline{\alpha}>\alpha>0$ we have $\overline{q}>0$ and similar computations to that above show that in this case

$$E(\overline{q}-U|U\leq\overline{q})=\frac{\alpha}{\overline{\alpha}}\frac{\alpha}{\delta_{\alpha,\lambda}^{1/\lambda}}\frac{\Gamma(2/\lambda)}{\Gamma(1/\lambda)}-\frac{(1-\alpha)}{\overline{\alpha}}\frac{(1-\alpha)}{\delta_{\alpha,\lambda}^{1/\lambda}}\times$$
$$\frac{\Gamma(2/\lambda)}{\Gamma(1/\lambda)}I\left(I^{-1}\left(1-\frac{1-\overline{\alpha}}{1-\alpha},1/\lambda\right)/\sqrt{2},2/\lambda\right)+\overline{q}$$

where now $\overline{q}=\left[\frac{(1-\alpha)^{\lambda}}{\delta_{\alpha,\lambda}\sqrt{\lambda}}\right]^{1/\lambda}\left[I^{-1}\left(1-\frac{1-\overline{\alpha}}{1-\alpha},1/\lambda\right)\right]^{1/\lambda}$ from Lemma 2. This completes the proof of Proposition 1.

Proof of Proposition 2: I start by showing that $\hat{\beta}_T$, obtained as a solution to the problem $\max_{\beta\in B}L_T(\beta)$ with $L_T(\beta)$ as defined in equation (9), is a consistent estimate of β_0. In order to do so, I use the MLE consistency result by Newey and McFadden (1994, p. 2131) and show that all the assumptions of their Theorem 2.5 hold. I first need to show that the identification condition (i) of Theorem 2.5 holds, i.e. if $\beta\neq\beta_0$ then $f_X(\cdot|\beta)\neq f_X(\cdot|\beta_0)$. I prove the converse of the above implication: consider the case where $f_X(\cdot|\beta)=f_X(\cdot|\beta_0)$. This implies that these two APD densities have modes that coincide; since the APD density is unimodal we necessarily have $\theta=\theta_0$. Further, note that $f_X(\cdot|\beta)=f_X(\cdot|\beta_0)$ implies $F_X(\cdot|\beta)=F_X(\cdot|\beta_0)$ and so $\alpha=F_X(\theta|\beta)=F_X(\theta_0|\beta_0)=\alpha_0$. In order to show that $\lambda=\lambda_0$, consider the equality $\alpha^{-1}f_X(\theta|\beta)ES_X(\alpha)=\alpha_0^{-1}f_X(\theta_0|\beta)ES_X(\alpha_0)$; using the results of Proposition 1, we have that $\alpha^{-1}f_X(\theta|\beta)ES_X(\alpha)=\lambda\Gamma(2/\lambda)[\Gamma(1/\lambda)]^{-2}$. The function $\lambda\mapsto\lambda\Gamma(2/\lambda)[\Gamma(1/\lambda)]^{-2}$ is strictly monotone decreasing on $(1/2,+\infty)$, hence $\lambda\Gamma(2/\lambda)[\Gamma(1/\lambda)]^{-2}=\lambda_0\Gamma(2/\lambda_0)[\Gamma(1/\lambda_0)]^{-2}$ implies $\lambda=\lambda_0$. Finally, comparing the values that the APD densities achieve at their mode we have that $f_X(\theta|\beta)=f_X(\theta_0|\beta_0)$, together with $\theta=\theta_0,\alpha=\alpha_0$ and $\lambda=\lambda_0$, implies $\phi=\phi_0$. Hence $\beta=\beta_0$ which shows identification. The compactness condition (ii) of Theorem 2.5 is ensured by considering a compact parameter set B. The continuity condition (iii) of Theorem 2.5 is trivially verified since $\ln f_X(X_t|\beta)$ is continuous at each $\beta\in B$ with probability one (indeed, discontinuity arises only when $\theta=X_t$ which is of measure zero). Finally, the boundednes condition (iv) of Theorem 2.5 requires that $E[\sup_{\beta\in B}|\ln f_X(X_t|\beta)|]<\infty$. Recall that

DOI: 10.1002/jae

$$\ln f_X(X_t|\beta) = -\ln\phi + \ln\frac{\delta_{\alpha,\lambda}^{1/\lambda}}{\Gamma(1+1/\lambda)} - \frac{\delta_{\alpha,\lambda}}{\phi^\lambda}\cdot\left[\frac{|X_t-\theta|^\lambda}{\alpha^\lambda}\mathbb{I}(X_t\le\theta) + \frac{|X_t-\theta|^\lambda}{(1-\alpha)^\lambda}\mathbb{I}(x_t>\theta)\right],$$

Note that, for every $\beta \in B$, we have

$$\begin{aligned}
|\ln f_X(X_t|\beta)| &\le |\ln\phi| + \left|\ln\frac{\delta_{\alpha,\lambda}^{1/\lambda}}{\Gamma(1+1/\lambda)}\right| + \delta_{\alpha,\lambda}\frac{[\alpha^\lambda+(1-\alpha)^\lambda]}{\alpha^\lambda(1-\alpha)^\lambda}\left|\frac{X_t-\theta}{\phi}\right|^\lambda \\
&\le |\ln\phi| + \left|\ln\frac{\delta_{\alpha,\lambda}^{1/\lambda}}{\Gamma(1+1/\lambda)}\right| + 2\phi^{-\lambda}n_\lambda[|X_t|^\lambda + |\theta|^\lambda],
\end{aligned} \tag{17}$$

where n_λ is a positive constant such that $(a+b)^\lambda \le n_\lambda(a^\lambda + b^\lambda)$ for all $a, b > 0$. Hence $E[\sup_{\beta\in B}|\ln f_X(X_t|\beta)|] \le C_1 + C_2 \max\{1, E[|X_t|^\lambda\}$, where $C_1 = \sup_{\beta\in B}\{|\ln\phi| + \left|\ln\frac{\delta_{\alpha,\lambda}^{1/\lambda}}{\Gamma(1+1/\lambda)}\right| + 2\phi^{-\lambda}n_\lambda|\theta|^\lambda\}$, $C_2 = sup_{\beta\varepsilon B}\{2\phi^{-\lambda}n_\lambda\}$ and $\overline{\lambda} = sup_{\beta\varepsilon B}\lambda$. By compactness of B we have $C_1 < \infty, C_2 < \infty$, and $\overline{\lambda} < \infty$; it remains to be shown that $E[|X_t|^\lambda] < \infty$. From Equation (2.7) in Ayebo and Kozubowski (2003), a simple change of variable $\theta = 0$, $\sigma = \left[\frac{\alpha^{\overline{\lambda}} + (1-\alpha)^{\overline{\lambda}}}{2\alpha^{\lambda/2}(1-\alpha)^{\lambda/2}}\right]^{1/\overline{\lambda}}$ and $\kappa = \left[\frac{1-\alpha}{\alpha}\right]^{1/2}$ shows that the $\overline{\lambda}$-moment of the absolute value of a standard APD random variable $U_t = (X_t - \theta_0)/\phi_0$ with parameters (α_0, λ_0) equals $E(|U_t|^{\overline{\lambda}}) = \frac{\Gamma((1+\lambda)/\lambda_0)}{\Gamma(1/\lambda_0)}\delta_{\alpha_0,\lambda_0}^{-\overline{\lambda}/\lambda_0}[\alpha_0^{1+\overline{\lambda}} + (1-\alpha_0)^{1+\overline{\lambda}}]$. We therefore get that

$$E(|U_t|^{\overline{\lambda}}) \le \frac{\Gamma((1+\overline{\lambda})/\lambda_0)}{\Gamma(1/\lambda_0)}2\delta_{\alpha_0,\lambda_0}^{-\overline{\lambda}/\lambda_0}, \tag{18}$$

and $E(|X_t|^{\overline{\lambda}}) \le n_{\overline{\lambda}}\left[|\theta_0|^{\overline{\lambda}} + \phi_0^{\overline{\lambda}}\frac{\Gamma((1+\overline{\lambda})/\lambda_0)}{\Gamma(1/\lambda_0)}2\delta_{\alpha_0,\lambda_0}^{-\overline{\lambda}/\lambda_0}\right] < \infty$, as desired. Applying the result of Theorem 2.5 in Newey and McFadden (1994, p 2131) I thus show that $\hat{\beta}_T$ is consistent, i.e. $\hat{\beta}_T \xrightarrow{p} \beta_0$.

I now show that $\hat{\beta}_T$ is asymptotically normal with asymptotic covariance matrix $J(\beta_0)^{-1}$, where $J(\beta_0) = E[[\nabla_\beta \ln f_X(X_t|\beta_0)][\nabla_\beta \ln f_X(X_t|\beta_0)]']$. In order to do so, I use the asymptotic normality result for the MLE contained in Theorem 7.1 of Newey and McFadden (1994, p. 2185). It is important to note that the main difficulty in applying the existing asymptotic normality results lies in the fact that the objective function here is not everywhere differentiable. The first condition to be satisfied for the asymptotic normality to hold is the maximum condition (i): $\beta_0 = \arg\max_{\beta\in B} E[\ln f_X(X_t|\beta)]$. This condition is trivially satisfied by assuming that $X_1, \ldots, X_T$ are i.i.d. from the APD distribution with parameter β_0 (i.e., there is no distributional misspecification). The interior condition (ii) of Theorem 7.1 is equivalent to the assumption $\beta_0 \in \mathring{B}$ (interior of B). The twice differentiability condition (iii) also holds with the 4×4 Hessian matrix of second derivatives, $H(\beta_0) = E[\Delta_{\beta\beta} \ln f_X(X_t|\beta_0)]$, being nonsingular. I checked the nonsingularity condition by first computing analytic expressions of the elements of $\Delta_{\beta\beta} \ln f_X(x|\beta_0)$, and then numerically integrating them with respect to the four-parameter APD probability density with the true parameter β_0. My numerical computations of $H(\beta_0)$ have shown that for $\beta_0 \in B = [0.01, 0.99] \times [0.5, 5] \times [-10, 10] \times [0.1, 10]$, we have $\det(H) \neq 0$. Due to the length of analytic expressions for different elements of $H(\beta_0)$ I choose not to report them here. I now show that condition (iv) of Theorem 7.1 is satisfied, i.e., that $\sqrt{T} D_T(\beta_0) \xrightarrow{d} N(0, J(\beta_0))$, where $D_T(\beta_0)$ is a gradient of $L_T(\beta)$ at β_0, i.e., $D_T(\beta_0) = T^{-1} \sum_{t=1}^{T} \nabla_\beta \ln f_X(x_t|\beta_0)$. For that, I use a standard Lindeberg–Levy central limit Theorem (CLT) for i.i.d. sequences (see, e.g., Theorem 5.2 in White, 2001, p. 114) for which I need to show that all the elements of the asymptotic covariance matrix $J(\beta_0)$ are finite. Note that we have $|[\nabla_\beta \ln f_X(X_t|\beta_0)][\nabla_\beta \ln f_X(X_t|\beta_0)]'| = |\frac{\partial}{\partial \beta_{i_0}} \ln f_X(X_t|\beta_0)| \cdot |\frac{\partial}{\partial \beta_{j_0}} \ln f_X(X_t|\beta_0)|$, where the indices $1 \leq i_0, j_0 \leq 4$ are such that $\max_{1\leq i,j\leq 4} |\frac{\partial}{\partial \beta_i} \ln f_X(X_t|\beta_0) \cdot \frac{\partial}{\partial \beta_j} \ln f_X(X_t|\beta_0)| = |\frac{\partial}{\partial \beta_{i_0}} \ln f_X(X_t|\beta_0) \cdot \frac{\partial}{\partial \beta_{j_0}} \ln f_X(X_t|\beta_0)|$. Hence, by norm equivalence we know that there exist a positive constant c, such that $|[\nabla_\beta \ln f_X(X_t|\beta_0)][\nabla_\beta \ln f_X(X_t|\beta_0)]'| \leq c^2 \cdot |\nabla_\beta \ln f_X(X_t|\beta_0)|^2$. Then, all the elements of $J(\beta_0)$ are finite if $E\left[|\frac{\partial}{\partial \beta_i} \ln f_X(X_t|\beta_0)|^2\right] < \infty$ for $1 \leq i \leq 4$: based on the results from Lemma 5, we have $|\frac{\partial}{\partial \alpha} \ln f_X(X_t|\beta_0)|^2 \leq 4\left[\frac{\partial \delta_{\alpha_0,\lambda_0}}{\partial \alpha}\right]^2 \times \left[\frac{1}{\lambda_0^2 \delta_{\alpha_0,\lambda_0}^2} + \frac{|X_t - \theta_0|^{2\lambda_0}}{\phi_0^{2\lambda_0}} \frac{4}{\delta_{\alpha_0,\lambda_0}^2}\right] + 4\lambda_0^2 \delta_{\alpha_0,\lambda_0}^2 \frac{|X_t - \theta_0|^{2\lambda_0}}{\phi_0^{2\lambda_0}} \frac{4}{\alpha_0^2(1-\alpha_0)^2 \delta_{\alpha_0,\lambda_0}^2}$, where $\left[\frac{\partial \delta_{\alpha_0,\lambda_0}}{\partial \alpha}\right]^2 \leq \frac{16\lambda_0^2 \delta_{\alpha_0,\lambda_0}^2}{\alpha_0^2(1-\alpha_0)^2}$. Hence

$$|\frac{\partial}{\partial \alpha} \ln f_X(X_t|\beta_0)|^2 \leq \frac{64}{\alpha_0^2(1-\alpha_0)^2}[1 + 4.25\lambda_0^2 |U_t|^{2\lambda_0}] \tag{19}$$

where as previously $U_t = (X_t - \theta_0)/\phi_0$ denotes a standard APD random variable with parameters (α_0, λ_0). By using (19) together with the moment inequality in (18) we then have

$$E\left[|\frac{\partial}{\partial \alpha} \ln f_X(X_t|\beta_0)|^2\right] \leq \frac{64}{\alpha_0^2(1-\alpha_0)^2}\left[1 + 8.5\lambda_0^2 \frac{\Gamma((1+2\lambda_0)/\lambda_0)}{\delta_{\alpha_0,\lambda_0}^2 \Gamma(1/\lambda_0)}\right] < \infty \tag{20}$$

By using the same reasoning as above, we have

$$\begin{aligned} |\frac{\partial}{\partial \lambda} \ln f_X(X_t|\beta_0)|^2 \leq{}& \frac{4}{\lambda_0^4}[(\ln \delta_{\alpha_0,\lambda_0})^2 + \frac{\lambda_0^2}{\delta_{\alpha_0,\lambda_0}^2}(\frac{\partial \delta_{\alpha_0,\lambda_0}}{\partial \lambda})^2 + (\Psi(1 + 1/\lambda_0))^2] + 4[(\frac{\partial \delta_{\alpha_0,\lambda_0}}{\partial \lambda})^2 \\ &+ \delta_{\alpha_0,\lambda_0}^2 (\ln \frac{|X_t - \theta_0|}{\phi_0})^2] \frac{|X_t - \theta_0|^{2\lambda_0}}{\phi_0^{2\lambda_0}} \frac{4}{\delta_{\alpha_0,\lambda_0}^2} + \delta_{\alpha_0,\lambda_0}^2 \frac{|X_t - \theta_0|^{2\lambda_0}}{\phi_0^{2\lambda_0}} [\ln(\alpha_0(1-\alpha_0))]^2 \frac{4}{\delta_{\alpha_0,\lambda_0}^2} \end{aligned}$$

where $\left[\frac{\partial \delta_{\alpha_0,\lambda_0}}{\partial \lambda}\right]^2 = 4\delta_{\alpha_0,\lambda_0}^2 [\ln(\alpha_0(1-\alpha_0))]^2$. Hence

$$\begin{aligned} |\frac{\partial}{\partial \lambda} \ln f_X(X_t|\beta_0)|^2 \leq{}& \frac{4}{\lambda_0^4}\{[\ln \delta_{\alpha_0,\lambda_0}]^2 + 4\lambda_0^2 [\ln(\alpha_0(1-\alpha_0))]^2 + (\Psi(1 + 1/\lambda_0))^2\} \\ &+ 16\{5[\ln(\alpha_0(1-\alpha_0))]^2 + [\ln |U_t|]^2\}|U_t|^{2\lambda_0} \end{aligned}$$

Note that $[\ln|U_t|]^2 \leq 2(|U_t|^2 + 1/|U_t|^2)$, so that

$$|\frac{\partial}{\partial\lambda}\ln f_X(X_t|\beta_0)|^2 \leq \frac{4}{\lambda_0^4}\{[\ln\delta_{\alpha_0,\lambda_0}]^2 + 4(\lambda_0^2+20)[\ln(\alpha_0(1-\alpha_0))]^2 + (\Psi(1+1/\lambda_0))^2\} + 32\{|U_t|^{2(\lambda_0+1)} + |U_t|^{2(\lambda_0-1)}\}$$

Hence, by inequality (18) we have

$$E\left[|\frac{\partial}{\partial\lambda}\ln f_X(X_t|\beta_0)|^2\right] \leq \frac{4}{\lambda_0^4}\{[\ln\delta_{\alpha_0,\lambda_0}]^2 + 4(\lambda_0^2+20)[\ln(\alpha_0(1-\alpha_0))]^2 + (\Psi(1+1/\lambda_0))^2\} + \frac{64}{\Gamma(1/\lambda_0)}\left\{\frac{\Gamma((3+2\lambda_0)/\lambda_0)}{\delta_{\alpha_0,\lambda_0}^{2(\lambda_0+1)/\lambda_0}} + \frac{\Gamma((2\lambda_0-1)/\lambda_0)}{\delta_{\alpha_0,\lambda_0}^{2(\lambda_0-1)/\lambda_0}}\right\} < \infty \tag{21}$$

Similarly, we have $|\frac{\partial}{\partial\theta}\ln f_X(X_t|\beta_0)|^2 \leq 4\lambda_0^2\frac{|X_t-\theta_0|^{2(\lambda_0-1)}}{\phi_0^{2\lambda_0}} = \frac{4\lambda_0^2}{\phi_0^2}|U_t|^{2(\lambda_0-1)}$, and

$$E\left[|\frac{\partial}{\partial\theta}\ln f_X(X_t|\beta_0)|^2\right] \leq \frac{8\lambda_0^2}{\phi_0^2}\frac{\Gamma((2\lambda_0-1)/\lambda_0)}{\delta_{\alpha_0,\lambda_0}^{2(\lambda_0-1)/\lambda_0}\Gamma(1/\lambda_0)} < \infty \tag{22}$$

Note that $\lambda_0 \geq 1/2$ so $\Gamma((2\lambda_0-1)/\lambda_0) \geq 0$, which is required for the nonsingularity of $J(\beta_0)$

Finally, $|\frac{\partial}{\partial\phi}\ln f_X(X_t|\beta_0)|^2 \leq \frac{2}{\phi_0^2}\left[1 + 4\lambda_0^2\frac{|X_t-\theta_0|^{2\lambda_0}}{\phi_0^{2\lambda_0}}\right] = \frac{2}{\phi_0^2}[1 + 4\lambda_0^2|U_t|^{2\lambda_0}]$, so that by using again the result of inequality (18) we get

$$E\left[|\frac{\partial}{\partial\phi}\ln f_X(X_t|\beta_0)|^2\right] \leq \frac{2}{\phi_0^2}\left[1 + 8\lambda_0^2\frac{\Gamma((1+2\lambda_0)/\lambda_0)}{\delta_{\alpha_0,\lambda_0}^2\Gamma(1/\lambda_0)}\right] < \infty \tag{23}$$

Inequalities (20), (21), (22) and (23) imply that all the elements of $J(\beta_0)$ are finite, therefore I can use Theorem 5.2 in White (2001, p. 114) to show that condition (iv) of Theorem 7.1 is satisfied. Finally, the stochastic differentiablity condition (v) of the same theorem can be shown to hold by using the results obtained by Andrews (1994) for the special case $\lambda_0 = 1$ and extending them to any $\lambda_0 > 1/2$. I can now apply the results of Theorem 7.1 in Newey and McFadden (1994) to show that $\sqrt{T}(\hat{\beta}_T - \beta_0)\xrightarrow{d}N(0, J(\beta_0)^{-1})$, since in the maximum likelihood case $H(\beta_0) = -J(\beta_0)$. This completes the proof of the asymptotic normality result of Proposition 2.

It remains to be shown that my estimator $J_T(\hat{\beta}_T)$ of $J(\beta_0)$ is consistent:

$$J_T(\hat{\beta}_T) = T^{-1}\sum_{t=1}^{T}(\nabla_\beta \ln f_X(x_t|\hat{\beta}_T))(\nabla_\beta \ln f_X(x_t|\hat{\beta}_T))'\xrightarrow{p}J(\beta_0) \tag{24}$$

For that, I use Theorem 4.4 in Newey and McFadden (1994). First, note that each element of the gradient $\nabla_\beta \ln f_X(X_t|\beta)$ is continuous with probability one—the only point of discontinuity is at θ and the probability of the corresponding event $\{X_t = \theta\}$ equals zero. Hence, the only condition I

DOI: 10.1002/jae

need for (24) to hold is that in some neighborhood $\mathcal{N}$ of β_0 we have $E[\sup_{\beta\in\mathcal{N}}|\nabla_\beta \ln f_X(X_t|\beta)|^2] < \infty$, i.e., that $E\left[\sup_{\beta\in\mathcal{N}}|\frac{\partial}{\partial\beta_i}\ln f_X(X_t|\beta)|^2\right] < \infty$ for $1 \le i \le 4$. From (19) we have

$$
\begin{aligned}
|\frac{\partial}{\partial\alpha}\ln f_X(X_t|\beta)|^2 &\le \frac{64}{\alpha^2(1-\alpha)^2}\left[1 + 4.25\lambda^2|\frac{X_t-\theta}{\phi}|^{2\lambda}\right] \\
&\le \frac{64}{\alpha^2(1-\alpha)^2}\left\{1 + 4.25\lambda^2\frac{\phi_0^{2\lambda}}{\phi^{2\lambda}}c_\lambda\left[|U_t|^{2\lambda} + |\frac{\theta-\theta_0}{\phi_0}|^{2\lambda}\right]\right\}
\end{aligned}
$$

where $U_t = (X_t - \theta_0)/\phi_0$ is a standard APD with shape parameters (α_0, λ_0) and c_λ is a positive constant such that $(x+y)^{2\lambda} \le c_\lambda(x^{2\lambda} + y^{2\lambda})$ for all $(x, y) \in \mathbb{R}^2$. From the moment inequality in (18) we know that $E(|U_t|^\lambda) \le 2\delta_{\alpha_0,\lambda_0}^{-\lambda/\lambda_0}\Gamma((1+\lambda)/\lambda_0)/\Gamma(1/\lambda_0)$, so we have

$$
\begin{aligned}
&E\left[\sup_{\beta\in\mathcal{N}}|\frac{\partial}{\partial\alpha}\ln f_X(X_t|\beta)|^2\right] \\
&\le \sup_{\beta\in\mathcal{N}}\frac{64}{\alpha^2(1-\alpha)^2}\left\{1 + 4.25\lambda^2\frac{\phi_0^{2\lambda}}{\phi^{2\lambda}}c_\lambda\left[2\frac{\Gamma((1+2\lambda)/\lambda_0)}{\delta_{\alpha_0,\lambda_0}^{2\lambda/\lambda_0}\Gamma(1/\lambda_0)} + |\frac{\theta-\theta_0}{\phi_0}|^{2\lambda}\right]\right\} < \infty
\end{aligned}
\tag{25}
$$

Similarly, we have $|\frac{\partial}{\partial\lambda}\ln f_X(X_t|\beta)|^2 \le \frac{4}{\lambda^4}\{[\ln\delta_{\alpha,\lambda}]^2 + 4(\lambda^2+20)[\ln(\alpha(1-\alpha))]^2 + (\Psi(1+1/\lambda))^2\} + 32\left\{c_{\lambda+1}\frac{\phi_0^{2(\lambda+1)}}{\phi^{2(\lambda+1)}}\left[|U_t|^{2(\lambda+1)} + |\frac{\theta-\theta_0}{\phi_0}|^{2(\lambda+1)}\right] + c_{\lambda-1}\frac{\phi_0^{2(\lambda-1)}}{\phi^{2(\lambda-1)}} \times \left[|U_t|^{2(\lambda-1)} + |\frac{\theta-\theta_0}{\phi_0}|^{2(\lambda-1)}\right]\right\}$. Hence, by inequality (18) we have

$$
\begin{aligned}
E[\sup_{\beta\in\mathcal{N}}|\frac{\partial}{\partial\lambda}\ln f_X(X_t|\beta)|^2] \le{}& \frac{4}{\lambda^4}\{[\ln\delta_{\alpha,\lambda}]^2 + 4(\lambda^2+20)[\ln(\alpha(1-\alpha))]^2 + (\Psi(1+1/\lambda))^2\} \\
&+ \frac{32}{\Gamma(1/\lambda_0)}\{c_{\lambda+1}\frac{\phi_0^{2(\lambda+1)}}{\phi^{2(\lambda+1)}}[2\frac{\Gamma((3+2\lambda)/\lambda_0)}{\delta_{\alpha_0,\lambda_0}^{2(\lambda+1)/\lambda_0}} + |\frac{\theta-\theta_0}{\phi_0}|^{2(\lambda+1)}] \\
&+ c_{\lambda-1}\frac{\phi_0^{2(\lambda-1)}}{\phi^{2(\lambda-1)}}[2\frac{\Gamma((2\lambda-1)/\lambda_0)}{\delta_{\alpha_0,\lambda_0}^{2(\lambda-1)/\lambda_0}} + |\frac{\theta-\theta_0}{\phi_0}|^{2(\lambda-1)}]\} < \infty
\end{aligned}
\tag{26}
$$

Similarly, we have $|\frac{\partial}{\partial\theta}\ln f_X(X_t|\beta)|^2 \le \frac{4\lambda^2}{\phi^2}c_{\lambda-1}\frac{\phi_0^{2(\lambda-1)}}{\phi^{2(\lambda-1)}}\left[|U_t|^{2(\lambda-1)} + |\frac{\theta-\theta_0}{\phi_0}|^{2(\lambda-1)}\right]$, so

$$
E\left[\sup_{\beta\in\mathcal{N}}|\frac{\partial}{\partial\theta}\ln f_X(X_t|\beta)|^2\right] \le \frac{4\lambda^2}{\phi^2}c_{\lambda-1}\frac{\phi_0^{2(\lambda-1)}}{\phi^{2(\lambda-1)}}\left[2\frac{\Gamma((2\lambda-1)/\lambda_0)}{\delta_{\alpha_0,\lambda_0}^{2(\lambda-1)/\lambda_0}\Gamma(1/\lambda_0)} + |\frac{\theta-\theta_0}{\phi_0}|^{2(\lambda-1)}\right] < \infty \tag{27}
$$

Finally, we have $|\frac{\partial}{\partial\phi}\ln f_X(X_t|\beta)|^2 \leq \frac{2}{\phi^2}\left\{1+4\lambda^2\frac{\phi_0^{2\lambda}}{\phi^{2\lambda}}c_\lambda\left[|U_t|^{2\lambda}+|\frac{\theta-\theta_0}{\phi_0}|^{2\lambda}\right]\right\}$, so that by using again the result of inequality (18) we get

$$E\left[\sup_{\beta\in\mathcal{N}}|\frac{\partial}{\partial\phi}\ln f_X(X_t|\beta)|^2\right] \leq \frac{2}{\phi^2}\left\{1+4\lambda^2\frac{\phi_0^{2\lambda}}{\phi^{2\lambda}}c_\lambda\left[2\frac{\Gamma((1+2\lambda)/\lambda_0)}{\delta^2_{\alpha_0,\lambda_0}\Gamma(1/\lambda_0)}+|\frac{\theta-\theta_0}{\phi_0}|^{2\lambda}\right]\right\} < \infty \tag{28}$$

Inequalities (25), (26), (27) and (28) imply that $E[\sup_{\beta\in\mathcal{N}}|\nabla_\beta \ln f_X(X_t|\beta)|^2] < \infty$. Hence, I can apply Theorem 4.4 in Newey and McFadden (1994) to show that $J_T(\hat{\beta}_T)\xrightarrow{p}J(\beta_0)$. Given the nonsingularity of $J(\beta_0)$ and the continuity of the inverse function away from zero, it follows that $J_T(\hat{\beta}_T)^{-1}$ is consistent for $J(\beta_0)^{-1}$, which completes the proof of Proposition 2.

ACKNOWLEDGMENTS

I would like to thank the Editor, Tim Bollerslev, and three anonymous Referees for useful comments, which led to a considerably improved version of the paper. Thanks to Yong Bao, Christian Bontemps, Jean-Marie Dufour, Nour Meddahi, Denis Pelletier, George Tauchen and all the participants at the Montreal (2004) CIRANO/CIREQ conference in Financial Econometrics for their comments and suggestions. All computations and simulations are performed in Matlab and the corresponding programs are available upon request from the author.

REFERENCES

Aas K, Haff IH. 2006. The generalized hyperbolic skew Student's *t*-distribution. *Journal of Financial Econometrics* **4**: 275–309.

Andrews OWK. 1994. Empirical process ethods in econometrics. In *Handbook of Econometrics*. Vol. 4, Engle RF, McFadden DL (eds). Elsevier Science: Amsterdam; 2248–2294.

Artzner P, Delbaen F, Eber JM, Heath D. 1999. Coherent measures of risk. *Mathematical Finance* **9**: 203–228.

Ayebo A, Kozubowski TJ. 2003. An asymmetric generalization of Gaussian and Laplace laws. *Journal of Probability and Statistical Science* **1**: 187–210.

Bain LJ, Engelhardt M. 1973. Interval estimation for the two-parameter double exponential distribution. *Technometrics* **15**: 875–887.

Balakrishnan N, Basu AP. 1995. *The Exponential Distribution: Theory, Methods and Applications*. Gordon & Breach: New York.

Balakrishnan N, Chandramouleeswaran MP, Ambagaspitiya RS. 1996. BLUE's of location and scale parameters of Laplace distribution based on Type-II censored samples and associated inference. *Microelectronic Reliability* **36**: 371–374.

Bassett G, Koenker R, Kordas G. 2004. Pessimistic portfolio allocation and Choquet expected utility. *Journal of Financial Econometrics* **2**: 477–492.

Bawa VS. 1978. Safety first, stochastic dominance and optimal portfolio choice. *Journal of Financial and Quantitative Analysis* **13**: 255–271.

Bawa VS, Lindenberg EB. 1977. Capital market equilibrium in a mean-lower partial moment framework. *Journal of Financial Economics* **5**: 189–200.

Birnbaum A, Mike V. 1970. Asymptotically robust estimators of location. *Journal of the American Statistical Association* **65**: 1265–1282.

Bollerslev T. 1986. Generalized autoregressive conditional heteroskedasticity. *Journal of Econometrics* **31**(3): 307–328.

Bollerslev T, Wooldridge JM. 1992. Quasi-maximum likelihood estimation and inference in dynamic models with time-varying covariances. *Econometric Reviews* **11**: 143–179.
Buchinsky M. 1995. Estimating the asymptotic covariance matrix for quantile regression models: a Monte Carlo study. *Journal of Econometrics* **68**: 303–338.
Chen XI. 2005. Nonparametric estimation of expected shortfall. Mimeo, Iowa State University.
Engle RF. 1982. Autoregressive conditional heteroskedasticity with estimates of the variance of United Kingdom inflation. *Econometrica* **50**: 987–1007.
Feldstein MS. 1969. Mean-variance analysis in the theory of liquidity preference and portfolio selection. *Review of Economic Studies* **36**(1): 5–12.
Fermanian JD, Scaillet O. 2005. Sensitivity analysis of VaR and expected shortfall for portfolios under netting agreements. *Journal of Banking and Finance* **29**: 927–958.
Fernandez C, Osiewalski J, Steel MFJ. 1995. Modeling and inference with v-spherical distributions. *Journal of the American Statistical Association* **90**: 1331–1340.
Fernandez C, Steel MJF. 1998. On Bayesian modeling of fat tails and skewness. *Journal of the American Statistical Association* **93**: 359–371.
Fitzenberger B. 1997. The moving blocks bootstrap and robust inference for linear least squares and quantile regressions. *Journal of Econometrics* **82**: 235–287.
Follmer H, Schied A. 2002. Convex measures of risk and trading constraints. *Finance and Stochastics* **6**(4): 429–447.
Giot P, Laurent S. 2004. Modelling daily value-at-risk using realized volatility and ARCH type models. *Journal of Empirical Finance* **11**: 379–398.
Govindarajulu Z. 1966. Best linear estimates under symmetric censoring of the parameters of double exponential distribution. *Journal of the American Statistical Association* **61**: 248–258.
Hanoch G, Levy H. 1969. The efficiency analysis of choices involving risk. *Review of Economic Studies* **36**: 335–346.
Hansen BE. 1994. Autoregressive conditional density estimation. *International Economic Review* **35**: 705–730.
Jakuszenkow H. 1979. Estimation of the variance in the generalized Laplace distribution with quadratic loss function. *Demonstratio Mathematica* **3**: 581–591.
Johnson ME. 1979. Computer generation of the exponential power distributions. *Journal of Statistical Computation and Simulation* **9**: 239–240.
Johnson NL, Kotz S, Balakrishnan N. 1994. *Continuous Univariate distributions* (2nd edn). Wiley: New York.
Koenker R, Bassett G. 1978. Regression quantiles. *Econometrica* **46**: 33–50.
Komunjer I. 2005. Quasi-maximum likelihood estimation for conditional quantiles. *Journal of Econometrics* **128**: 137–164.
Kotz S, Kozubowski TJ, Podgorski K. 2001. *The Laplace Distribution and Generalizations: A Revisit with Applications to Communications, Economics, Engineering, and Finance*. Birkhauser: Boston, MA.
Kuester K, Mittnik S, Paolella MS. 2006. Value-at-risk prediction: a comparison of alternative strategies. *Journal of Financial Econometrics* **4**: 53–89.
Kupiec P. 1995. Techniques for verifying the accuracy of risk measurement models. *Journal of Derivatives* **3**: 73–84.
Lingappaiah GS. 1988. On two-piece double exponential distribution. *Journal of the Korean Statistical Society* **17**: 46–55.
Lopez J. 1997. Regulatory evaluation of value-at-risk models. *Federal Reserve Bank of New York Research Paper* 9710.
Markowitz HM. 1952. Portfolio selection. *Journal of Finance* **6**: 77–91.
McDonald JB. 1991. Parametric models for partially adaptive estimation with skewed and leptokurtic residuals. *Economics Letters* **37**: 273–278.
McDonald JB. 1997. Probability distributions for financial models. In *Handbook of Statistics*, Vol. 14, Maddala GS, Rao CR (eds). Elsevier Science: Amsterdam; 427–461.
McDonald JB, Newey WK. 1988. Partially adaptive estimation of regression models via the generalized t distribution. *Econometric Theory* **4**: 428–457.
McNeil AJ, Frey R. 2000. Estimation of tail-related risk measures for heteroskedastic time series: an extreme value approach. *Journal of Empirical Finance* **7**: 271–300.

Nelson DB. 1991. Conditional heteroskedasticity in asset returns: a new approach. *Econometrica* **59**: 347–370.

Newey WK, McFadden DL. 1994. Large sample estimation and hypothesis testing. In *Handbook of Econometrics*, Vol. 4, Engle RF, McFadden DL (eds). Elsevier Science: Amsterdam; 2113–2247.

Paolella MS. 2006. *Intermediate Probability: A Computational Approach*. Wiley: Chichester.

Patton AJ. 2004. On the out-of-sample importance of skewness and asymmetric dependence for asset allocation. *Journal of Financial Econometrics* **2**: 130–168.

Scaillet O. 2004. Nonparametric estimation and sensitivity analysis of expected shortfall. *Mathematical Finance* **14**: 115–129.

Scaillet O. 2005. Nonparametric Estimation of Conditional Expected Shortfall. *Revue Assurances et Gestion des Risques/Insurance and Risk Management Journal* **74**: 639–660.

Straumann D. 2005. *Estimation in Conditionally Heteroskedastic Time Series Models*. Springer: Berlin.

White H. 2001. *Asymptotic Theory for Econometricians*. Academic Press: San Diego, CA.

DOI: 10.1002/jae