

# Asymmetric Power Distribution: Theory and Applications to Risk Measurement

Ivana Komunjer<sup>\*,†</sup>

Caltech

## Abstract

Theoretical literature in finance has shown that quantifying the risk of financial time series amounts to measuring their expected shortfall, also known as tail Value at Risk. Unfortunately, little empirical work has been devoted to the problem of modeling and inference of such risk measures and, in particular, to their estimation. In this paper, we construct a parametric estimator for the expected shortfall based on a new family of densities, which we call the Asymmetric Power Distribution (APD).

The APD family extends the Generalized Power Distribution to cases where the data exhibits asymmetry. We provide a detailed description of the properties of an APD random variable, such as its quantiles, moments and moment related parameters. Moreover, we discuss the problem of simulation of such random variables and provide maximum likelihood estimates of the APD density parameters. The study of asymptotic properties of the latter falls outside the standard framework due to the non-differentiability of the APD log-likelihood.

---

<sup>\*</sup> *Address:* Division of the Humanities and Social Sciences, Caltech MC 228-77, 1200 E California Boulevard, Pasadena, CA 91125 (*komunjer@hss.caltech.edu*)

<sup>†</sup> All computations were performed using MATLABv6.5 software and the corresponding algorithms are available from the author upon request.

An empirical application to six 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. Under a more general assumption that the return innovations are APD, we are able to compute expected shortfalls and corresponding confidence intervals and thus compare the riskiness of the series examined.

**Keywords:** Asymmetric Power Distribution, expected shortfall, risk.

## 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. It is therefore not surprising that a large literature has been devoted to studying the properties of risk measures such as volatility, Value at Risk (VaR) or expected shortfall (also known as tail VaR). An important number of theoretical results is today available on these three measures of risk (see, e.g. Markowitz 1952, Bawa 1978, Hanoch and Levy 1969, Artzner, Delbaen, Eber and Heath 1999, Follmer and Schied 2003). Unfortunately, empirical literature has primarily focused on the first two leaving the problem of estimation and inference of expected shortfall mainly unanswered. In this paper, we propose a parametric estimator for the expected shortfall based on a new family of densities, which we call the Asymmetric Power Distribution (APD). The main feature of the APD family is that it combines the flexible tail decay property with the asymmetry, which makes it particularly suited for modeling the behavior of financial returns.

Starting with the work of Markowitz (1952), virtually all risk-return models over the past 40 years, have used volatility as the main measure of risk. The shortcomings of volatility as a risk measure for the purposes of portfolio optimization are however well established in the literature. Early on, Hanoch and Levy (1969), e.g., have shown that the volatility was a good risk measure only if the returns were elliptically distributed. This serious limitation of volatility

has further prompted researchers to define more generally valid risk measures, as witnessed, e.g., by the work of Bawa (1978). More recently, Artzner, Delbaen, Eber and Heath (1999) and Follmer and Schied (2003), have raised the question if VaR qualifies as such a measure. Their common finding is that VaR - defined as a prespecified quantile of the distribution of portfolio returns - is not an adequate measure of risk. Unlike the VaR, the expected shortfall (or tail VaR) - defined as expected portfolio tail return - has been shown to have all necessary characteristics of a good risk measure.<sup>1</sup> These recent findings agree with the results on Lower Partial Moments (LPMs), derived by the early literature.<sup>2</sup> For example, Bawa (1978) has shown that portfolios with low expected shortfall second order stochastically dominate those with high expected shortfall.

Despite a theoretical appeal of the expected shortfall for the purposes of risk measurement, relatively little work has been done on its estimation. We try to fill this literature gap by proposing a parametric estimator for the expected shortfall of a given variable of interest. An alternative approach to estimation would be to estimate the latter non-parametrically. Such an approach would, however, involve non-parametric density estimation, and would as such inherit all of its difficulties, which is why we choose not to pursue this line of research here. A yet different approach would consist in estimating the expected shortfall semi-parametrically, by using an M-estimator for example. The related literature, however, proposes no estimators for LPMs in general and for the expected shortfall in particular, despite its success with estimating quantiles. Left are therefore fully parametric methods, on which we focus in this paper.

The main drawback of any parametric estimator for the expected shortfall is that it imposes strong constraints on the shape of the density of interest, and in particular on its tails. For example, a double-exponential (Laplace) assumption imposes exponential decay of the density (see, e.g., Govindarajulu 1966, Birnbaum and Mike 1970, Bain and Engelhardt 1973, Sheynin

---

<sup>1</sup>Requirements for coherence or convexity of a given risk measure can be found in Artzner *et al.* (1999) and Follmer and Schied (2003).

<sup>2</sup>Given  $n \in \mathbb{N}$ , the LPM of order  $n$  of a real random variable  $X$ , evaluated at some  $x \in \mathbb{R}$ , is defined as  $LPM_n(X, x) \equiv E[(x - X)^n | X \leq x]$ .

1977, Jakuszenkow 1979, Lingappaiah 1988, Johnson, Kotz and Balakrishnan 1994 p.193, Balakrishnan and Basu 1995, Balakrishnan, Chandramouleeswaran and Ambagaspitiya 1996, Kotz, Kozubowski and Podgorski 2001), while a Gaussian implies exponential square decay. An example of density which allows for a flexible decay parameter has been known in the literature as the Generalized Power Distribution (GPD).<sup>3</sup> In finance, the GPD family has already been used by Nelson (1991), e.g., in the context of exponential ARCH volatility modeling. Despite its flexibility in modeling the tail behavior, the GPD 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.

In this paper, we propose a new family of distributions, which we call the Asymmetric Power Distribution (APD), and which contains the GPD as a special symmetric case. In what follows, we show that members of the APD family are able to generate a wide variety of values for the first four moments of the variable of interest, and are hence suited for modelling financial return series. More importantly, we derive analytic expressions for the expected shortfall of an APD random variable and construct the corresponding maximum likelihood estimator.

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. Section 3 derives different moments and moment related parameters, such as the expected shortfall. Section 4 describes how to simulate an APD random variable and how to recover its true parameters by using a maximum likelihood approach. Finally, Section 5 gives an empirical application to several daily financial return series and concludes the paper. Technical details are provided in the appendices at the end of the paper.

---

<sup>3</sup>Also known as the Exponential Power Distribution (EPD) or the Generalized Error Distribution (GED).

## 2 Definition and Basic Properties

The new family of distributions proposed in this paper combines the flexible tail decay property of the GPD family, measured by a parameter denoted  $\lambda$ , with the asymmetry, quantified by a parameter  $\alpha$ ,  $0 < \alpha < 1$ . Hence, it can be viewed as a generalization of the GPD family, which corresponds to the special case  $\alpha = 1/2$ , to any degree of asymmetry. We 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  $f : \mathbb{R} \rightarrow \mathbb{R}_+^*$ ,  $u \mapsto f(u)$  such that

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

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

We can easily verify that  $\forall u \in \mathbb{R}$ ,  $f(u) \geq 0$  and  $\int_{\mathbb{R}} f(u) du = 1$ , which ensure that  $f$  is a probability density.<sup>4</sup> The function  $f$  is moreover continuously differentiable on  $\mathbb{R}^*$ . The parameter  $\lambda$  controls the tail decay whereas  $\alpha$  measures the degree of asymmetry.

When  $\alpha$  equals one half, the APD pdf defined in equation (1) is symmetric around zero. In this important special case  $f$  reduces to the standard GPD density (see, e.g., Johnson, Kotz and Balakrishnan 1994 p.194-195, Kotz, Kozubowski and Podgorski 2001 p.219). The GPD family, indexed by a single parameter  $\lambda$ , 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  $\lambda$  decreases. Special cases of the GPD include: uniform ( $\lambda = \infty$ ), Gaussian ( $\lambda = 2$ ) and Laplace ( $\lambda = 1$ ) distributions.

---

<sup>4</sup>Note that for any  $\alpha$  and  $\lambda$  such that  $0 < \alpha < 1$  and  $\lambda > 0$ , we have  $0 < 2\alpha^\lambda(1-\alpha)^\lambda \leq \alpha^{2\lambda} + (1-\alpha)^{2\lambda} < \alpha^\lambda + (1-\alpha)^\lambda$ , so that  $0 < \delta_{\alpha,\lambda} < 1$ .

When  $\alpha$  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 (also known as two-piece double exponential distribution), obtained when  $\lambda = 1$  (see, e.g., Johnson, Kotz and Balakrishnan 1994 p.193), and the two-piece normal distribution, obtained when  $\lambda = 2$  (see, e.g., Johnson, Kotz and Balakrishnan 1994, vol 1 p.173 and vol 2 p.190). 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  $\lambda$ .

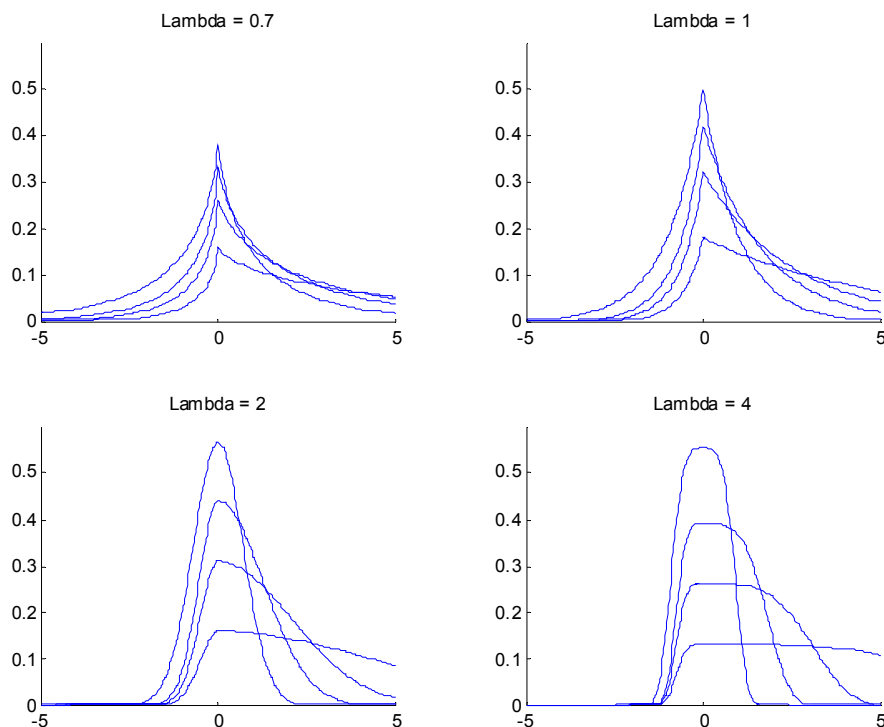


Figure 1: APD density:  $X = u$  and  $Y = f(u)$  for  $\alpha = 0.1, 0.2, 0.3, 0.5$  and  $\lambda = 0.7, 1, 2, 4$

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

as

$$X \equiv \theta + \phi U, \quad (2)$$

with a location parameter  $\theta$ ,  $\theta \in \mathbb{R}$ , and a positive scale  $\phi$ ,  $\phi > 0$ . It is straightforward to show that  $X$  has density  $f_X$ ,

$$f_X(x) = \phi^{-1} f(\phi^{-1}[x - \theta]), \text{ for any } x \in \mathbb{R}, \quad (3)$$

where  $f$  is as defined in equation (1). Note that when  $\theta = 0$  and  $\phi = 1$ ,  $X$  reduces to a standard APD random variable  $U$  in Definition 1, and we have  $f_X(x) = f(x)$ . In the most general case, the APD density  $f_X$  depends on the four parameters  $\alpha$ ,  $\lambda$ ,  $\theta$  and  $\phi$ , with  $0 < \alpha < 1$ ,  $\lambda > 0$ ,  $\theta \in \mathbb{R}$  and  $\phi > 0$ .

We now derive expressions for the cumulative distribution function (cdf)  $F$  of a standard APD random variable  $U$  and for its quantile function  $F^{-1}$ . The expressions for the cdf  $F_X$  of  $X$  and its inverse  $F_X^{-1}$  are then easily obtained from

$$F_X(x) = F(\phi^{-1}[x - \theta]), \text{ for any } x \in \mathbb{R}, \quad (4)$$

and

$$F_X^{-1}(v) = \theta + \phi^{-1} F^{-1}(v), \text{ for any } v \in (0, 1). \quad (5)$$

In the following lemma we derive  $F$ .

**Lemma 2 (APD cdf)** *For given values of  $\alpha$  and  $\lambda$ ,  $0 < \alpha < 1$  and  $\lambda > 0$ , let  $U$  be a standard APD random variable with pdf  $f$  as defined in equation (1). For any  $u \in \mathbb{R}$ , the cumulative distribution function  $F$  of  $U$  then equals*

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

where  $\delta_{\alpha, \lambda}$  is as in Definition 1 and  $I(x, \gamma)$  is Pearson's (1922) incomplete gamma function,  $I(x, \gamma) \equiv [\Gamma(\gamma)]^{-1} \int_0^{x\sqrt{\gamma}} t^{\gamma-1} \exp(-t) dt$ .

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, Kotz and Balakrishnan 1994 p.193).

We now derive expressions for different quantiles  $q_v$  of  $U$ , where  $q_v \equiv F^{-1}(v)$  and  $v$  corresponds to some probability level,  $v \in (0, 1)$ . In the following lemma we derive the quantile function  $F^{-1}$  and give a probabilistic interpretation for the parameter  $\alpha$ .

**Lemma 3 (APD quantiles)** *For any probability  $v$ ,  $v \in (0, 1)$ , the  $v$ -quantile of the standard APD random variable  $U$ ,  $F^{-1}(v)$ , equals*

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

where  $\delta_{\alpha,\lambda}$  is as in Definition 1 and  $I^{-1}(y, \gamma)$  is the inverse of the Pearson's (1922) incomplete gamma function, i.e.  $x = I^{-1}(y, \gamma)$  is equivalent to  $y = I(x, \gamma)$ . In particular,  $F^{-1}(\alpha) = 0$ .

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  $v > \alpha$ .

It is important to note that, for any  $\alpha$  and  $\lambda$  fixed,  $0 < \alpha < 1$  and  $\lambda > 0$ , the  $\alpha$ -quantile of a standard APD random variable equals zero. Therefore, the parameter  $\alpha$  has a probabilistic interpretation: it corresponds to the probability  $v$  such that the  $v$ -quantile of  $U$  equals zero, i.e.  $\alpha = F(0)$ . Using the equality (5), this property can be generalized to any APD random variable  $X$  in (2), yielding  $\alpha = F_X(\theta)$ . In other words, the probability  $\alpha$  is such that the location  $\theta$  of the APD density  $f_X$  corresponds exactly to the  $\alpha$ -quantile of  $X$ . For example, in the symmetric case where  $\theta$  is the median of  $X$ , the probability  $\alpha$  equals one half.



### 3 Moments and Related Parameters

We now turn to the study of different moments and related parameters of an APD random variable. The following lemma derives non-centered moments of the standard APD random variable  $U$ . In principle, those can be used in order to fit the parameters of the APD density by a Generalized Method of Moments (GMM), however, we choose not to follow this approach here.

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

$$E(U^r) = \frac{\Gamma((1+r)/\lambda)}{\Gamma(1/\lambda)} \frac{(1-\alpha)^{1+r} + (-1)^r \alpha^{1+r}}{\delta_{\alpha,\lambda}^{r/\lambda}}. \quad (6)$$

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

$$E(U) = \frac{\Gamma(2/\lambda)}{\Gamma(1/\lambda)} [1 - 2\alpha] \delta_{\alpha,\lambda}^{-1/\lambda}, \quad (7)$$

$$\text{var}(U) = \frac{\Gamma(3/\lambda)\Gamma(1/\lambda)[1 - 3\alpha + 3\alpha^2] - \Gamma(2/\lambda)^2[1 - 2\alpha]^2}{[\Gamma(1/\lambda)]^2} \delta_{\alpha,\lambda}^{-2/\lambda}. \quad (8)$$

When  $\alpha = 1/2$  and  $\lambda > 0$ , the random variable  $U$  has the GPD density and results on its moments are already established in the literature (see, e.g., Johnson, Kotz and Balakrishnan 1994 p.194-195, Kotz, Kozubowski and Podgorski 2001 p.219). In the asymmetric Laplace case, obtained when  $0 < \alpha < 1$  and  $\lambda = 1$ , the third and fourth centered moments of  $U$  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 unsuitable 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  $\text{ku}(U) = 6$ . Table 1 summarizes all the moment results.

[Table 1]

In Figure 2 we plot the first four moments of a standard APD random variable  $U$ . Expressions for different centered moments of  $X$  follow directly from equation (6): for example, we have  $E(X) = \theta + \phi E(U)$ ,  $\text{var}(X) = \phi^2 \text{var}(U)$ ,  $\text{sk}(X) = \text{sk}(U)$  and  $\text{ku}(X) = \text{ku}(U)$ .

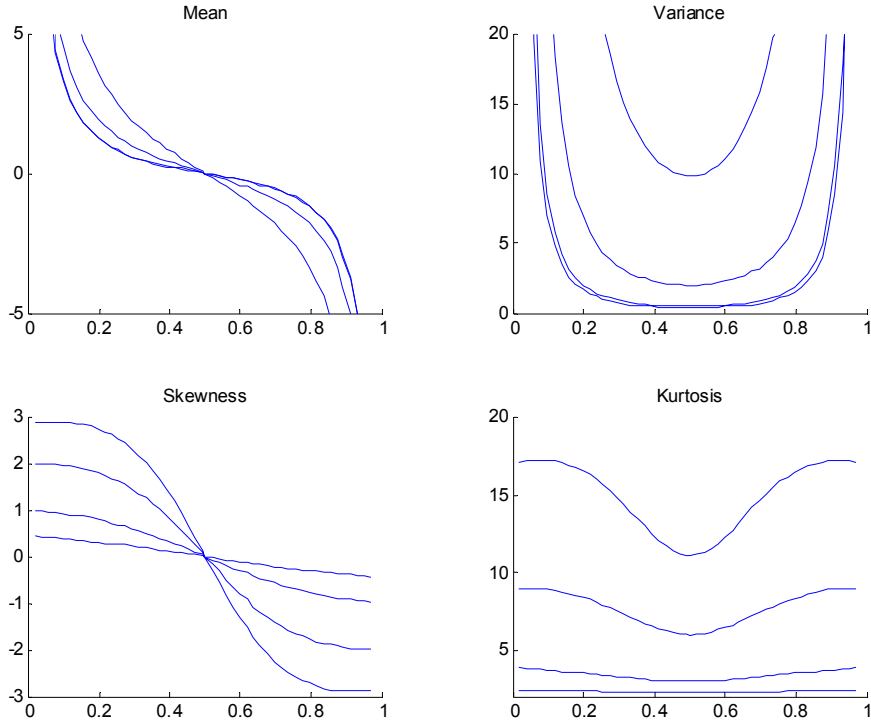


Figure 2: Moments of a standard APD random variable:  $X = \alpha$ ,  $0 < \alpha < 1$ , and  $Y = E(U)$ ,  $\text{var}(U)$ ,  $\text{sk}(U)$  and  $\text{ku}(U)$ , for  $\lambda = .7, 1, 2, 4$ .

We now derive an important moment related parameter of the standard APD random variable  $U$ : its Lower Partial Moment (LPM) of order  $n$ ,  $n \in \mathbb{N}$ . The  $n$ th order LPM of  $U$  computed at a given value  $q$ ,  $LPM_n(U, q)$ , is defined as

$$LPM_n(U, q) \equiv E[(q - U)^n | U \leq q], \quad (9)$$

for  $q \in \mathbb{R}$ . An important special case of the above quantity is the  $\bar{\alpha}$ -expected shortfall of  $U$ ,  $LPM_1(U, F^{-1}(\bar{\alpha}))$ , obtained when  $n = 1$  and when the value  $q$  corresponds to the  $\bar{\alpha}$ -quantile of  $U$ ,  $0 < \bar{\alpha} < 1$ . For example, when  $\bar{\alpha} = \alpha$ , the  $\alpha$ -expected shortfall,  $E(-U | U \leq 0)$ , corresponds to the lower-tail mean value of  $U$  evaluated at zero. In other words, it is the expected value

of  $U$  conditional on the fact that  $U$  is lower than its  $\alpha$ -quantile, i.e. zero. Given that we are particularly interested in measuring the left-tail of  $U$ , we derive analytic expressions for  $\bar{\alpha}$ -expected shortfalls with probabilities  $\bar{\alpha}$  smaller than  $\alpha$ .

**Lemma 5 (LPM)** *For any probability  $\bar{\alpha}$  smaller than  $\alpha$ ,  $0 < \bar{\alpha} \leq \alpha$ , the  $\bar{\alpha}$ -expected shortfall of the standard APD random variable  $U$ , equals*

$$E(\bar{q} - U|U \leq \bar{q}) = \frac{\alpha}{\bar{\alpha}} \frac{\alpha}{\delta_{\alpha,\lambda}^{1/\lambda}} \frac{\Gamma(2/\lambda)}{\Gamma(1/\lambda)} \left[ 1 - I\left(\frac{I^{-1}(1 - \frac{\bar{\alpha}}{\alpha}, 1/\lambda)}{\sqrt{2}}, 2/\lambda\right) \right] + \bar{q}, \quad (10)$$

where  $\bar{q}$  corresponds to the  $\bar{\alpha}$ -quantile of  $U$ , i.e.  $\bar{q} = -\left[\frac{\alpha^\lambda}{\delta_{\alpha,\lambda}\sqrt{\lambda}}\right]^{1/\lambda} [I^{-1}(1 - \frac{\bar{\alpha}}{\alpha}, 1/\lambda)]^{1/\lambda}$ ,  $I(x, \gamma)$  is the Pearson's (1922) incomplete gamma function,  $I^{-1}(y, \gamma)$  its inverse and the constant  $\delta_{\alpha,\lambda}$  is as in Definition 1. In particular, the  $\alpha$ -expected shortfall of  $U$ ,  $E(-U|U \leq 0)$ , is given by

$$E(-U|U \leq 0) = \frac{\Gamma(2/\lambda)}{\Gamma(1/\lambda)} \frac{\alpha}{\delta_{\alpha,\lambda}^{1/\lambda}}. \quad (11)$$

Taking into account the location-scale property of the pdf  $f_X$ , the above results are easily transposable to any APD random variable  $X$  in (2). For example, the  $\alpha$ -expected shortfall of  $X$  is given by  $E(\theta - X|X \leq \theta) = [\Gamma(2/\lambda)/\Gamma(1/\lambda)]\phi\alpha\delta_{\alpha,\lambda}^{-1/\lambda}$ . In the asymmetric Laplace case ( $\lambda = 1$ ), the equation (11) becomes  $E(-U|U \leq 0) = 1/[2(1 - \alpha)]$ . Similarly, we have  $E(\theta - X|X \leq \theta) = \phi/[2(1 - \alpha)]$ .

Finally, we use the moment results in equations (7) and (8) to compute a density  $f_Z$  of a standardized APD random variable  $Z$ .

**Lemma 6 (Standardized  $\alpha$ -Quantile density)** *Let  $Z$  be an APD random variable which is standardized, i.e.  $E[Z] = 0$  and  $\text{var}(Z) = 1$ . Then the density of  $Z$ , denoted  $f_Z$ , is given by*

$$f_Z(z) = \begin{cases} \frac{\lambda}{\mu} \frac{\Gamma(2/\lambda)}{\Gamma(1/\lambda)^2} \exp\left[-\left(\frac{\Gamma(2/\lambda)}{\alpha\Gamma(1/\lambda)}\right)^\lambda \left|\frac{z}{\mu} + 1 - 2\alpha\right|^\lambda\right], & \text{if } z \leq -(1 - 2\alpha)\mu, \\ \frac{\lambda}{\mu} \frac{\Gamma(2/\lambda)}{\Gamma(1/\lambda)^2} \exp\left[-\left(\frac{\Gamma(2/\lambda)}{(1 - \alpha)\Gamma(1/\lambda)}\right)^\lambda \left|\frac{z}{\mu} + 1 - 2\alpha\right|^\lambda\right], & \text{if } z > -(1 - 2\alpha)\mu, \end{cases} \quad (12)$$

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

The above quantity  $-(1-2\alpha)\mu$  corresponds to the  $\alpha$ -quantile of the random variable  $Z$ , i.e.  $F_Z^{-1}(-(1-2\alpha)\mu) = \alpha$ , where  $F_Z$  is the cdf of  $Z$ . Note that in the special case  $\alpha = 1/2$ , the density  $f_Z$  reduces to the standardized GPD density,  $f_Z(z) = [\lambda/(\Lambda\Gamma(1/\lambda)2^{1+1/\lambda})] \exp[-|z/\Lambda|^\lambda/2]$ , for  $z \in \mathbb{R}$ , where  $\Lambda = [2^{-2/\lambda}\Gamma(1/\lambda)/\Gamma(3/\lambda)]^{1/2}$  (see, e.g., Nelson, 1991).<sup>5</sup>

As previously, we derive the  $\alpha$ -expected shortfall of the standardized APD random variable  $Z$ ,

$$E[(2\alpha - 1)\mu - Z | Z \leq (2\alpha - 1)\mu] = \alpha\mu, \quad (13)$$

with  $\mu$  as in Lemma 6.<sup>6</sup>

In particular, provided consistent estimates of  $\alpha$  and  $\lambda$ , we can use the equality (13) in order to estimate the  $\alpha$ -expected shortfall of  $Z$ . Before proceeding with the estimation of the latter, we need a consistent estimator for the parameter  $(\alpha, \lambda)$  of an APD distribution. In the next section, we propose one such estimator and derive its asymptotic distribution.

## 4 Simulation and Maximum Likelihood Estimation

In this section we 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 cases of Monte Carlo studies, for example.

Random variates from the APD family can be obtained by direct transformation of Gamma variates. For given values of  $\alpha$  and  $\lambda$ ,  $0 < \alpha < 1$  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  $V \equiv (W/\delta_{\alpha,\lambda})^{1/\lambda}$ ; (3) generate a random sign variable  $S$  equal to  $+1$  with probability  $(1-\alpha)$  and to  $-1$  with probability  $\alpha$ ; finally (4) let  $U \equiv -\alpha W \cdot 1(S \leq 0) + (1-\alpha)W \cdot 1(S > 0)$ .

<sup>5</sup>When  $\alpha = 1/2$  we have  $\mu = \Gamma(2/\lambda)/[4\Gamma(3/\lambda)\Gamma(1/\lambda)]^{1/2}$  and  $\Lambda = \mu^{2^{1+1/\lambda}}\Gamma(2/\lambda)/\Gamma(1/\lambda)$ .

<sup>6</sup>Note that we can write  $Z = \theta_Z + \phi_Z U$ , where the location of  $Z$  equals  $\theta_Z = -(1-2\alpha)\mu$  and its scale is  $\phi_Z = \Gamma(1/\lambda)[\Gamma(2/\lambda)]^{-1}\delta_{\alpha,\lambda}^{1/\lambda}\mu$ , with  $\mu$  as in Lemma 6.

It is straightforward to show that such random variable  $U$  has density  $f$  as defined in equation (1) and is hence standard APD distributed. Note that this method requires a Gamma random number generator that accepts values of the shape parameter greater than zero.

Alternatively, having determined the expressions of the standard APD cdf  $F$  and its inverse  $F^{-1}$  in Lemmas 2 and 3 respectively, standard APD random variates can be generated by using an inversion method. The inversion method can be summarized as follows: (1) generate a Uniform variate  $V$  in  $(0, 1)$ ; then (2) let  $U \equiv F^{-1}(V)$ . Note that the inversion method only requires a uniform random number generator.

We now turn to the problem of estimating the parameters of the APD density  $f_X$ , 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  $\beta$  denote the parameter vector,  $\beta \equiv (\alpha, \lambda, \theta, \phi)'$ . We follow the usual convention and let  $\beta_0$  be the true value of  $\beta$  which needs to be estimated,  $\beta_0 \equiv (\alpha_0, \lambda_0, \theta_0, \phi_0)'$ . In this paper, we focus on the method of maximum likelihood, leaving the discussion of other methods of estimation, such as the GMM mentioned in Section 3, outside its scope.

Recall that for any given  $\alpha$  and  $\lambda$ ,  $0 < \alpha < 1$  and  $\lambda > 0$ , the APD random variable  $X$  in equation (2) has density  $f_X$  as given by equation (3). Let then  $X_1, \dots, X_T$  be an i.i.d. random sample from an APD distribution with density  $f_X$  parametrized by  $\beta$ , and let  $x_1, \dots, x_T$  be the corresponding observations. The APD normalized log-likelihood  $L_T(\beta)$ ,  $L_T(\beta) \equiv T^{-1} \sum_{t=1}^T \ln f_X(x_t|\beta)$ , takes the form

$$L_T(\beta) = -\ln \phi + \frac{1}{\lambda} \ln \delta_{\alpha, \lambda} - \ln \Gamma(1 + 1/\lambda) - \frac{\delta_{\alpha, \lambda}}{\phi^\lambda} T^{-1} \sum_{t=1}^T \left[ \frac{|x_t - \theta|^\lambda}{\alpha^\lambda} 1(x_t \leq \theta) + \frac{|x_t - \theta|^\lambda}{(1 - \alpha)^\lambda} 1(x_t > \theta) \right], \quad (14)$$

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

Let  $\hat{\beta}_T$  be a solution to the problem  $\max_{\beta \in B} L_T(\beta)$  where  $B$  is a compact parameter set,  $B \subset (0, 1) \times \mathbb{R}_+^* \times \mathbb{R} \times \mathbb{R}_+^*$ . The standard asymptotic normality results for maximum likelihood estimators (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 we will 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  $L_0(\beta) \equiv E[\ln f_X(X_t|\beta)]$ , 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 7 (MLE)** *Let  $X_1, \dots, X_T$  be i.i.d. from the APD distribution with an unknown parameter  $\beta_0$ ,  $\beta_0 \in \hat{B}$  with  $B$  compact. Then, the MLE  $\hat{\beta}_T$  of  $\beta_0$  is consistent and asymptotically normal,*

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

where  $J$  is the Fisher Information matrix,  $J \equiv E[(\nabla_\beta \ln f_X(X_t|\beta))(\nabla_\beta \ln f_X(X_t|\beta))']$ , and can be consistently estimated by  $\hat{J}_T$ ,  $\hat{J}_T \equiv T^{-1} \sum_{t=1}^T (\nabla_\beta \ln f_X(x_t|\hat{\beta}_T))(\nabla_\beta \ln f_X(x_t|\hat{\beta}_T))'$ . Analytic expressions of the scores,  $S_t \equiv \nabla_\beta \ln f_X(X_t|\beta)$ , are provided in the Appendix A.

We study the performance of the above MLE - small sample bias and 95% confidence interval empirical coverage - by conducting a Monte Carlo experiment. For a given value of parameter  $\beta_0 \equiv (\alpha_0, \lambda_0, \theta_0, \phi_0)'$ , we generate  $N = 10000$  replications of the sequence  $x_1, \dots, x_T$  from the APD random variable  $X$  with density  $f_X$ . The simulations are performed by using Matlab gamma random number generator with default seed values, which are obtained when the state of Matlab pseudo-random number generator is set to zero.<sup>7</sup> The parameter  $\alpha_0$  is taken to be equal to 0.1, 0.25 and 0.5, while  $\lambda_0$  takes the values 0.7, 1, 2 and 4. The parameters  $\theta_0$  and  $\phi_0$  are held fixed to 0 and 1, respectively, in all of the performed replications. The sample size  $T$  is chosen to be 100, 500, 1000 and 5000. For each replicate  $n$ ,  $1 \leq n \leq N$ , the true

---

<sup>7</sup>All computations are performed in Matlab version 6.5.0. release 13 (June 2002).

parameter  $\beta_0$  is estimated by  $\hat{\beta}_{T,n} \equiv (\hat{\alpha}_{T,n}, \hat{\lambda}_{T,n}, \hat{\theta}_{T,n}, \hat{\phi}_{T,n})'$  and for each of the components of  $\hat{\beta}_{T,n}$  we compute the 95% confidence intervals, using the covariance matrix estimator  $\hat{J}_{T,n}$  defined in Proposition 7. In Tables 2 - 5 we report the mean value of the MLEs as well as the empirical levels of the 95% confidence intervals.

[Tables 2 - 5]

As expected, the mean values of the Monte Carlo MLEs converge with the sample size  $T$  to the true values of  $\beta_0$  in all of the configurations studied in this experiment. Also, the empirical levels of the 95% confidence intervals for  $\hat{\alpha}_{T,n}$ ,  $\hat{\lambda}_{T,n}$  and  $\hat{\phi}_{T,n}$  converge with  $T$  to the nominal coverage, i.e. 95%. However, in the case of  $\hat{\theta}_{T,n}$  the convergence is particularly slow for the true values of  $\lambda$  smaller than one.

## 5 Empirical Application

We study six financial time series obtained from the Center for Research in Security Prices (CRSP) during a period from 01/02/1990 to 12/31/2002. These consist of daily prices of two indices: S&P500 and NASDAQ, two individual securities: IBM and Microsoft, and two exchange rates: British pound (BP/USD) and Japanese yen (JY/USD), expressed in terms of the US dollar. For each series of prices  $P_t$  we construct the series of log-returns,  $Y_t \equiv 100 \ln(P_t/P_{t-1})$ , which we adjust to take into account events such as stock splits on individual securities. In Table 6 we report first four unconditional moments of returns  $Y_t$ .

[Table 6]

As can be seen from Table 6, all series exhibit high values of kurtosis, ranging from 5.649 (BP/USD) to 138.579 (Microsoft). Skewness of the series studied here are generally negative.

As a first step of our analysis, we estimate a standard GARCH(1,1) model for the returns  $Y_t$ ,

$$Y_t = \mu + \sigma_t \varepsilon_t, \quad (15)$$

$$\sigma_t^2 = \omega_0 + \omega_1 \sigma_{t-1}^2 + \omega_2 Y_{t-1}^2, \quad (16)$$

where the innovations  $\varepsilon_t$  are assumed to be independent and identically distributed with some distribution  $D$  which has zero mean and unit variance, i.e.  $E(\varepsilon_t) = 0$  and  $E(\varepsilon_t^2) = 1$ . The parameter vector  $\omega \equiv (\omega_0, \omega_1, \omega_2)'$  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 for the time series of squared returns  $Y_t^2$ . We estimate  $\omega$  and  $\mu$  by using a Gaussian quasi-maximum likelihood estimator (QMLE)  $(\hat{\omega}'_T, \hat{\mu}'_T)'$ , which is consistent and asymptotically normal (see, e.g. Bollerslev and Wooldridge, 1992). Table 7 reports the values of the QMLE  $(\hat{\omega}'_T, \hat{\mu}'_T)'$  and its consistent standard errors.

[Table 7]

In Table 8 we report the first four unconditional moments of the residuals  $\hat{\varepsilon}_t \equiv \hat{\sigma}_t^{-1}(y_t - \hat{\mu}_T)$ , where  $\hat{\sigma}_t^2 = \hat{\omega}_{0T} + \hat{\omega}_{1T}\hat{\sigma}_{t-1}^2 + \hat{\omega}_{2T}y_{t-1}^2$ .

[Table 8]

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 five cases out of six, their kurtosis lies outside the interval [6, 9].

We now use the maximum likelihood result of the previous section to estimate the parameters of the distribution of residuals, by assuming it belongs to the APD family. Given that the innovations  $\varepsilon_t$  are standardized, their density is parametrized by only two parameters  $\alpha$  and  $\lambda$ , as in (12). We estimate the true value of  $(\alpha, \lambda)$  by the maximum likelihood approach



described in the previous section. For each estimate  $(\hat{\alpha}_T, \hat{\lambda}_T)'$ , we compute the location  $\hat{\theta}_T$  and scale  $\hat{\phi}_T$  of the density of  $\varepsilon_t$ , by using the restrictions of zero mean and unit variance. Consistent standard errors for  $(\hat{\alpha}_T, \hat{\lambda}_T)'$  are obtained from  $\hat{J}_T$  defined in Proposition 7. We report the estimation results in Table 9.

[Table 9]

As can be seen from Table 9, values of the asymmetry parameter  $\alpha$  range from 0.462 (Microsoft) to 0.586 (NASDAQ). A simple Wald test of the restriction  $\alpha = 1/2$  shows that in three cases out of six (NASDAQ, Microsoft and JY/USD), the value of  $\alpha$  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 six cases the values of the exponent  $\lambda$  are found to be significantly different from both 1 and 2, thus invalidating the assumptions that the innovations  $\varepsilon_t$  are double-exponential (Laplace) or normally distributed. Note that this conclusion holds even if we allow for asymmetry in the density of  $\varepsilon_t$ . Consistent with our previous results on the asymmetry parameter, we find that any time  $\alpha$  is not significantly different from zero, the same holds for the location parameter  $\theta$ . This result is as we would expect since any symmetric distribution with mean zero, necessarily has its median equal to zero. Finally, we find that in three cases out of six (IBM, BP/USD and JY/USD), the value of the scale parameter  $\phi$  is not significantly different from one (with 95% probability), although this conclusion holds weakly (with 90% probability) in the case of IBM.

At last, we are able to compute the  $\alpha$ -expected shortfall for each of the six series of innovations,  $LPM_1(\varepsilon_t, \theta)$ , which we report in Table 10. Figure 4 plots time series of conditional  $\alpha$ -expected shortfalls of returns  $Y_t$ . We construct the latter by using the fact that in conditional heteroskedasticity models such as GARCH(1,1) defined in equations (15) – (16), we have  $LPM_1(Y_t, \mu + \theta\sigma_t) = \sigma_t LPM_1(\varepsilon_t, \theta)$ .

[Table 10 and Figure 4]

It is important to note that in order to compare the riskiness of two different portfolios with returns  $Y_{1t}$  and  $Y_{2t}$  we need to compare, at each point in time  $t$ ,  $LPM_1(Y_{1t}, q_{1t})$  and  $LPM_1(Y_{2t}, q_{2t})$ , with  $q_{1t} = q_{2t}$ . In other words, we need to make sure that the two LPMs are computed at a same value.

## 6 Conclusion

This paper introduces a new family of probability distributions called the Asymmetric Power Distribution (APD). The APD family generalizes the GPD distributions to cases where the density function is asymmetric. For a given value of the asymmetry parameter  $\alpha$ ,  $0 < \alpha < 1$ , the mode of the APD probability density at zero no longer corresponds to the distribution median, as in the GPD case, but rather to its  $\alpha$ -quantile.

Expressions for quantiles, moments and moment related parameters of APD distributed random variables are also derived. The simulation of APD random variables is easily performed by using gamma or uniform random number generators. The estimation of the four parameters  $\alpha$ ,  $\lambda$ ,  $\theta$  and  $\phi$  of the APD distribution is conducted by maximum likelihood. The APD MLE is shown to be consistent and asymptotically normally distributed with the asymptotic covariance matrix equal to the inverse of the Fisher information matrix  $J$ . A consistent estimator of  $J$  is easily obtained by computing the covariance matrix of the APD scores, whose analytic expressions are provided in the paper. We confirm our theoretical findings in a Monte Carlo experiment which shows that: (1) the small sample bias of the MLE disappears, and (2) the empirical coverage of estimated confidence intervals converges to the nominal value, as the sample size increases.

An empirical application to six financial market series reveals that daily return innovations tend to be asymmetric, with asymmetry parameter  $\alpha$  significantly different from one half. For all the series examined, the exponent parameter  $\lambda$  is found to be within  $[1.21, 1.55]$ . This invalidates the assumption that the return innovations are Laplace (double-exponential) or normally distributed, even if we allow the latter to be asymmetric. Based on the above estimates for  $\alpha$

and  $\lambda$ , we are able to compute the  $\alpha$ -expected shortfalls and their standard errors.

Robustness of our results to departures from an APD density assumption, i.e. an extension to a quasi-maximum likelihood framework, and relative performances of semi-, non- and fully parametric estimators for the expected shortfall are important open questions which are left for future research.

## References

- [1] Andrews, D.W.K., (1994), “Empirical Process Methods in Econometrics”, in *Handbook of Econometrics*, 4, 2247-2294.
- [2] Artzner, P., Delbaen, F., Eber, J.M. and Heath, D. (1999), “Coherent Measures of Risk”, *Mathematical Finance*, 9, 203-228.
- [3] Bain, L.J. and Engelhardt, M. (1973), “Interval Estimation for the Two-parameter Double Exponential Distribution”, *Technometrics*, 15, 875-887.
- [4] Balakrishnan, N. and Basu, A.P. (1995), *The Exponential distribution: theory, methods and applications*, United States: Gordon and Breach.
- [5] Balakrishnan, N., Chandramouleeswaran, M.P. and Ambagaspitiya, R.S. (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.
- [6] Bawa, V.S. (1978), “Safety First, Sochastic Dominance and Optimal Portfolio Choice”, *Journal of Financial and Quantitative Analysis*, 13, 255-271.
- [7] Birnbaum, A. and Mike, V. (1970), “Asymptotically Robust Estimators of Location”, *Journal of the American Statistical Association*, 65, 1265-1282.

- [8] Bollerslev, T., and Wooldridge, J. M. (1992), “Quasi-maximum likelihood estimation and inference in dynamic models with time-varying covariances”, *Econometric Reviews*, 11, 143-179.
- [9] Follmer, H. and Schied, A. (2002), “ Convex measures of risk and trading constraints”, *Finance and Stochastics*, forthcoming.
- [10] 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.
- [11] Hanoch, G. and Levy, H. (1969), “The Efficiency Analysis of Choices Involving Risk”, *Review of Economic Studies*, 36, 335-346.
- [12] Jakuszenkow, H. (1979), “Estimation of the variance in the generalized Laplace distribution with quadratic loss function”, *Demonstratio Mathematica*, 3, 581-591.
- [13] Johnson, N.L., Kotz, S. and Balakrishnan, N. (1994), *Continuous univariate distributions*, New York : Wiley & Sons, 2nd ed.
- [14] Kotz, S., Kozubowski, T J. and Podgorski, K. (2001), *The Laplace Distribution and Generalizations: A revisit with Applications to Communications, Economics, Engineering, and Finance*, Birkhauser: Boston.
- [15] Lingappaiah, G.S. (1988), “On Two-Piece Double Exponential Distribution”, *Journal of the Korean Statistical Society*, 17, 46-55.
- [16] Markowitz, H. M. (1952), “Portfolio Selection”, *Journal of Finance*, 6, 77-91.
- [17] Nelson, D. B. (1991), “Conditional Heteroskedasticity in Asset Returns: A New Approach”, *Econometrica*, 59, 347-370.

- [18] Newey, W. K. and McFadden, D. L. (1994), “Large sample estimation and hypothesis testing”, in R. F. Engle and D. L. Newey (eds.), *Handbook of Econometrics*, 4, 2113-2247, Elsevier Science.
- [19] Sheynin, O.B. (1977), “Laplace’s Theory of Errors”, *Archive for History of Exact Sciences*, 17, 1-61.

## Appendix A

**Lemma 8 (Scores)** For any  $\beta = (\alpha, \lambda, \theta, \phi)' \in B$ , the score,  $S_t \equiv \nabla_{\beta} \ln f_X(X_t|\beta)$ , of an APD random variable  $X_t$  with density  $f_X(x|\beta) \equiv \phi^{-1}f(\phi^{-1}[x-\theta])$ , where  $f$  is as defined in equation (1), is given by

$$\begin{aligned}
S_{t,\alpha} &\equiv \frac{\partial}{\partial \alpha} \ln f_X(X_t|\beta) = \frac{\partial \delta_{\alpha,\lambda}}{\partial \alpha} \left\{ \frac{1}{\lambda \delta_{\alpha,\lambda}} - \frac{|X_t - \theta|^\lambda}{\phi^\lambda} \left[ \frac{1}{\alpha^\lambda} \mathbf{1}(X_t \leq \theta) + \frac{1}{(1-\alpha)^\lambda} \mathbf{1}(X_t > \theta) \right] \right\} \\
&\quad + \lambda \delta_{\alpha,\lambda} \frac{|X_t - \theta|^\lambda}{\phi^\lambda} \left[ \frac{1}{\alpha^{\lambda+1}} \mathbf{1}(X_t \leq \theta) - \frac{1}{(1-\alpha)^{\lambda+1}} \mathbf{1}(X_t > \theta) \right] \\
S_{t,\lambda} &\equiv \frac{\partial}{\partial \lambda} \ln f_X(X_t|\beta) = \frac{1}{\lambda^2} [-\ln \delta_{\alpha,\lambda} + \frac{\lambda}{\delta_{\alpha,\lambda}} \frac{\partial \delta_{\alpha,\lambda}}{\partial \lambda} + \Psi(1 + 1/\lambda)] \\
&\quad - \left[ \frac{\partial \delta_{\alpha,\lambda}}{\partial \lambda} + \delta_{\alpha,\lambda} \ln \frac{|X_t - \theta|}{\phi} \right] \frac{|X_t - \theta|^\lambda}{\phi^\lambda} \left[ \frac{1}{\alpha^\lambda} \mathbf{1}(X_t \leq \theta) + \frac{1}{(1-\alpha)^\lambda} \mathbf{1}(X_t > \theta) \right] \\
&\quad + \delta_{\alpha,\lambda} \frac{|X_t - \theta|^\lambda}{\phi^\lambda} \left[ \frac{\ln \alpha}{\alpha^\lambda} \mathbf{1}(X_t \leq \theta) + \frac{\ln(1-\alpha)}{(1-\alpha)^\lambda} \mathbf{1}(X_t > \theta) \right] \\
S_{t,\theta} &\equiv \frac{\partial}{\partial \theta} \ln f_X(X_t|\beta) = \lambda \frac{|X_t - \theta|^{\lambda-1}}{\phi^\lambda} \left[ \frac{\delta_{\alpha,\lambda}}{(1-\alpha)^\lambda} - 2 \cdot \mathbf{1}(X_t \leq \theta) \right] \\
S_{t,\phi} &\equiv \frac{\partial}{\partial \phi} \ln f_X(X_t|\beta) = -\frac{1}{\phi} \left\{ 1 - \lambda \delta_{\alpha,\lambda} \frac{|X_t - \theta|^\lambda}{\phi^\lambda} \left[ \frac{1}{\alpha^\lambda} \mathbf{1}(X_t \leq \theta) + \frac{1}{(1-\alpha)^\lambda} \mathbf{1}(X_t > \theta) \right] \right\},
\end{aligned}$$

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

## Appendix B (proofs can be omitted from publication)

**Notation:**

if  $V$  is a real  $n$ -vector,  $V \equiv (V_1, \dots, V_n)'$ , then  $|V|$  denotes the  $L_2$ -norm of  $V$ , i.e.  $|V|^2 \equiv V'V = \sum_{i=1}^n V_i^2$ . If  $M$  is a real  $n \times n$ -matrix,  $M \equiv (M_{ij})_{1 \leq i,j \leq n}$ , then  $|M|$  denotes the  $L_\infty$ -norm of  $M$ , i.e.  $|M| \equiv \max_{1 \leq i,j \leq n} |M_{ij}|$ .

**Proof of Lemma 8.** Recall that we have

$$\ln f_X(x|\beta) = -\ln \phi + \frac{1}{\lambda} \ln \delta_{\alpha,\lambda} - \ln \Gamma(1 + 1/\lambda) - \frac{\delta_{\alpha,\lambda}}{\phi^\lambda} \left[ \frac{|x - \theta|^\lambda}{\alpha^\lambda} \mathbf{1}(x \leq \theta) + \frac{|x - \theta|^\lambda}{(1-\alpha)^\lambda} \mathbf{1}(x > \theta) \right],$$

where  $\delta_{\alpha,\lambda} \equiv \frac{2\alpha^\lambda(1-\alpha)^\lambda}{\alpha^\lambda + (1-\alpha)^\lambda}$ . Straightforward though tedious computations show that:

$$\begin{aligned} \frac{\partial}{\partial \alpha} \ln f_X(x|\beta) &= \frac{1}{\lambda\delta_{\alpha,\lambda}} \frac{\partial \delta_{\alpha,\lambda}}{\partial \alpha} - \frac{1}{\phi^\lambda} \frac{\partial \delta_{\alpha,\lambda}}{\partial \alpha} \left[ \frac{|x-\theta|^\lambda}{\alpha^\lambda} 1(x \leq \theta) + \frac{|x-\theta|^\lambda}{(1-\alpha)^\lambda} 1(x > \theta) \right] \\ &\quad - \frac{\delta_{\alpha,\lambda}}{\phi^\lambda} \left[ -\lambda \frac{|x-\theta|^\lambda}{\alpha^{\lambda+1}} 1(x \leq \theta) + \lambda \frac{|x-\theta|^\lambda}{(1-\alpha)^{\lambda+1}} 1(x > \theta) \right] \\ &= \frac{\partial \delta_{\alpha,\lambda}}{\partial \alpha} \left\{ \frac{1}{\lambda\delta_{\alpha,\lambda}} - \frac{|x-\theta|^\lambda}{\phi^\lambda} \left[ \frac{1}{\alpha^\lambda} 1(x \leq \theta) + \frac{1}{(1-\alpha)^\lambda} 1(x > \theta) \right] \right\} \\ &\quad + \lambda \delta_{\alpha,\lambda} \frac{|x-\theta|^\lambda}{\phi^\lambda} \left[ \frac{1}{\alpha^{\lambda+1}} 1(x \leq \theta) - \frac{1}{(1-\alpha)^{\lambda+1}} 1(x > \theta) \right], \end{aligned}$$

where  $\frac{\partial \delta_{\alpha,\lambda}}{\partial \alpha} = \lambda \delta_{\alpha,\lambda} \left[ \frac{(1-2\alpha)}{\alpha(1-\alpha)} - \frac{\alpha^{\lambda-1} - (1-\alpha)^{\lambda-1}}{\alpha^\lambda + (1-\alpha)^\lambda} \right]$ .

Similarly, we have

$$\begin{aligned} \frac{\partial}{\partial \lambda} \ln f_X(x|\beta) &= -\frac{1}{\lambda^2} \ln \delta_{\alpha,\lambda} + \frac{1}{\lambda\delta_{\alpha,\lambda}} \frac{\partial \delta_{\alpha,\lambda}}{\partial \lambda} + \frac{1}{\lambda^2} \frac{\Gamma'(1+1/\lambda)}{\Gamma(1+1/\lambda)} \\ &\quad - \frac{\partial \delta_{\alpha,\lambda}}{\partial \lambda} \left[ \frac{|x-\theta|}{\phi} \right]^\lambda \left[ \frac{1}{\alpha^\lambda} 1(x \leq \theta) + \frac{1}{(1-\alpha)^\lambda} 1(x > \theta) \right] \\ &\quad - \delta_{\alpha,\lambda} \ln \left( \frac{|x-\theta|}{\phi} \right) \left[ \frac{|x-\theta|}{\phi} \right]^\lambda \left[ \frac{1}{\alpha^\lambda} 1(x \leq \theta) + \frac{1}{(1-\alpha)^\lambda} 1(x > \theta) \right] \\ &\quad + \delta_{\alpha,\lambda} \left[ \frac{|x-\theta|}{\phi} \right]^\lambda \left[ \frac{\ln \alpha}{\alpha^\lambda} 1(x \leq \theta) + \frac{\ln(1-\alpha)}{(1-\alpha)^\lambda} 1(x > \theta) \right] \\ &= \frac{1}{\lambda^2} \left[ -\ln \delta_{\alpha,\lambda} + \frac{\lambda}{\delta_{\alpha,\lambda}} \frac{\partial \delta_{\alpha,\lambda}}{\partial \lambda} + \Psi \left( 1 + \frac{1}{\lambda} \right) \right] \\ &\quad - \left[ \frac{\partial \delta_{\alpha,\lambda}}{\partial \lambda} + \delta_{\alpha,\lambda} \ln \left( \frac{|x-\theta|}{\phi} \right) \right] \frac{|x-\theta|^\lambda}{\phi^\lambda} \left[ \frac{1}{\alpha^\lambda} 1(x \leq \theta) + \frac{1}{(1-\alpha)^\lambda} 1(x > \theta) \right] \\ &\quad + \delta_{\alpha,\lambda} \frac{|x-\theta|^\lambda}{\phi^\lambda} \left[ \frac{\ln \alpha}{\alpha^\lambda} 1(x \leq \theta) + \frac{\ln(1-\alpha)}{(1-\alpha)^\lambda} 1(x > \theta) \right], \end{aligned}$$

where  $\frac{\partial \delta_{\alpha,\lambda}}{\partial \lambda} = \delta_{\alpha,\lambda} \left\{ \ln[\alpha(1-\alpha)] - \frac{\alpha^\lambda \ln \alpha + (1-\alpha)^\lambda \ln(1-\alpha)}{\alpha^\lambda + (1-\alpha)^\lambda} \right\}$  and  $\Psi(1+1/\lambda) \equiv \frac{\Gamma'(1+1/\lambda)}{\Gamma(1+1/\lambda)}$  is a digamma function, i.e. the logarithmic derivative of the gamma function  $\Gamma$ . Note that there are several different analytic expressions for  $\Psi$ . For example, for  $z > 0$ , we have  $\Psi(z+1) = \frac{d}{dz} (z-1)!$ , so that by using a series expansion of the factorial function,  $\Psi(z+1) = -\gamma + \sum_{n=1}^{\infty} \frac{z}{n(n+z)}$ ,

where  $\gamma$  is the Euler-Mascheroni constant, i.e.  $\gamma \simeq 0.577$ . Finally,

$$\begin{aligned}
\frac{\partial}{\partial \theta} \ln f_X(x|\beta) &= -\lambda \frac{\delta_{\alpha,\lambda}}{\phi^\lambda} \left[ \frac{|x-\theta|^{\lambda-1}}{\alpha^\lambda} 1(x \leq \theta) - \frac{|x-\theta|^{\lambda-1}}{(1-\alpha)^\lambda} 1(x > \theta) \right] \\
&= -\lambda \frac{\delta_{\alpha,\lambda}}{\phi^\lambda} |x-\theta|^{\lambda-1} \left[ \frac{1}{\alpha^\lambda} 1(x \leq \theta) - \frac{1}{(1-\alpha)^\lambda} 1(x > \theta) \right] \\
&= -\lambda \frac{\delta_{\alpha,\lambda}}{\phi^\lambda} |x-\theta|^{\lambda-1} \left[ \frac{(1-\alpha)^\lambda + \alpha^\lambda}{\alpha^\lambda(1-\alpha)^\lambda} 1(x \leq \theta) - \frac{1}{(1-\alpha)^\lambda} \right] \\
&= \lambda \frac{|x-\theta|^{\lambda-1}}{\phi^\lambda} \left[ \frac{\delta_{\alpha,\lambda}}{(1-\alpha)^\lambda} - 2 \cdot 1(x \leq \theta) \right],
\end{aligned}$$

and

$$\begin{aligned}
\frac{\partial}{\partial \phi} \ln f_X(x|\beta) &= -\frac{1}{\phi} + \lambda \frac{\delta_{\alpha,\lambda}}{\phi^{1+\lambda}} \left[ \frac{|x-\theta|^\lambda}{\alpha^\lambda} 1(x \leq \theta) + \frac{|x-\theta|^\lambda}{(1-\alpha)^\lambda} 1(x > \theta) \right] \\
&= -\frac{1}{\phi} \left\{ 1 - \lambda \delta_{\alpha,\lambda} \frac{|x-\theta|^\lambda}{\phi^\lambda} \left[ \frac{1}{\alpha^\lambda} 1(x \leq \theta) + \frac{1}{(1-\alpha)^\lambda} 1(x > \theta) \right] \right\}.
\end{aligned}$$

This completes the proof of Lemma 8. ■

**Proof of Lemma 2.** By definition  $F(u) \equiv \int_{-\infty}^u f(v)dv$  where  $f$  is given by equation (1).

We have

$$F(u) = \frac{\delta_{\alpha,\lambda}^{1/\lambda}}{\Gamma(1+1/\lambda)} \cdot \left\{ \int_{-\infty}^u \exp\left[-\frac{\delta_{\alpha,\lambda}}{\alpha^\lambda} |v|^\lambda\right] 1(v \leq 0) dv + \int_{-\infty}^u \exp\left[-\frac{\delta_{\alpha,\lambda}}{(1-\alpha)^\lambda} |v|^\lambda\right] 1(v > 0) dv \right\},$$

where  $\delta_{\alpha,\lambda} = \frac{2\alpha^\lambda(1-\alpha)^\lambda}{\alpha^\lambda + (1-\alpha)^\lambda}$ . First consider the case  $u \leq 0$ :  $F(u)$  then becomes

$$\begin{aligned}
F(u) &= \frac{\delta_{\alpha,\lambda}^{1/\lambda}}{\Gamma(1+1/\lambda)} \cdot \int_{-\infty}^u \exp\left[-\frac{\delta_{\alpha,\lambda}}{\alpha^\lambda} |v|^\lambda\right] dv \\
&= \frac{\delta_{\alpha,\lambda}^{1/\lambda}}{\Gamma(1+1/\lambda)} \cdot \frac{\alpha}{\lambda \delta_{\alpha,\lambda}^{1/\lambda}} \cdot \int_b^{+\infty} t^{1/\lambda-1} \exp(-t) dt,
\end{aligned}$$

where  $b \equiv \frac{\delta_{\alpha,\lambda}}{\alpha^\lambda} |u|^\lambda$ . Hence,

$$F(u) = \alpha \left[ 1 - I\left(\frac{\delta_{\alpha,\lambda}}{\alpha^\lambda} \sqrt{\lambda} |u|^\lambda, 1/\lambda\right) \right],$$

where  $I(x, \gamma)$  is Pearson's (1922) incomplete gamma function,

$$I(x, \gamma) \equiv \frac{1}{\Gamma(\gamma)} \int_0^{x\sqrt{\gamma}} t^{\gamma-1} \exp(-t) dt.$$



In the same manner, for  $u > 0$ ,  $F(u)$  becomes

$$\begin{aligned}
F(u) &= \frac{\delta_{\alpha,\lambda}^{1/\lambda}}{\Gamma(1+1/\lambda)} \cdot \left\{ \int_{-\infty}^0 \exp\left[-\frac{\delta_{\alpha,\lambda}}{\alpha^\lambda} |v|^\lambda\right] dv + \int_0^u \exp\left[-\frac{\delta_{\alpha,\lambda}}{(1-\alpha)^\lambda} |v|^\lambda\right] dv \right\} \\
&= \alpha + (1-\alpha) I\left(\frac{\delta_{\alpha,\lambda}}{(1-\alpha)^\lambda} \sqrt{\lambda} |u|^\lambda, 1/\lambda\right) \\
&= 1 - (1-\alpha) \left[1 - I\left(\frac{\delta_{\alpha,\lambda}}{(1-\alpha)^\lambda} \sqrt{\lambda} |u|^\lambda, 1/\lambda\right)\right].
\end{aligned}$$

Hence

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

which proves the Lemma 2. ■

**Proof of Lemma3.** In order to compute the inverse of  $F$  we again separately consider the two cases  $u \leq 0$  and  $u > 0$ . For all  $(u, v) \in \mathbb{R} \times (0, 1)$  we have

$$\begin{aligned}
u &\leq 0 \text{ AND } F(u) = v \\
&\iff u \leq 0 \text{ AND } v = \alpha \left[1 - I\left(\frac{\delta_{\alpha,\lambda}}{\alpha^\lambda} \sqrt{\lambda} |u|^\lambda, 1/\lambda\right)\right] \\
&\iff u \leq 0 \text{ AND } |u| = \left[\frac{\alpha^\lambda}{\delta_{\alpha,\lambda} \sqrt{\lambda}}\right]^{1/\lambda} \left[I^{-1}\left(1 - \frac{v}{\alpha}, 1/\lambda\right)\right]^{1/\lambda} \\
&\iff v \leq \alpha \text{ AND } u = H(v),
\end{aligned}$$

where the constant  $\delta_{\alpha,\lambda}$  is as in Definition 1,  $H(v) \equiv -\left[\frac{\alpha^\lambda}{\delta_{\alpha,\lambda} \sqrt{\lambda}}\right]^{1/\lambda} \left[I^{-1}\left(1 - \frac{v}{\alpha}, 1/\lambda\right)\right]^{1/\lambda}$  and  $I^{-1}(y, \gamma)$  is the inverse of the Pearson's (1922) incomplete gamma function. Similarly,

$$u > 0 \text{ AND } F(u) = v \iff v > \alpha \text{ AND } u = K(v),$$

where  $K(v) \equiv \left[\frac{(1-\alpha)^\lambda}{\delta_{\alpha,\lambda} \sqrt{\lambda}}\right]^{1/\lambda} \left[I^{-1}\left(1 - \frac{1-v}{1-\alpha}, 1/\lambda\right)\right]^{1/\lambda}$ . Thus, for any  $0 < \alpha < 1$ , the inverse  $F^{-1}$  of  $F$  is given by

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

which completes the proof of Lemma 3. ■

**Proof of Lemma 4.** For every  $r \in \mathbb{R}^+$ , we have

$$\begin{aligned}
E[U^r] &= \int_{-\infty}^{+\infty} u^r f(u) du \\
&= \frac{\delta_{\alpha,\lambda}^{1/\lambda}}{\Gamma(1+1/\lambda)} \cdot \left\{ \int_{-\infty}^0 u^r \exp\left[-\frac{\delta_{\alpha,\lambda}}{\alpha^\lambda} |u|^\lambda\right] du + \int_0^{+\infty} u^r \exp\left[-\frac{\delta_{\alpha,\lambda}}{(1-\alpha)^\lambda} |u|^\lambda\right] du \right\} \\
&= \frac{\delta_{\alpha,\lambda}^{1/\lambda}}{\Gamma(1+1/\lambda)} \cdot \left\{ \int_0^{+\infty} (-1)^r v^r \exp\left[-\frac{\delta_{\alpha,\lambda}}{\alpha^\lambda} v^\lambda\right] dv + \int_0^{+\infty} u^r \exp\left[-\frac{\delta_{\alpha,\lambda}}{(1-\alpha)^\lambda} u^\lambda\right] du \right\},
\end{aligned}$$

where we have let  $v = -u$ . Note that

$$\begin{aligned}
\int_0^{+\infty} v^r \exp\left[-\frac{\delta_{\alpha,\lambda}}{\alpha^\lambda} v^\lambda\right] dv &= \int_0^{+\infty} v^r \exp\left[-\frac{\delta_{\alpha,\lambda}}{\alpha^\lambda} v^\lambda\right] dv \\
&= \left(\frac{\alpha}{\delta_{\alpha,\lambda}^{1/\lambda}}\right)^{r+1} \frac{1}{\lambda} \int_0^{+\infty} t^{(1+r)/\lambda-1} \exp[-t] dt \\
&= \left(\frac{\alpha}{\delta_{\alpha,\lambda}^{1/\lambda}}\right)^{r+1} \frac{1}{\lambda} \Gamma\left(\frac{1+r}{\lambda}\right),
\end{aligned}$$

where  $t \equiv (\delta_{\alpha,\lambda}/\alpha^\lambda) \cdot v^\lambda$  so that  $v = (\alpha/\delta_{\alpha,\lambda}^{1/\lambda}) \cdot t^{1/\lambda}$  and  $dv = (\alpha/\delta_{\alpha,\lambda}^{1/\lambda}) \cdot t^{1/\lambda-1}/\lambda$ . Hence,

$$\begin{aligned}
E[U^r] &= \frac{\delta_{\alpha,\lambda}^{1/\lambda}}{\lambda\Gamma(1+1/\lambda)} \cdot \left\{ (-1)^r \left[\frac{\delta_{\alpha,\lambda}}{\alpha^\lambda}\right]^{-(1+r)/\lambda} \Gamma\left(\frac{1+r}{\lambda}\right) + \left[\frac{\delta_{\alpha,\lambda}}{(1-\alpha)^\lambda}\right]^{-(1+r)/\lambda} \Gamma\left(\frac{1+r}{\lambda}\right) \right\} \\
&= \frac{\Gamma((1+r)/\lambda)}{\Gamma(1/\lambda)} \cdot \delta_{\alpha,\lambda}^{-r/\lambda} [(-1)^r \alpha^{1+r} + (1-\alpha)^{1+r}],
\end{aligned}$$

which completes the proof of Lemma 4. ■

**Proof of Lemma 5.** For a given probability  $\bar{\alpha}$ ,  $0 < \bar{\alpha} \leq \alpha$ , let  $\bar{q}$  denote the  $\bar{\alpha}$ -quantile of  $U$ , i.e.  $\bar{q} \equiv F^{-1}(\bar{\alpha})$ . We then have, for any  $u \in \mathbb{R}$

$$\Pr(U \leq u | U \leq \bar{q}) = \frac{\Pr(U \leq u, U \leq \bar{q})}{\Pr(U \leq \bar{q})},$$

so that  $\Pr(U \leq u | U \leq \bar{q}) = 1$ , if  $u \geq \bar{q}$  and  $\Pr(U \leq u | U \leq \bar{q}) = F(u)/\bar{\alpha}$ , otherwise. Hence, the  $\bar{\alpha}$ -expected shortfall of  $U$  equals

$$E(\bar{q} - U | U \leq \bar{q}) = \frac{1}{\bar{\alpha}} \int_{-\infty}^{\bar{q}} (\bar{q} - u) f(u) du. \tag{17}$$

Recall that  $\bar{\alpha} \leq \alpha$  so that  $\bar{q} = F^{-1}(\bar{\alpha}) \leq F^{-1}(\alpha) = 0$ . In that case the equation (17) above becomes

$$\begin{aligned} E(\bar{q} - U|U \leq \bar{q}) &= \frac{1}{\bar{\alpha}} \frac{\delta_{\alpha,\lambda}^{1/\lambda}}{\Gamma(1+1/\lambda)} \int_{-\infty}^{\bar{q}} (\bar{q} - u) \exp[-\frac{\delta_{\alpha,\lambda}}{\alpha^\lambda}(-u)^\lambda] du \\ &= \frac{1}{\bar{\alpha}} \frac{\delta_{\alpha,\lambda}^{1/\lambda}}{\Gamma(1+1/\lambda)} \int_0^{+\infty} v \exp[-\frac{\delta_{\alpha,\lambda}}{\alpha^\lambda}(v - \bar{q})^\lambda] dv \end{aligned}$$

where we have set  $v \equiv \bar{q} - u$ . Note that we have  $u \leq \bar{q} \leq 0$  so that  $v \geq 0 \geq \bar{q}$ . Now let  $s \equiv \frac{\delta_{\alpha,\lambda}}{\alpha^\lambda}(v - \bar{q})^\lambda$  so that  $v = [\frac{\alpha^\lambda}{\delta_{\alpha,\lambda}}s]^{1/\lambda} + \bar{q}$  and  $dv = \frac{1}{\lambda}[\frac{\alpha^\lambda}{\delta_{\alpha,\lambda}}]^{1/\lambda} s^{1/\lambda-1} ds$ . The integral above becomes

$$E(\bar{q} - U|U \leq \bar{q}) = \frac{1}{\bar{\alpha}} \frac{\alpha}{\lambda \Gamma(1+1/\lambda)} \int_b^{+\infty} (\frac{\alpha}{\delta_{\alpha,\lambda}^{1/\lambda}} s^{1/\lambda} + \bar{q}) s^{1/\lambda-1} \exp(-s) ds,$$

where  $b \equiv \frac{\delta_{\alpha,\lambda}}{\alpha^\lambda}(-\bar{q})^\lambda$ . Hence,

$$\begin{aligned} E(\bar{q} - U|U \leq \bar{q}) &= \frac{1}{\bar{\alpha}} \frac{\alpha}{\lambda \Gamma(1+1/\lambda)} \left[ \frac{\alpha}{\delta_{\alpha,\lambda}^{1/\lambda}} \int_b^{+\infty} s^{2/\lambda-1} \exp(-s) ds + \bar{q} \int_b^{+\infty} s^{1/\lambda-1} \exp(-s) ds \right] \\ &= \frac{\alpha}{\bar{\alpha}} \left\{ \frac{\alpha}{\delta_{\alpha,\lambda}^{1/\lambda}} \frac{\Gamma(2/\lambda)}{\Gamma(1/\lambda)} [1 - I(\frac{\delta_{\alpha,\lambda}}{\alpha^\lambda}(-\bar{q})^\lambda \sqrt{\lambda/2}, 2/\lambda)] + \bar{q} [1 - I(\frac{\delta_{\alpha,\lambda}}{\alpha^\lambda}(-\bar{q})^\lambda \sqrt{\lambda}, 1/\lambda)] \right\}. \end{aligned}$$

Recall from Lemma 3 that  $\bar{q} = -[\frac{\alpha^\lambda}{\delta_{\alpha,\lambda}\sqrt{\lambda}}]^{1/\lambda} \cdot [I^{-1}(1 - \frac{\bar{\alpha}}{\alpha}, 1/\lambda)]^{1/\lambda}$  so that

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

which shows that (10) holds. In the special case where  $\bar{\alpha} = \alpha$  so that  $\bar{q} = 0$  we have

$$E(U|U \leq 0) = -\frac{\alpha}{\delta_{\alpha,\lambda}^{1/\lambda}} \frac{\Gamma(2/\lambda)}{\Gamma(1/\lambda)},$$

which proves (11) and thus completes the proof of Lemma 5. ■

**Proof of Lemma 6.** In order to derive the density  $f_Z$  of a standardized APD random variable  $Z \equiv \theta_Z + \phi_Z U$  we use the moment results in (7) and (8). Using  $E[Z] = 0$  and  $\text{var}(Z) = 1$  together with the equations (7) and (8) we get that

$$\phi_Z = \Gamma(1/\lambda) [\Gamma(3/\lambda) \Gamma(1/\lambda) [1 - 3\alpha + 3\alpha^2] - \Gamma(2/\lambda)^2 [1 - 2\alpha]^2]^{-1/2} \delta_{\alpha,\lambda}^{1/\lambda},$$

and

$$\theta_Z = -\Gamma(2/\lambda)[\Gamma(1/\lambda)]^{-1}(1 - 2\alpha)\phi_Z\delta_{\alpha,\lambda}^{-1/\lambda}.$$

We let

$$\mu \equiv \Gamma(2/\lambda)\{\Gamma(3/\lambda)\Gamma(1/\lambda)[1 - 3\alpha + 3\alpha^2] - \Gamma(2/\lambda)^2[1 - 2\alpha]^2\}^{-1/2},$$

which is positive due to the positivity of the variance in (8), i.e.  $\mu > 0$ . By using the above notation the scale of  $Z$  then equals  $\phi_Z = \Gamma(1/\lambda)[\Gamma(2/\lambda)]^{-1}\delta_{\alpha,\lambda}^{1/\lambda}\mu$  and its location is given by  $\theta_Z = -(1 - 2\alpha)\mu$ . We can now substitute the values of  $\theta_Z$  and  $\phi_Z$  in the general expression of an APD density  $f_Z(z) = f_Z(\theta_Z + \phi_Z u) = f(u)/\phi_Z$  where  $f$  is as defined in (1). Hence  $f_Z(z) = \phi_Z^{-1}f(\phi_Z^{-1}[z - \theta_Z])$ , i.e.

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

By noting that  $\Gamma(1 + 1/\lambda) = \Gamma(1/\lambda) \cdot 1/\lambda$  we get the equation (12) which completes the proof of Lemma 6. ■

**Proof of Proposition 7.** We 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 (14), is a consistent estimate of  $\beta_0$ . In order to do so we use the MLE consistency result by Newey and McFadden (1994, p 2131) and show that all the assumptions of their Theorem 2.5 hold. We first need to show that the identification condition (i) of Theorem 2.5 holds, i.e. if  $\beta \neq \beta_0$  then for every  $t = 1, \dots, T$  we have  $f_X(X_t|\beta) \neq f_X(X_t|\beta_0)$  (meaning  $\Pr(\{f_X(X_t|\beta) \neq f_X(X_t|\beta_0)\}) > 0$ ), where  $X_1, \dots, X_T$  are i.i.d. random variables drawn from an APD distribution with density  $f_X$  parametrized by  $\beta$ . We prove the converse of the above implication: consider the case where  $f_X(X_t|\beta) = f_X(X_t|\beta_0)$  (meaning  $\Pr(\{f_X(X_t|\beta) = f_X(X_t|\beta_0)\}) = 1$ ). In that case

$$\frac{1}{\phi}f\left(\frac{X_t - \theta}{\phi}|\alpha, \lambda\right) = \frac{1}{\phi_0}f\left(\frac{X_t - \theta_0}{\phi_0}|\alpha_0, \lambda_0\right),$$

where  $f(\cdot|\alpha, \lambda)$  is the standard APD density in (1) with parameters  $\alpha$  and  $\lambda$ . Note that  $X_t = \theta_0 + \phi_0 U_t$ , where  $U_1, \dots, U_T$  are i.i.d. standard APD random variables with density

$f(\cdot|\alpha_0, \lambda_0)$ . Hence, the above equality becomes

$$\frac{\phi_0}{\phi} f\left(\frac{\theta_0 - \theta}{\phi} + \frac{\phi_0}{\phi} U_t | \alpha, \lambda\right) = f(U_t | \alpha_0, \lambda_0). \quad (18)$$

Let  $V_t \equiv \frac{\theta_0 - \theta}{\phi} + \frac{\phi_0}{\phi} U_t$ . Then  $V_t$  is APD distributed with density

$$f_V(v) = \frac{\phi}{\phi_0} f\left(\frac{v - \frac{\theta_0 - \theta}{\phi}}{\frac{\phi_0}{\phi}} | \alpha_0, \lambda_0\right). \quad (19)$$

Using the equation (18) we then have that  $f_V(v) = f(v|\alpha, \lambda)$  which combined with (19) yields  $\alpha = \alpha_0$ ,  $\lambda = \lambda_0$ ,  $\phi/\phi_0 = 1$  and  $(\theta_0 - \theta)/\phi = 0$ . 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 boundedness condition (iv) of Theorem 2.5 requires that for all  $\beta \in B$ ,  $E[|\ln f_X(X_t|\beta)|] \leq K < \infty$  (see Lemma 2.4 in Newey and McFadden, 1994, p 2129). Recall that

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

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

$$\begin{aligned} |\ln f_X(X_t|\beta)| &\leq |\ln \phi| + \left| \ln \frac{\delta_{\alpha, \lambda}^{1/\lambda}}{\Gamma(1 + 1/\lambda)} \right| + \delta_{\alpha, \lambda} \frac{[\alpha^\lambda + (1 - \alpha)^\lambda]}{\alpha^\lambda (1 - \alpha)^\lambda} \cdot \left| \frac{X_t - \theta}{\phi} \right|^\lambda \\ &\leq |\ln \phi| + \left| \ln \frac{\delta_{\alpha, \lambda}^{1/\lambda}}{\Gamma(1 + 1/\lambda)} \right| + 2 \left| \frac{X_t - \theta}{\phi} \right|^\lambda. \end{aligned} \quad (20)$$

By using an argument similar to the one in Lemma 4 we can show that the  $\lambda$ -moment of the absolute value of a standard APD random variable  $U_t = (X_t - \theta)/\phi$  with parameters  $(\alpha_0, \lambda_0)$  is given by

$$E(|U_t|^\lambda) = \frac{\Gamma((1 + \lambda)/\lambda_0)}{\Gamma(1/\lambda_0)} \cdot \delta_{\alpha_0, \lambda_0}^{-\lambda/\lambda_0} [\alpha_0^{1+\lambda} + (1 - \alpha_0)^{1+\lambda}]. \quad (21)$$

Indeed,

$$\begin{aligned}
E(|U_t|^\lambda) &= \int_{-\infty}^{+\infty} |u|^\lambda f(u) du \\
&= \frac{\delta_{\alpha_0, \lambda_0}^{1/\lambda_0}}{\Gamma(1 + 1/\lambda_0)} \cdot \left\{ \int_{-\infty}^0 (-u)^\lambda \exp\left[-\frac{\delta_{\alpha_0, \lambda_0}}{\alpha_0^{\lambda_0}} |u|^{\lambda_0}\right] du + \int_0^{+\infty} u^\lambda \exp\left[-\frac{\delta_{\alpha_0, \lambda_0}}{(1 - \alpha_0)^{\lambda_0}} |u|^{\lambda_0}\right] du \right\} \\
&= \frac{\delta_{\alpha_0, \lambda_0}^{1/\lambda_0}}{\Gamma(1 + 1/\lambda_0)} \cdot \left\{ \int_0^{+\infty} v^\lambda \exp\left[-\frac{\delta_{\alpha_0, \lambda_0}}{\alpha_0^{\lambda_0}} v^{\lambda_0}\right] dv + \int_0^{+\infty} u^\lambda \exp\left[-\frac{\delta_{\alpha_0, \lambda_0}}{(1 - \alpha_0)^{\lambda_0}} u^{\lambda_0}\right] du \right\},
\end{aligned}$$

where, in the first integral of the right hand side of the above equality, we let  $v = -u$ . Hence,

$$\begin{aligned}
E(|U_t|^\lambda) &= \frac{\delta_{\alpha_0, \lambda_0}^{1/\lambda_0}}{\lambda_0 \Gamma(1 + 1/\lambda_0)} \cdot \left\{ \left[\frac{\delta_{\alpha_0, \lambda_0}}{\alpha_0^{\lambda_0}}\right]^{-(1+\lambda)/\lambda_0} \Gamma\left(\frac{1 + \lambda}{\lambda_0}\right) + \left[\frac{\delta_{\alpha_0, \lambda_0}}{(1 - \alpha_0)^{\lambda_0}}\right]^{-(1+\lambda)/\lambda_0} \Gamma\left(\frac{1 + \lambda}{\lambda_0}\right) \right\} \\
&= \frac{\Gamma((1 + \lambda)/\lambda_0)}{\Gamma(1/\lambda_0)} \cdot \delta_{\alpha_0, \lambda_0}^{-\lambda/\lambda_0} [\alpha_0^{1+\lambda} + (1 - \alpha_0)^{1+\lambda}],
\end{aligned}$$

which shows that (21) holds.

We therefore get that

$$E(|U_t|^\lambda) \leq \frac{\Gamma((1 + \lambda)/\lambda_0)}{\Gamma(1/\lambda_0)} \cdot 2\delta_{\alpha_0, \lambda_0}^{-\lambda/\lambda_0}. \quad (22)$$

Hence, the inequality (20) becomes

$$E[|\ln f_X(X_t|\beta)|] \leq |\ln \phi| + \left| \ln \frac{\delta_{\alpha, \lambda}^{1/\lambda}}{\Gamma(1 + 1/\lambda)} \right| + 4\delta_{\alpha_0, \lambda_0}^{-\lambda/\lambda_0} \frac{\Gamma((1 + \lambda)/\lambda_0)}{\Gamma(1/\lambda_0)}.$$

Let  $K \equiv \sup_{\beta \in B} \{|\ln \phi| + \left| \ln \frac{\delta_{\alpha, \lambda}^{1/\lambda}}{\Gamma(1 + 1/\lambda)} \right| + 4\delta_{\alpha_0, \lambda_0}^{-\lambda/\lambda_0} \frac{\Gamma((1 + \lambda)/\lambda_0)}{\Gamma(1/\lambda_0)}\}$ . By compactness of  $B$  we have  $K < \infty$ , so that for every  $\beta \in B$ , we have  $E[|\ln f_X(X_t|\beta)|] \leq K < \infty$ . We can now apply the result of Theorem 2.5 in Newey and McFadden (1994, p 2131) to show that  $\hat{\beta}_T$  is consistent, i.e.  $\hat{\beta}_T \xrightarrow{P} \beta_0$ .

We now proceed and show that  $\hat{\beta}_T$  is asymptotically normal with asymptotic covariance matrix  $J^{-1}$  with  $J$  as defined in Proposition 7, i.e.  $J = E[[\nabla_\beta \ln f_X(X_t|\beta_0)][\nabla_\beta \ln f_X(X_t|\beta_0)]']$ : in order to do so we 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  $X_1, \dots, X_T$  are i.i.d. from the APD distribution with parameter  $\beta_0$  (i.e. there is no distributional misspecification). The interior condition (ii) of Theorem 7.1 is equivalent to the assumption  $\beta_0 \in \overset{\circ}{B}$  (interior of  $B$ ). The twice differentiability condition (iii) also holds with the  $4 \times 4$  Hessian matrix of second derivatives,  $H \equiv E[\Delta_{\beta\beta} \ln f_X(X_t|\beta_0)]$ , being nonsingular. We 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  $\beta_0$ . Our numerical computations of  $H$  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  $H$  we choose not to report them here. We now show that condition (iv) of Theorem 7.1 is satisfied, i.e. that  $\sqrt{T}D_T \xrightarrow{d} N(0, J)$ , where  $D_T$  is a gradient of  $L_T(\beta)$  at  $\beta_0$ , i.e.  $D_T \equiv T^{-1} \sum_{t=1}^T \nabla_{\beta} \ln f_X(x_t|\beta_0)$ . For that, we use a standard Lindeberg-Levy Central Limit Theorem (CLT) for iid sequences (see, e.g., Theorem 5.2 in White, 2001, p 114) for which we need to show that all the elements of the asymptotic covariance matrix  $J$  are finite. Note that we have

$$|[\nabla_{\beta} \ln f_X(X_t|\beta_0)][\nabla_{\beta} \ln f_X(X_t|\beta_0)]'| = \left| \frac{\partial}{\partial \beta_{i_0}} \ln f_X(X_t|\beta_0) \right| \cdot \left| \frac{\partial}{\partial \beta_{j_0}} \ln f_X(X_t|\beta_0) \right|,$$

where  $1 \leq i_0, j_0 \leq 4$  are such that  $\max_{1 \leq i, j \leq 4} \left| \frac{\partial}{\partial \beta_i} \ln f_X(X_t|\beta_0) \cdot \frac{\partial}{\partial \beta_j} \ln f_X(X_t|\beta_0) \right| = \left| \frac{\partial}{\partial \beta_{i_0}} \ln f_X(X_t|\beta_0) \cdot \frac{\partial}{\partial \beta_{j_0}} \ln f_X(X_t|\beta_0) \right|$ . Hence, by norm equivalence we know that there exist a positive constant  $c$ , such that

$$|[\nabla_{\beta} \ln f_X(X_t|\beta_0)][\nabla_{\beta} \ln f_X(X_t|\beta_0)]'| \leq c^2 \cdot |\nabla_{\beta} \ln f_X(X_t|\beta_0)|^2.$$

In order to show that all the elements of  $J$  are finite, it suffices to show that  $E\left[\left|\frac{\partial}{\partial \beta_i} \ln f_X(X_t|\beta_0)\right|^2\right] < \infty$  for  $1 \leq i \leq 4$ : based on the results from Lemma 8, we have

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

where  $[\frac{\partial \delta_{\alpha_0, \lambda_0}}{\partial \alpha}]^2 \leq \frac{16\lambda_0^2 \delta_{\alpha_0, \lambda_0}^2}{\alpha_0^2(1-\alpha_0)^2}$ . Hence

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

where as previously  $U_t = (X_t - \theta_0)/\phi_0$  denotes a standard APD random variable with parameters  $(\alpha_0, \lambda_0)$ . By using the moment inequality in (22) we therefore have

$$E[|\frac{\partial}{\partial \alpha} \ln f_X(X_t|\beta_0)|^2] \leq \frac{64}{\alpha_0^2(1-\alpha_0)^2} [1 + 4\lambda_0^2 \frac{\Gamma((1+2\lambda_0)/\lambda_0)}{\delta_{\alpha_0, \lambda_0}^2 \Gamma(1/\lambda_0)}] < \infty. \quad (23)$$

By using the same reasoning as above, we have

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

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

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

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

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

Hence, by inequality (22) we have

$$\begin{aligned} E[|\frac{\partial}{\partial \lambda} \ln f_X(X_t|\beta_0)|^2] &\leq \frac{4}{\lambda_0^4} \{[\ln \delta_{\alpha_0, \lambda_0}]^2 + 4(\lambda_0^2 + 20)[\ln(\alpha_0(1-\alpha_0))]^2 + (\Psi(1+1/\lambda_0))^2\} \\ &\quad + \frac{64}{\Gamma(1/\lambda_0)} \left\{ \frac{\Gamma((3+2\lambda_0)/\lambda_0)}{\delta_{\alpha_0, \lambda_0}^{2(\lambda_0+1)/\lambda_0}} + \frac{\Gamma((2\lambda_0-1)/\lambda_0)}{\delta_{\alpha_0, \lambda_0}^{2(\lambda_0-1)/\lambda_0}} \right\} \\ &< \infty. \end{aligned} \quad (24)$$

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

$$E\left[\left|\frac{\partial}{\partial\theta} \ln f_X(X_t|\beta_0)\right|^2\right] \leq \frac{8\lambda_0^2}{\phi_0^2} \frac{\Gamma((2\lambda_0 - 1)/\lambda_0)}{\delta_{\alpha_0, \lambda_0}^{2(\lambda_0 - 1)/\lambda_0} \Gamma(1/\lambda_0)} < \infty. \quad (25)$$

Finally,

$$\left|\frac{\partial}{\partial\phi} \ln f_X(X_t|\beta_0)\right|^2 \leq \frac{2}{\phi_0^2} \left[1 + 4\lambda_0^2 \frac{|X_t - \theta_0|^{2\lambda_0}}{\phi_0^{2\lambda_0}}\right] = \frac{2}{\phi_0^2} [1 + 4\lambda_0^2 |U_t|^{2\lambda_0}],$$

so that by using again the result of inequality (22) we get

$$E\left[\left|\frac{\partial}{\partial\phi} \ln f_X(X_t|\beta_0)\right|^2\right] \leq \frac{2}{\phi_0^2} \left[1 + 8\lambda_0^2 \frac{\Gamma((1 + 2\lambda_0)/\lambda_0)}{\delta_{\alpha_0, \lambda_0}^{2\lambda_0} \Gamma(1/\lambda_0)}\right] < \infty. \quad (26)$$

Inequalities (23)–(26) imply that all the elements of  $J$  are finite, therefore we can use Theorem 5.2 in White (2001, p 114) to show that condition (iv) of Theorem 7.1 is satisfied. Finally, the stochastic differentiability 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  $\lambda_0 > 0$ . We can now apply the results of Theorem 7.1 in Newey and McFadden (1994) to show that  $\sqrt{T}(\hat{\beta}_T - \beta_0) \xrightarrow{d} N(0, J^{-1})$ , since in the maximum likelihood case  $H = -J$ . This completes the proof of Proposition 7. ■

## Tables and Figures

Table 1: Moments of a standard APD random variable  $U$

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

Table 2: MLE probability limits and empirical levels ( $T = 250$ )

$\alpha_0$	$\lambda_0$	$plim \alpha_T$	$plim \lambda_T$	$plim \theta_T$	$plim \varphi_T$	$p_{T\alpha}$	$p_{T\lambda}$	$p_{T\theta}$	$p_{T\varphi}$
0.1	0.7	0.1005	0.7084	0.0201	1.0054	0.8936	0.9570	0.2402	0.9390
0.25	0.7	0.2496	0.7067	0.0062	1.0074	0.8987	0.9546	0.1793	0.9317
0.5	0.7	0.4986	0.7045	-0.0011	1.0042	0.8983	0.9566	0.1162	0.9373
0.1	1	0.0987	1.0209	0.0032	0.9787	0.9097	0.9552	0.8458	0.9161
0.25	1	0.2496	1.0186	0.0039	0.9965	0.9190	0.9569	0.8530	0.9389
0.5	1	0.5000	1.0160	-0.0007	0.9992	0.9204	0.9520	0.8343	0.9412
0.1	2	0.0939	2.0727	-0.0332	0.9297	0.9198	0.9582	0.8999	0.9168
0.25	2	0.2480	2.0811	-0.0026	0.9806	0.9368	0.9669	0.9348	0.9309
0.5	2	0.4987	2.0607	-0.0040	0.9817	0.9484	0.9696	0.9500	0.9592
0.1	4	0.0848	4.3378	-0.1107	0.8411	0.9934	0.9682	0.8378	0.9913
0.25	4	0.2403	4.3202	-0.0324	0.9474	0.9287	0.9746	0.9229	0.9139
0.5	4	0.4987	4.3453	-0.0025	0.9385	0.9452	0.9712	0.9432	0.9680

NB: Monte Carlo results are obtained with  $N = 10000$  replications of the time series  $\{X_t\}$  with  $t = 1, \dots, T$ , and where  $X_t$ 's are iid APD distributed with TDGP values:  $\alpha_0, \lambda_0, \theta_0 = 0$  and  $\varphi_0 = 1$ .

Table 3: MLE probability limits and empirical levels ( $T = 500$ )

$\alpha_0$	$\lambda_0$	$plim \alpha_T$	$plim \lambda_T$	$plim \theta_T$	$plim \varphi_T$	$p_{T\alpha}$	$p_{T\lambda}$	$p_{T\theta}$	$p_{T\varphi}$
0.1	0.7	0.1003	0.7048	0.0108	1.0071	0.9009	0.9558	0.2349	0.9343
0.25	0.7	0.2498	0.7032	0.0029	1.0035	0.9053	0.9469	0.1935	0.9400
0.5	0.7	0.4997	0.7004	0.0000	0.9964	0.8985	0.9483	0.1276	0.9410
0.1	1	0.0994	1.0072	0.0007	0.9857	0.9207	0.9543	0.8945	0.9361
0.25	1	0.2495	1.0060	0.0005	0.9937	0.9415	0.9471	0.9019	0.9397
0.5	1	0.4998	1.0049	0.0004	0.9954	0.9307	0.9524	0.9047	0.9388
0.1	2	0.0971	2.0390	-0.0177	0.9659	0.9265	0.9533	0.9282	0.9278
0.25	2	0.2486	2.0396	-0.0032	0.9901	0.9467	0.9564	0.9454	0.9452
0.5	2	0.4999	2.0317	0.0008	0.9922	0.9360	0.9600	0.9408	0.9544
0.1	4	0.0943	4.1383	-0.0409	0.9376	0.9358	0.9513	0.9052	0.9322
0.25	4	0.2465	4.1550	-0.0093	0.9800	0.9432	0.9644	0.9472	0.9392
0.5	4	0.5000	4.1583	-0.0006	0.9710	0.9480	0.9648	0.9472	0.9736

NB: Monte Carlo results are obtained with  $N = 10000$  replications of the time series  $\{X_t\}$  with  $t = 1, \dots, T$ , and where  $X_t$ 's are iid APD distributed with TDGP values:  $\alpha_0, \lambda_0, \theta_0 = 0$  and  $\varphi_0 = 1$ .

Table 4: MLE probability limits and empirical levels ( $T = 1000$ )

$\alpha_0$	$\lambda_0$	$plim \alpha_T$	$plim \lambda_T$	$plim \theta_T$	$plim \varphi_T$	$p_{T\alpha}$	$p_{T\lambda}$	$p_{T\theta}$	$p_{T\varphi}$
0.1	0.7	0.1000	0.7025	0.0060	1.0007	0.9046	0.9581	0.2773	0.9388
0.25	0.7	0.2501	0.7016	0.0016	1.0016	0.9126	0.9470	0.2088	0.9430
0.5	0.7	0.5001	0.7011	0.0005	1.0018	0.9083	0.9395	0.1410	0.9339
0.1	1	0.0997	1.0048	0.0012	0.9961	0.9415	0.9516	0.9152	0.9343
0.25	1	0.2497	1.0034	0.0003	0.9983	0.9332	0.9496	0.9238	0.9401
0.5	1	0.4997	1.0017	-0.0026	0.9961	0.9368	0.9500	0.9192	0.9508
0.1	2	0.0989	2.0142	-0.0066	0.9864	0.9424	0.9583	0.9486	0.9417
0.25	2	0.2499	2.0202	0.0012	0.9967	0.9429	0.9582	0.9467	0.9463
0.5	2	0.5002	2.0164	-0.0002	0.9965	0.9488	0.9508	0.9440	0.9500
0.1	4	0.0971	4.0655	-0.0221	0.9688	0.9429	0.9582	0.9462	0.9426
0.25	4	0.2489	4.0591	-0.0024	0.9925	0.9496	0.9616	0.9488	0.9464
0.5	4	0.4993	4.0623	-0.0017	0.9862	0.9552	0.9624	0.9568	0.9672

NB: Monte Carlo results are obtained with  $N = 10000$  replications of the time series  $\{X_t\}$  with  $t = 1, \dots, T$ , and where  $X_t$ 's are iid APD distributed with TDGP values:  $\alpha_0, \lambda_0, \theta_0 = 0$  and  $\varphi_0 = 1$ .

Table 5: MLE probability limits and empirical levels ( $T = 5000$ )

$\alpha_0$	$\lambda_0$	$plim \alpha_T$	$plim \lambda_T$	$plim \theta_T$	$plim \varphi_T$	$p_{T\alpha}$	$p_{T\lambda}$	$p_{T\theta}$	$p_{T\varphi}$
0.1	0.7	0.1000	0.7001	0.0014	0.9988	0.9193	0.9519	0.4182	0.9566
0.25	0.7	0.2501	0.7006	0.0007	1.0015	0.9184	0.9482	0.3884	0.9478
0.5	0.7	0.5001	0.7006	0.0009	1.0017	0.9051	0.9544	0.2854	0.9528
0.1	1	0.1000	1.0010	0.0012	1.0001	0.9392	0.9520	0.9340	0.9409
0.25	1	0.2499	0.9995	0.0001	0.9980	0.9415	0.9466	0.9373	0.9471
0.5	1	0.5001	1.0010	0.0003	1.0004	0.9397	0.9426	0.9314	0.9494
0.1	2	0.0999	2.0025	-0.0003	0.9988	0.9505	0.9531	0.9494	0.9486
0.25	2	0.2498	2.0041	-0.0007	0.9989	0.9518	0.9526	0.9564	0.9495
0.5	2	0.4998	2.0022	-0.0004	0.9989	0.9564	0.9600	0.9548	0.9560
0.1	4	0.0995	4.0150	-0.0032	0.9950	0.9491	0.9528	0.9484	0.9496
0.25	4	0.2497	4.0140	-0.0010	0.9984	0.9496	0.9576	0.9476	0.9492
0.5	4	0.4999	4.0132	-0.0003	0.9974	0.9576	0.9528	0.9556	0.9576

NB: Monte Carlo results are obtained with  $N = 10000$  replications of the time series  $\{X_t\}$  with  $t = 1, \dots, T$ , and where  $X_t$ 's are iid APD distributed with TDGP values:  $\alpha_0, \lambda_0, \theta_0 = 0$  and  $\varphi_0 = 1$ .

Table 6: Descriptive statistics for the unconditional distribution of returns  $y_t$

<i>Return Series</i>	<i>Mean</i>	<i>Variance</i>	<i>Skewness</i>	<i>Kurtosis</i>	<i>T</i>
S&P500 Composite Index	0.027	1.105	-0.117	6.814	3278
NASDAQ Composite Index	0.033	2.606	-0.014	8.554	3278
IBM	0.036	4.508	0.037	9.021	3277
Microsoft	0.087	7.135	-5.214	138.579	3272
BP/USD Exchange Rate	0.000	0.331	0.243	5.649	3263
JY/USD Exchange Rate	-0.007	0.511	-0.562	7.177	3263

NB: number of observations:  $T$ .

Table 7: Gaussian QMLE of the GARCH(1,1) model

<i>Return Series</i>	$\omega_0$	$\omega_1$	$\omega_2$	$\mu$
S&P500 Composite Index	0.006 (0.002) [3.076]	0.933 (0.010) [94.164]	0.063 (0.009) [6.716]	0.053 (0.014) [3.806]
NASDAQ Composite Index	0.019 (0.005) [3.678]	0.891 (0.016) [56.082]	0.103 (0.015) [6.689]	0.088 (0.018) [4.808]
IBM	0.125 (0.038) [3.328]	0.908 (0.017) [54.162]	0.068 (0.012) [5.803]	0.074 (0.039) [1.908]
Microsoft	0.234 (0.059) [3.950]	0.866 (0.022) [38.492]	0.095 (0.016) [6.103]	0.159 (0.037) [4.254]
BP/USD Exchange Rate	0.004 (0.001) [3.522]	0.941 (0.009) [102.156]	0.047 (0.007) [6.717]	-0.003 (0.009) [-0.351]
JY/USD Exchange Rate	0.008 (0.002) [3.413]	0.941 (0.011) [88.544]	0.042 (0.007) [6.082]	-0.003 (0.014) [-0.180]

NB: Gaussian QMLE for the GARCH(1,1) model:  $y_t = \mu + \sigma_t \varepsilon_t$ , where  $\sigma_t^2 = \omega_0 + \omega_1 \sigma_{t-1}^2 + \omega_2 y_{t-1}^2$  and  $\varepsilon_t \sim D(0, I)$ . Consistent standard errors ( ) and t statistics [ ] are in parentheses. Values of the gaussian quasi log-likelihood: S&P500, 1.397; NASDAQ, 1.400; IBM, 1.371, Microsoft, 1.405, BP/USD, 1.394; JY/USD, 1.383.

Table 8: Descriptive statistics for the residuals

<i>Return Series</i>	<i>Mean</i>	<i>Variance</i>	<i>Skewness</i>	<i>Kurtosis</i>
S&P500 Composite Index	-0.027	1.000	-0.405	5.191
NASDAQ Composite Index	-0.034	0.999	-0.482	4.521
IBM	-0.025	1.000	-0.050	7.830
Microsoft	-0.021	1.000	0.131	4.711
BP/USD Exchange Rate	0.005	1.000	0.065	4.692
JY/USD Exchange Rate	0.008	1.000	-0.521	5.568

Table 9: MLE for the parameters of the APD density of the innovations

<i>Return Series</i>	$\alpha$	$\lambda$	$\theta$	$\varphi$
S&P500 Composite Index	0.512 (0.011) [45.511]	1.380 (0.042) [32.667]	0.035 (0.031) [1.143]	1.076 (0.035) [30.681]
NASDAQ Composite Index	0.586 (0.011) [53.461]	1.541 (0.047) [32.676]	0.263 (0.033) [7.877]	1.136 (0.033) [33.916]
IBM	0.486 (0.009) [53.241]	1.214 (0.031) [38.690]	-0.041 (0.024) [-1.751]	0.934 (0.034) [27.096]
Microsoft	0.462 (0.010) [46.187]	1.494 (0.043) [35.081]	-0.117 (0.031) [-3.773]	1.149 (0.035) [33.259]
BP/USD Exchange Rate	0.495 (0.010) [49.000]	1.319 (0.044) [29.913]	-0.016 (0.028) [-0.564]	1.028 (0.038) [27.171]
JY/USD Exchange Rate	0.530 (0.010) [51.045]	1.278 (0.041) [31.155]	0.090 (0.027) [3.327]	0.988 (0.037) [26.819]

NB: ADP MLE for the distribution of the residuals  $\varepsilon_t$ . Consistent standard errors ( ) and t statistics [ ] are in parentheses.

Table 10: MLE of the  $\alpha$ -expected shortfall of the innovations

<i>Return Series</i>	$LPM_1(\epsilon_t, \theta)$
S&P500 Composite Index	0.774 (0.018)
NASDAQ Composite Index	0.895 (0.016)
IBM	0.718 (0.014)
Microsoft	0.707 (0.017)
BP/USD Exchange Rate	0.743 (0.015)
JY/USD Exchange Rate	0.790 (0.017)

NB: Consistent standard errors obtained by simulation are reported in parentheses.