Modeling and forecasting expected shortfall with the generalized asymmetric Student-t and asymmetric exponential power distributions ☆

Dongming Zhu [a], John W. Galbraith [b,*]

[a] *School of Economics, Shanghai University of Finance and Economics, China*
[b] *Department of Economics, McGill University, Canada*

ABSTRACT

Financial returns typically display heavy tails and some degree of skewness, and conditional variance models with these features often outperform more limited models. The difference in performance may be especially important in estimating quantities that depend on tail features, including risk measures such as the expected shortfall. Here, using recent generalizations of the asymmetric Student-t and exponential power distributions to allow separate parameters to control skewness and the thickness of each tail, we fit daily financial return volatility and forecast expected shortfall for the S&P 500 index and a number of individual company stocks; the generalized distributions are used for the standardized innovations in a nonlinear, asymmetric GARCH-type model. The results provide evidence for the usefulness of the general distributions in improving fit and prediction of downside market risk of financial assets. Constrained versions, corresponding with distributions used in the previous literature, are also estimated in order to provide a comparison of the performance of different conditional distributions.

© 2011 Elsevier B.V. All rights reserved.

1. Introduction

Substantial progress has been made in the modeling and forecasting of financial asset returns since the ARCH and GARCH models were first developed in Engle (1982) and Bollerslev (1986). A key feature of this progress has been the incorporation of asymmetries and thicker tails into conditional variance models, both through the structures of the models themselves (see for example Engle and Ng, 1993, which will be used below) and through the use of thick-tailed distributions for the standardized innovations of the models. Bollerslev (1987) pioneered the use of the t-distribution for this purpose, yielding thicker-tailed predictive densities for asset returns than were available through the GARCH specification combined with Gaussian standardized innovations.[1]

Subsequent literature has explored the use of conditional distributions that are skewed as well as heavy-tailed. Hansen (1994) used a skewed t- for this purpose, and other skew extensions of the t-distribution have been proposed by Fernandez and Steel (1998), Theodossiou (1998), Branco and Dey (2001), Bauwens and Laurent (2002), Jones and Faddy (2003), Sahu et al. (2003), Azzalini and Capitanio (2003), Aas and Haff (2006) and others. An alternative to the skewed t- is the skewed exponential power distribution; see for example Fernandez et al. (1995), Ayebo and Kozubowski (2004) and Komunjer (2007). The availability of relatively long financial data

---

[1] Empirical evidence has clearly indicated that daily (for example) financial return data continue to exhibit conditional tail-fatness even after allowing for the GARCH effect; see for example Bollerslev et al. (1992).

sets has allowed the estimation and exploitation of these asymmetrical and heavy-tailed models, and empirical investigation has found that such extensions can be useful in characterizing the pattern of financial returns; see for example Mittnik and Paolella (2003) and Alberg et al. (2008).

These distributions constrain the modeling of asymmetry and tails by using two parameters which together control skewness and thickness of the left and right tails. Zhu and Zinde-Walsh (2009) and Zhu and Galbraith (2010) propose further generalizations which allow a separate parameter for each of these quantities, so that two parameters control thickness of the two tails and a third allows skewness to change independently of the tail parameters; the former extends the skewed exponential power distribution in this way, and the latter extends the skewed Student-*t*. Another approach that allows more general characterizations is to use mixtures; Rombouts and Bouaddi (2009) successfully use mixtures of exponential power distributions to fit tails. Because there is the potential that tail thickness differs non-negligibly between the left and right tails (for example, because market declines may involve more extreme price movements than market advances), these extensions can allow better fitting of the rates of tail decay and therefore can facilitate better estimation and forecasting of downside risk, for which the left tail alone is relevant. With increasingly large data sets describing financial asset returns, these less-constrained forms offer the potential for improvement in the subtlety of our fit to data, and therefore in the accuracy of forecasts of quantities related to risk.

The present paper explores the use of these distributions for modeling the conditional distribution of asset returns, and in particular for forecasting downside risk through the expected shortfall, a measure which is sensitive to losses in the extreme tail of the distribution of returns. The asymmetric exponential power distribution (AEPD) proposed by Zhu and Zinde-Walsh (2009), and the generalized asymmetric Student-*t* (AST) distribution proposed by Zhu and Galbraith (2010), are used to model the standardized innovations in the nonlinear asymmetric NGARCH model of Engle and Ng (1993). We find evidence of the potential to improve both fits to financial return data and forecasts of the expected shortfall using these distributions. Also, because these distributions generalize a number of those available in the empirical literature on financial returns, and because we treat both the skewed *t*- and the skewed exponential power distribution, our results provide an overview and comparison of the risk forecasting performance of these conditional distribution classes and of a number of special cases which have previously appeared.

The next section of the paper reviews the two distributions, in density function forms that are convenient for estimation of conditional variance models, and summarizes results on expected shortfall for these distributions. Section 3 describes the models and the data that will be used, the performance of the models in fitting volatility of financial returns according to a number of measures, and reports statistical inference on the adequacy of the distribution models. Section 4 describes methods and gives empirical results for the problem of predicting downside risk, in the form of the expected shortfall. A final section concludes.

2. Asset return models and generalized distributions

We begin by characterizing the models of financial returns and the distributions to be used for modeling the standardized innovations, in a common notation; we summarize some results in Zhu and Zinde-Walsh (2009) and Zhu and Galbraith (2010).

2.1. Models for return processes

The return process $r = \{r_t\}$ is modeled as $r_t = m_t + \varepsilon_t$ with $\varepsilon_t = \sigma_t z_t$, where m_t and σ_t^2 are the conditional mean and variance of r_t given the information set available at time *t-1* (i.e., $m_t = E(r_t|\Omega_{t-1})$ and $\sigma_t^2 = E((r_t - m_t)^2|\Omega_{t-1})$), the z_t are i.i.d. innovations with zero mean and unit variance and Ω_{t-1} is the information set available at *t-1*. This model allows different features of the return process only through different specifications of m_t, σ_t^2 and through the distribution of z_t. The combination of an ARMA specification for the conditional mean and a GARCH specification for the conditional variance is now common in empirical finance; here we simply assume that the mean is $m_t = m$, for any *t*. As in Stentoft (2008) and Zhu and Zinde-Walsh (2009), here σ_t^2 is specified to be subject to the non-linear asymmetric GARCH (NGARCH) structure of Engle and Ng (1993), which allows for leverage; Stentoft (2008), in the context of option pricing models, compares the NGARCH with the restriction to GARCH and finds that the more general NGARCH structure tends to produce both better model fits and reductions in bias and RMSE in option pricing models.

The focus of our attention is the specification of the distribution of z_t, by which the conditional distribution of the return process is determined, so that the conditional expected shortfall can be computed and predicted. Here z_t is assumed to be a standardized AST or AEPD r.v., that is, $z_t \overset{d}{=} [X - \omega(\beta)] / \delta(\beta)$, where X has one of the standard AST or AEPD distributions (with a parameter vector β, $\omega(\beta) = E(X)$ and $\delta(\beta) = \sqrt{Var(X)}$) described in the next sub-sections. The return series $\{r_t\}$ is therefore an AST-NGARCH(1,1) or AEPD-NGARCH(1,1) process,

$$r_t = m + \sigma_t z_t, \quad z_t \sim \text{i.i.d.AST}(0,1), \text{ or } \sim \text{i.i.d.AEPD}(0,1), \tag{1}$$

$$\begin{aligned} \sigma_t^2 &= b_0 + b_1 \sigma_{t-1}^2 + b_2 \sigma_{t-1}^2 (z_{t-1} - c)^2 \\ &= b_0 + b_1 \sigma_{t-1}^2 + b_2 (r_{t-1} - m - c\sigma_{t-1})^2. \end{aligned} \tag{2}$$

The leverage effect is represented by the parameter *c* in Eq. (2); $c > 0$ implies a negative correlation between the innovations in the asset return and conditional volatility of the return. The use of AST or AEPD distributions allows us to capture better the significant asymmetry present in the conditional distribution of asset returns caused by the asymmetric effect of (good and bad) shocks.

We now introduce the distributions to be used for the standardized innovations z_t, in density function form.

2.2. The AEPD density

The standard AEPD density has the form:

$$
f_{AEP}(y;\beta) = \begin{cases} \left(\dfrac{\alpha}{\alpha^*}\right)K_{EP}(p_1)\exp\left(-\dfrac{1}{p_1}\left|\dfrac{y}{2\alpha^*}\right|^{p_1}\right), & \text{if } y \le 0; \\ \left(\dfrac{1-\alpha}{1-\alpha^*}\right)K_{EP}(p_2)\exp\left(-\dfrac{1}{p_2}\left|\dfrac{y}{2(1-\alpha^*)}\right|^{p_2}\right), & \text{if } y > 0, \end{cases}
\tag{3}
$$

where $\beta = (\alpha, p_1, p_2)$ is the parameter vector, $\alpha \in (0,1)$ is the skewness parameter, $p_1 > 0$ and $p_2 > 0$ are the left and right tail parameters respectively, $K_{EP}(p) \equiv 1/[2p^{1/p}\Gamma(1+1/p)]$ is the normalizing constant of the GED distribution, and α^* is defined as

$$
\alpha^* = \alpha K_{EP}(p_1) / [\alpha K_{EP}(p_1) + (1-\alpha)K_{EP}(p_2)].
\tag{4}
$$

Note that

$$
\frac{\alpha}{\alpha^*}K_{EP}(p_1) = \frac{1-\alpha}{1-\alpha^*}K_{EP}(p_2) = \alpha K_{EP}(p_1) + (1-\alpha)K_{EP}(p_2) \equiv B_{AEP}.
\tag{5}
$$

The AEPD density function is continuous at every point and unimodal with mode at $y = 0$, but is not differentiable at $y = 0$. The parameter α^* provides scale adjustments to the left and right parts of the density to ensure continuity under changes to the shape parameters (α, p_1, p_2). The special case in which $p_1 = p_2$ is equivalent to the skewed exponential power distributions given by Komunjer (2007), Theodossiou (1998) and Fernandez et al. (1995); for a detailed discussion see Zhu and Zinde-Walsh (2009). The general form of the AEPD density is expressed as $\frac{1}{\sigma}f_{AEP}\left(\frac{y-\mu}{\sigma};\beta\right)$, where μ and σ are the location (mode) and scale parameters, respectively.

Since the AEPD density is in general asymmetric, its mode ($y = 0$) is generally different from its mean, denoted by $\omega(\beta)$,

$$
\omega(\beta) \equiv E[Y_{AEP}] = \frac{1}{B_{AEP}}\left[(1-\alpha)^2\frac{p_2\Gamma(2/p_2)}{\Gamma^2(1/p_2)} - \alpha^2\frac{p_1\Gamma(2/p_1)}{\Gamma^2(1/p_1)}\right];
\tag{6}
$$

its standard deviation $\delta(\beta) \equiv \sqrt{Var[Y_{AEP}]}$ is given by

$$
[\delta(\beta)]^2 = \frac{1}{B_{AEP}^2}\left[(1-\alpha)^3\frac{p_2^2\Gamma(3/p_2)}{\Gamma^3(1/p_2)} + \alpha^3\frac{p_1^2\Gamma(3/p_1)}{\Gamma^3(1/p_1)}\right] - [\omega(\beta)]^2.
\tag{7}
$$

When the i.i.d. innovations z_t in Model (1) are assumed to be standardized AEPD random variables, their density is given by

$$
f(z;\beta) = \delta(\beta)f_{AEP}(\omega(\beta) + \delta(\beta)z;\beta).
$$

Zhu and Zinde-Walsh (2009, Fig. 1) provide graphical illustrations of the AEPD density for various parameter values, and Komunjer (2007) for the case in which $p_1 = p_2$.

2.3. The AST density

The AST distribution in its standard probability density function form is defined by

$$
f_{AST}(y;\beta) = \begin{cases} \dfrac{\alpha}{\alpha^*}K(v_1)\left[1 + \dfrac{1}{v_1}\left(\dfrac{y}{2\alpha^*}\right)^2\right]^{-(v_1+1)/2}, & \text{if } y \le 0; \\ \dfrac{1-\alpha}{1-\alpha^*}K(v_2)\left[1 + \dfrac{1}{v_2}\left(\dfrac{y}{2(1-\alpha^*)}\right)^2\right]^{-(v_2+1)/2}, & \text{if } y > 0, \end{cases}
\tag{8}
$$

where[2] $\beta = (\alpha, v_1, v_2)$, $\alpha \in (0,1)$, $v_1 > 0$ and $v_2 > 0$, $K(v) \equiv \Gamma((v+1)/2)/[\sqrt{\pi v}\Gamma(v/2)]$, and α^* is defined as $\alpha^* = \alpha K(v_1)/[\alpha K(v_1) + (1-\alpha)K(v_2)]$; as well,

$$
\frac{\alpha}{\alpha^*}K(v_1) = \frac{1-\alpha}{1-\alpha^*}K(v_2) = \alpha K(v_1) + (1-\alpha)K(v_2) \equiv B_{AST}.
\tag{9}
$$

---

[2] $\Gamma(\cdot)$ is the gamma function.

These parameters have interpretations similar to those in the definition of the AEPD density (Eq. 3). The mean and variance of the standard AST distribution are given as follows:

$$E[Y_{AST}] = 4B_{AST}\left[-\alpha^{*2}\frac{\upsilon_1}{\upsilon_1-1} + (1-\alpha^*)^2\frac{\upsilon_2}{\upsilon_2-1}\right] \equiv \omega(\beta), \tag{10}$$

$$Var[Y_{AST}] = 4\left[\frac{\alpha\alpha^{*2}\upsilon_1}{\upsilon_1-2} + \frac{(1-\alpha)(1-\alpha^*)^2\upsilon_2}{\upsilon_2-2}\right] - [\omega(\beta)]^2 \equiv [\delta(\beta)]^2. \tag{11}$$

The density of a standardized AST random variable $[Y_{AST} - \omega(\beta)])\delta(\beta)$ has the form $f(z; \beta) = \delta(\beta)f_{AST}(\omega(\beta) + \delta(\beta)z; \beta)$, which is one of the alternatives used for the distribution of z_t in Eq. (1).

We can write a general form of the AST density, with location and scale parameters μ and σ, as $\frac{1}{\sigma}f_{AST}\left(\frac{y-\mu}{\sigma}; \beta\right)$. Again, a re-scaled version is useful for computation:

$$f_{AST}^*(y; \theta) = \begin{cases} \frac{1}{\sigma}\left[1 + \frac{1}{\upsilon_1}\left(\frac{y-\mu}{2\alpha\sigma K(\upsilon_1)}\right)^2\right]^{-(\upsilon_1+1)/2}, & \text{if } y \leq \mu; \\ \frac{1}{\sigma}\left[1 + \frac{1}{\upsilon_2}\left(\frac{y-\mu}{2(1-\alpha)\sigma K(\upsilon_2)}\right)^2\right]^{-(\upsilon_2+1)/2}, & \text{if } y > \mu, \end{cases} \tag{12}$$

where $\theta = (\alpha, \upsilon_1, \upsilon_2, \mu, \sigma)^T$.

Fig. 1 illustrates AST densities using Eq. (12) with various values for α, υ_1, and υ_2. In each case the location parameter μ is 0 and the scale parameter σ is 1. In left-hand-side panels, υ_1 is held constant at 2 while υ_2 varies; in right-hand-side panels, $\upsilon_2 = 2$ while υ_1 varies. The central panels illustrate the $\alpha = 0.5$ case: the densities are of course symmetric in cases for which $\alpha = 0.5$, $\upsilon_1 = \upsilon_2$, which are among those depicted. For the restriction to $\upsilon_1 = \upsilon_2$ but with α unrestricted, the distribution is equivalent to the skewed Student-t distributions of Hansen (1994) and Fernandez and Steel (1998) (which are equivalent after re-parameterization; see Zhu and Galbraith, 2010, p.298).

2.4. Downside risk measurement in the AEPD and AST

The Expected Shortfall (ES) is a measure of risk of loss that has gained increasing prominence in recent financial literature, particularly because (unlike the Value at Risk) it is a *coherent* risk measure, and takes into account extreme negative returns.[3] However, since the expected shortfall is defined as a conditional mean as in Eq. (13), so that this risk measure is very sensitive to extreme negative returns, it is particularly important to model the left tail of the distribution of returns accurately.

We will define the expected shortfall for random variable Y at a point in the support of the distribution q as

$$ES(q) \equiv E(Y | Y < q); \tag{13}$$

note however that the expression is sometimes defined with a negative sign, in which case the expected shortfall becomes a positive number for negative returns. Note also that while we are defining the ES at a point q in the return distribution rather than at a probability $p = F(q)$, the point q is expressed as a percentage return in the empirical study below; for example we consider quantities such as $ES(-1\%)$, i.e. the expected return conditional on the return being less than -1%. The expression for ES in terms of a confidence level p, denoted by $ES^*(p)$, is given by substituting $q = VaR(p) = F^{-1}(p)$ into the expression for $ES(q)$.

The expected shortfall is computed analytically for the AEPD and AST distributions in Zhu and Zinde-Walsh (2009) and Zhu and Galbraith (2010) respectively. For the AEPD in standard form (Eq. 3), ES can be expressed as

$$ES_{AEP}(q; \beta) = \frac{2}{F_{AEP}(q)}\left\{-\alpha\alpha^* E(|Y_{EP}(p_1)|)\left[1-G\left(h_1(q); \frac{2}{p_1}\right)\right] + (1-\alpha)(1-\alpha^*)E(|Y_{EP}(p_2)|)G\left(h_2(q); \frac{2}{p_2}\right)\right\}, \tag{14}$$

where $\beta = (\alpha, p_1, p_2)$, $F_{AEP}(q)$ is the AEPD cdf, $Y_{EP}(p_j)$ is a r.v. having the standard GED distribution with parameter p_j, $G(\cdot; 2/p_j)$ is the gamma cdf with parameter $2/p_j$ $(j = 1, 2)$, the expressions for $F_{AEP}(q)$ and $E(|Y_{EP}(p)|)$ can be found in Zhu and Zinde-Walsh (2009), $E(|Y_{EP}(p)|) = p^{1/p}\Gamma(2/p)/\Gamma(1/p)$, and

$$h_1(q) = \frac{1}{p_1}\left|\frac{min\{q, 0\}}{2\alpha^*}\right|^{p_1}, \quad h_2(q) = \frac{1}{p_2}\left|\frac{max\{q, 0\}}{2(1-\alpha^*)}\right|^{p_2}.$$

For the AST distribution in standard form (Eq. 8),

$$ES_{AST}(q; \beta) = \frac{2}{F_{AST}(q)}\left\{-\alpha\alpha^* E(|Y_t(\upsilon_1)|)[1 + w_1(q)]^{(1-\upsilon_1)/2} + (1-\alpha)(1-\alpha^*)E(|Y_t(\upsilon_2)|)\left(1-[1 + w_2(q)]^{(1-\upsilon_2)/2}\right)\right\}, \tag{15}$$

---

[3] A risk measure is called coherent if it satisfies a set of reasonable properties proposed by Artzner et al. (1997, 1999): *translation invariance, subadditivity, positive homogeneity* and *monotonicity* (see also Bradley and Taqqu, 2003).
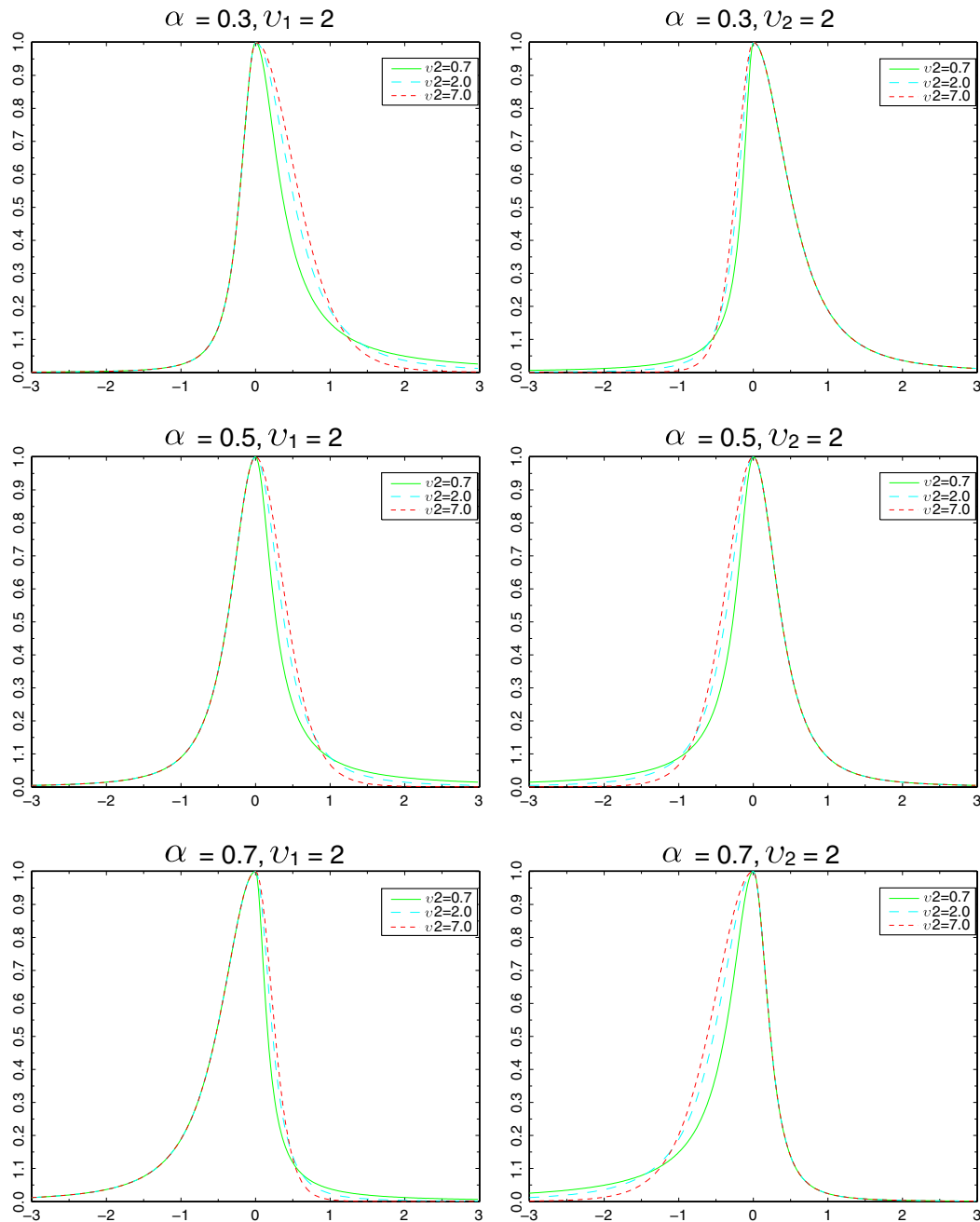
Fig. 1. AST densities for various parameter values $\mu = 0$, $\sigma = 1$.

where $\beta = (\alpha, v_1, v_2)$, $F_{\mathrm{AST}}(q)$ is the AST cdf, $Y_t(v_j)$ is a r.v. having standard Student-t distribution with parameter $v_j > 1$, the expressions for $F_{\mathrm{AST}}(q)$ and $E(|Y_t(v)|)$ can be found in Zhu and Galbraith (2010), $E(|Y_t(v)|) = \sqrt{v / \pi}\,\Gamma((v-1)/2)/\Gamma(v/2)$, and

$$ w_1(q) = \frac{1}{v_1}\left(\frac{min\{q, 0\}}{2\alpha^*}\right)^2, \quad w_2(q) = \frac{1}{v_2}\left(\frac{max\{q, 0\}}{2(1-\alpha^*)}\right)^2. $$

Substituting the expressions for $E(|Y_{\mathrm{EP}}(p_j)|)$ and $E(|Y_t(v_j)|)$ respectively into Eqs. (14) and (15) will lead to the results in Zhu and Zinde-Walsh (2009) and Zhu and Galbraith (2010).

Again, taking $q = VaR_{\mathrm{AEP}}(p) \equiv F_{\mathrm{AEP}}^{-1}(p)$ or $q = VaR_{\mathrm{AST}}(p) \equiv F_{\mathrm{AST}}^{-1}(p)$ as appropriate, we can also express the ES as a function of a confidence level $p \in (0, 1)$.
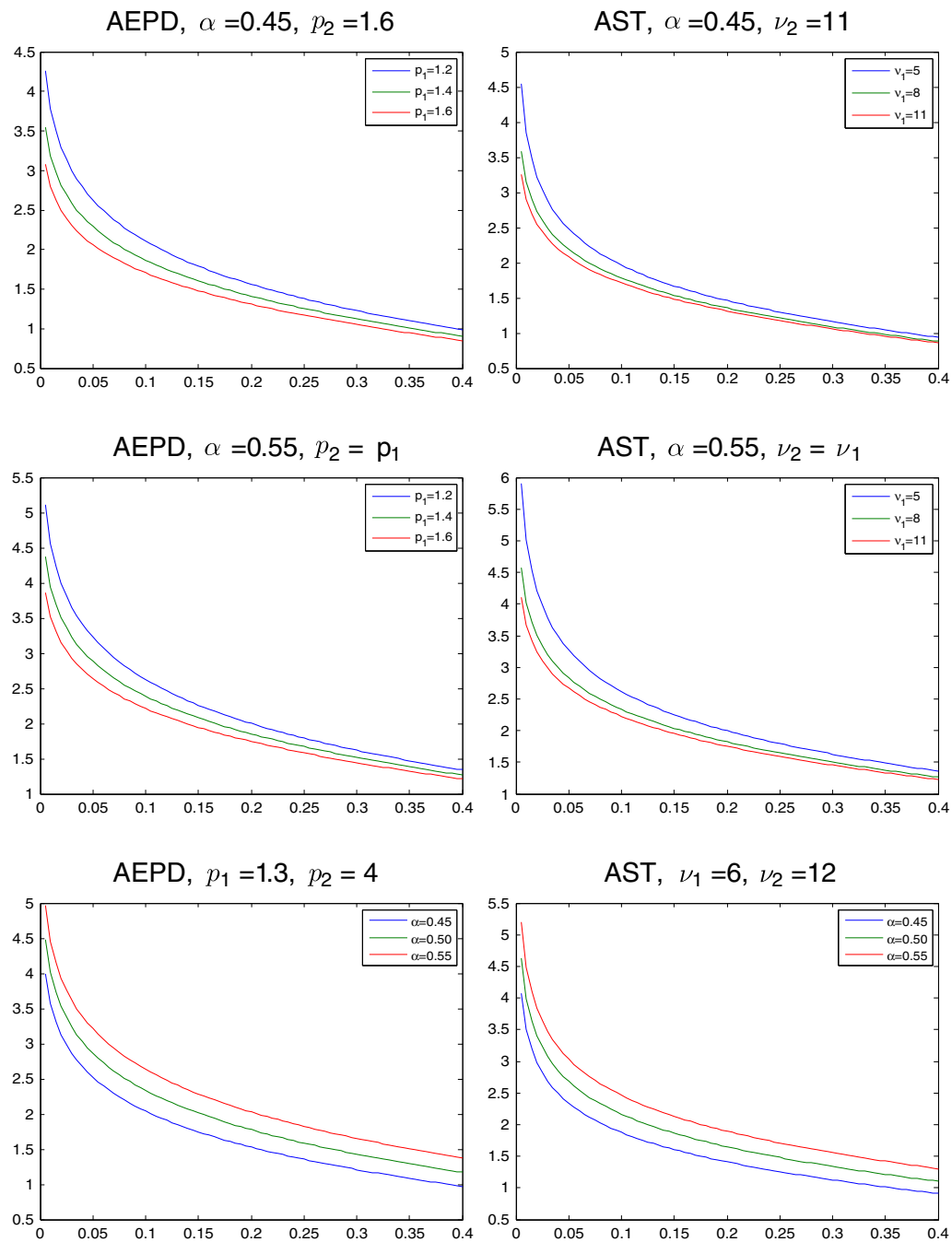
Fig. 2. Expected shortfall as a function of quantile, AEPD and AST $\mu = 0$, $\sigma = 1$.

Fig. 2 plots the ES expressed in this way for a number of examples of AEPD and AST distributions. Different values of the tail and asymmetry parameters are used for rough conformity with the range of values observed empirically (as reported below), and for comparability of the two distributions. The top panels treat cases in which α is below $\frac{1}{2}$; the middle panels show cases in which the two tail parameters are equal, corresponding therefore with the skewed exponential power distribution and skewed Student-t respectively: α is again chosen in these cases for rough conformity with empirical estimates reported below for these models. (Note that these graphics differ in form from those used by Komunjer (2007) for plotting ES.) While the patterns are fairly similar, the AST expected shortfalls tend to show slightly greater curvature, turning up somewhat more rapidly at very low quantiles (the extreme left tail). The lower panels illustrate the fact, visible in another form in Fig. 1, that higher values of the asymmetry parameter α are associated with thicker lower tails, hence greater expected shortfalls.

Table 1a
Parameter estimates for AEPD-NGARCH(1,1) models S&P 500 daily returns, 02.01.1990–31.12.2008.

	b_0	b_1	b_2	c	α	p_1	p_2
AEPD	.0103 (.0017)	.880 (.0088)	.057 (.0044)	.9947 (.0011)	.461 (.024)	1.31 (.053)	1.71 (.100)
AEPD $\alpha = \frac{1}{2}$	.0106 (.0017)	.878 (.0089)	.057 (.0044)	1.002 (.0011)		1.38 (.034)	1.59 (.056)
SEPD	.0108 (.0017)	.876 (.0091)	.057 (.0044)	1.012 (0.0011)	.536 (.0085)	1.47 (.036)	
GED	.0107 (.0017)	.877 (.0090)	.056 (.0044)	1.018 (.0011)		1.46 (.033)	

Table 1b
Parameter estimates for AST-NGARCH(1,1) models S&P 500 daily returns, 02.01.1990–31.12.2008.

	b_0	b_1	b_2	c	α	v_1	v_2
AST	.0099 (.0016)	.879 (.0088)	.056 (.0044)	1.013 (.0011)	.499 (.0171)	6.82 (.930)	16.7 (6.68)
AST, $\alpha = \frac{1}{2}$	.0099 (.0016)	.878 (.0087)	.056 (.0043)	1.013 (.0011)		6.82 (.715)	16.5 (4.96)
SST	.0103 (.0016)	.875 (.0091)	.057 (.0045)	1.024 (.0011)	.531 (.0104)	8.94 (.969)	
ST	.0101 (.0016)	.876 (.0092)	.056 (.0045)	1.031 (.0011)		8.74 (.878)	

3. Data, estimation and inference

We now consider empirical modeling and forecasting of the expected shortfall with GARCH-class models and compare forecast performance with error distributions given by either the general AEPD or AST distributions or a special case of one of these. In recording results from the special cases, we are able to provide an overview and summary of the forecast performance for a wide set of error distributions which nests many forms in the existing literature. For the AEPD, we estimate the general form and special cases corresponding with the skewed exponential power distribution (SEPD, in which $p_1 = p_2$ is imposed in the AEPD-), the generalized error distribution (GED), and the AEPD with α restricted to $\frac{1}{2}$ Correspondingly, for the AST distribution, we estimate the general form and special cases corresponding with the skewed Student-t (SST, in which $v_1 = v_2$ is imposed in the AST), the usual Student-t (ST), and the AST with α restricted to $\frac{1}{2}$ to represent asymmetry arising purely from different left and right tail behavior.

We model daily returns on the S&P 500 composite index and several individual company stocks (Adobe Systems, Bank of America, JP Morgan, Pfizer, and Starbucks) chosen to represent different industries and lengths of data history; we include two financial companies because of the particularly interesting effects of recent events on this sector. The data sets end at December 31 2008; for the S&P 500 we take two samples, one beginning on January 2, 1965 (11,076 observations), and the other for a shorter period which excludes the crash of 1987 and the immediate aftermath; the shorter sample begins January 2, 1990 (4791 obs.).[4]

3.1. Model fit and comparison

For each model, maximum likelihood estimates are obtained by numerical maximization of the corresponding likelihood function.[5] Let the parameter vector of the likelihood be ϕ, where $\phi = (m, b_0, b_1, b_2, c, \alpha, p_1, p_2)$ for the AEPD and $\phi = (m, b_0, b_1, b_2, c, \alpha, v_1, v_2)$ for the AST. The general form of the likelihood is then

$$l_T(\phi; r) = \sum_{t=1}^{T} \left\{ \log \delta - \log \sigma_t + \log f_Y\left(\omega + \delta \frac{r_t - m}{\sigma_t}; \beta\right)\right\},\tag{16}$$

where $f_Y(\cdot; \beta)$ is the AEPD density with parameters $\beta = (\alpha, p_1, p_2)'$, the AST density function with parameters $\beta = (\alpha, v_1, v_2)'$, or a constrained version of one of these. The parameters $\omega \equiv \omega(\beta)$ and $\delta \equiv \delta(\beta)$ denote the mean and standard deviation of $f_Y(\cdot; \beta)$ respectively; their expressions are given in Eqs. (6)–(7) for the AEPD and in Eqs. (10)–(11) for the AST. Since we model a return series as a stationary AEPD-NGARCH(1,1) or AST-NGARCH(1,1) process, the unconditional mean and variance are in either case $E(r_t) = m$ and $Var(r_t) = b_0/[1 - b_1 - b_2(1 + c^2)]$. To examine purely the effects of different distributional specifications of innovations we constrain the unconditional moments $E(r_t)$ and $Var(r_t) = b_0/[1 - b_1 - b_2(1 + c^2)]$ in all eight models (the general AEPD, AST and their restricted cases), to equal the same values, the sample mean and variance of returns respectively, which are their consistent estimates.

The ML estimates of the parameters of the AEPD, AST and the nested distribution classes mentioned above, and their standard errors, are presented for the post-crash sample of S&P 500 Composite Index data in Tables 1a and 1b; for individual company stocks we report below a more limited set of results on fit and forecast performance.

---

[4] Data come from CRSP (Center for Research on Security Prices, University of Chicago). The return r_t in period t is defined as $r_t = 100 \times (P_t - P_{t-1})/P_{t-1}$, where P_t is the security price or index level at time t. Other sample sizes are: Adobe, 5646; Bank of America, 6760; J.P. Morgan, 10052; Pfizer, 16632; Starbucks, 4161.

[5] The ML estimation is implemented in Matlab 7 with the command 'fmincon' and initial values $\phi_0 = (\text{mean}(r), b_0, 0.9, 0.05, 0, 0.5, 6, 6)$ for the AST-NGARCH (1,1) model and $\phi_0 = (\text{mean}(r), b_0, 0.9, 0.05, 0, 0.5, 2, 2)$ for the AEP-NGARCH(1,1) model, where b_0 is given by the sample variance of return data multiplied by $1 - b_1 - b_2 = 0.05$. Code for estimation is available from the authors.

Table 2
Likelihood ratio tests of $H_A : \alpha = \frac{1}{2}$ and H_B: equal tail parameters.

Company or index	class	$\hat{\alpha}$	H_A: stat.	$H_A: p-$	H_B: stat.	$H_B: p-$
S&P 500, 1990–	AEPD	0.461	2.08	0.15	7.14	.008
S&P 500, 1990–	AST	0.499	.005	0.94	6.95	.008
S&P 500, 1965–	AEPD	0.462	3.80	0.05	10.9	.001
S&P 500, 1965–	AST	0.498	.032	0.86	5.54	.019
Adobe Systems	AEPD	0.490	10.4	.001	5.08	.024
Adobe Systems	AST	0.436	21.1	<.001	1.24	.265
Bank of America	AEPD	0.388	15.4	<.001	3.28	.070
Bank of America	AST	0.460	9.50	.002	4.32	.038
JP Morgan	AEPD	0.450	7.90	.005	0.92	.338
JP Morgan	AST	0.460	15.3	<.001	1.75	.185
Pfizer	AEPD	0.485	22.1	<.001	0.54	.464
Pfizer	AST	0.460	24.5	<.001	3.14	.076
Starbucks	AEPD	0.423	5.42	.02	0.475	.491
Starbucks	AST	0.425	20.3	<.001	1.30	.255

Estimates of the parameters b_i and c of the NGARCH model are similar across all specifications; again, the mean m is in each case set to the sample mean of returns rather than being estimated jointly with the other parameters. Note that, while the parameter α of the general AST is estimated to be almost exactly 0.5 on this data set, values on the other data sets analyzed are typically well below 0.5. These values, and likelihood ratio tests of the hypothesis $H_0 : \alpha = \frac{1}{2}$ are presented in Table 2; on individual company data the restriction to $\alpha = \frac{1}{2}$ is invariably rejected at conventional levels, usually very strongly. The estimated values fall around 0.45, suggesting that values of α somewhat below 0.5 are an empirical regularity on equity return data, and reinforcing the suggestion that relaxing the restriction to $\alpha = \frac{1}{2}$ provides a genuine improvement in the description of such data.

Goodness-of-fit measures provide indications of the quality of data description provided by the models, which in sufficiently large samples may predict relative forecast performance. However, since expected shortfall forecasting depends on fitting the lower tail of the data, overall goodness-of-fit measures may not reveal the relevant features of the models. Below we will consider whether the goodness-of-fit measures provide a useful indication of which models will turn out, in out-of-sample comparisons, to produce the best downside risk forecasts. Following Mittnik and Paolella (2003) and Zhu and Zinde-Walsh (2009), we report four goodness-of-fit criteria; these are described in greater detail in the latter study. The first is the maximized log-likelihood value (L), an overall indicator of goodness of fit which, of course, does not embody any penalty for additional parameters. The second and the third are the modified Akaike information criterion (AICC) of Hurvich and Tsai (1989) and the Schwarz information criterion (SIC) of Schwarz (1978); these are defined in the references just given. Finally we report the Anderson and Darling (1952) statistic defined as $AD = sup_{-\infty < x < +\infty} \sqrt{T} \left| F_T(x) - \hat{F}(x) \right| / \sqrt{\hat{F}(x)\left(1 - \hat{F}(x)\right)}$, where $\hat{F}(x)$ is the estimated parametric cdf of the innovations, and $F_T(x)$ is the empirical cdf of the (*ex post*) innovations $\hat{z}_t = (r_t - \hat{m}) / \hat{\sigma}_t$, i.e. $F_T(x) = \ell / T$ where there are ℓ of these innovations less than or equal to x. Here, because the innovations are assumed to have zero mean and unit variance, their estimated cdf $\hat{F}(x)$ can be expressed as $\hat{F}(x) = F_Y\left(\hat{\omega} + \hat{\delta}x; \hat{\beta}\right)$, where $F_Y(\cdot; \beta)$ is the cdf of the AEPD or AST with the three parameters β (fewer in restricted cases), $\hat{\beta}$ is the ML estimate of β, and $\hat{\omega}$ and $\hat{\delta}$ are given by $\hat{\omega} = \omega\left(\hat{\beta}\right)$ and $\hat{\delta} = \delta\left(\hat{\beta}\right)$, respectively the estimated mean and standard deviation of $F_Y(\cdot; \beta)$. Since the AD statistic is actually the sup-norm of the normalized difference between $F_T(x)$ and $\hat{F}(x)$ and thus gives appropriate weight to the tails of the distribution, it can be used to measure goodness of fit in the tails.

Table 3 displays the four measures of goodness-of-fit for the estimated AEPD-NGARCH(1,1) and AST-NGARCH(1,1) models on the 1990 − sample of S&P 500 index data; for other samples, Tables 4a and 4b below reports the model ranked best by each measure.

As the estimate of α for the AST in Table 1b would suggest, the unrestricted AST and AST with $\alpha = \frac{1}{2}$ are essentially indistinguishable on this data set by most criteria, although the AICC chooses the restricted version, and the AD statistic the unrestricted. Either of these is preferred by all statistics to the SST (one tail parameter) and the standard Student-*t*. For the AEPD, all criteria prefer the fully general version except the AICC, which is equal to the value with $\alpha = \frac{1}{2}$ to five digits.

Table 3
Goodness-of-fit measures for all NGARCH(1,1) models S&P 500 daily returns, 02.01.1990–31.12.2008.

	L	AICC	SIC	AD
AEPD	−6210.2	12436.	12420.	15.6
AEPD, $\alpha = \frac{1}{2}$	−6211.2	12436.	12422.	19.8
SEPD	−6213.7	12442.	12428.	27.2
GED	−6222.4	12457.	12445.	42.8
AST	−6209.5	12435.	12419.	3.53
AST, $\alpha = \frac{1}{2}$	−6209.5	12433.	12419.	3.57
SST	−6213.0	12440.	12426.	6.08
ST	−6218.2	12448.	12436.	7.28

Table 4a
Preferred AEPD-class model by various criteria.

Data series	AICC	SIC	AD	MAE$_{-1.2\%, j=1}$	MAE$_{-1.2\%, j=5}$
S&P 500, 1990–	A	A	A	A	A
S&P 500, 1965–	A	A	A	A-D	A-D
Adobe Systems	A	A	A	A*	A*
Bank of America	A	A*	A	A	D*
JP Morgan	C	A	B	A*	A*
Pfizer	C	A	D	A*	A*
Starbucks	C	A	D	A*	A*

A: General AEPD; B: AEPD with $\alpha = 0.5$; C: SEPD; D: GED. '*' indicates a more favorable value than for the best element of the other model class. The notation A-D indicates that results are essentially indistinguishable within a model class; the '*' nonetheless indicates which model class provides superior results.

Table 4b
Preferred AST-class model by various criteria.

Data series	AICC	SIC	AD	MAE$_{-1.2\%, j=1}$	MAE$_{-1.2\%, j=5}$
S&P 500, 1990–	B*	A*	A*	A*	A*
S&P 500, 1965–	B*	A*	A*	A-D*	A-D*
Adobe Systems	C*	A*	A*	B	A
Bank of America	A*	A	A*	D*	D
JP Morgan	C*	A*	A*	A	D
Pfizer	A*	A*	A*	D	D
Starbucks	C*	A*	A*	C	C

A: General AST; B: AST with $\alpha = 0.5$; C: SST; D: ST. '*' indicates a more favorable value than for the best element of the other model class. The notation A-D indicates that results are essentially indistinguishable within a model class; the '*' nonetheless indicates which model class provides superior results.

In comparing the two classes, AEPD and AST, the AST is preferred in each case to the corresponding AEPD distribution, by any of the goodness-of-fit criteria.

Statistical inference takes the form of LR tests (LR $= 2(L^u - L^r)$, where L^u and L^r are respectively unrestricted and restricted log-likelihood values) of the restriction $\alpha = \frac{1}{2}$ and of the restriction to equal tail parameters ($p_1 = p_2$ or $v_1 = v_2$). On the S&P 500 data (for either sample size, as Table 2 below indicates), and for each class of distribution, the test fails to reject $\alpha = \frac{1}{2}$, but does reject the restriction to one tail parameter rather than two. However, on the individual company data (Table 2) the restriction to $\alpha = \frac{1}{2}$ is invariably rejected, and the restriction to one tail parameter is rejected in a minority of cases (four of ten at the 10% level, and only two of ten at the 5% level). Of course, the tail parameters are (being dependent on the relatively sparse extreme observations) relatively difficult to estimate precisely, so that distinguishing different left and right tail parameters may require very large samples.

Tables 4a and 4b report a summary of the best-ranked models, by the measures described earlier, for each of the AEPD and AST classes. The criteria are generally in agreement in favoring the most general model within each class, and in favoring the AST class over the AEPD. Specifically, the AD statistic invariably chooses the most general AST specification as the preferred model; the AICC also invariably prefers the AST class, but within each class often chooses a restricted model rather than the most general; the SIC always chooses the most general member of each class, and in all but one case (Bank of America) chooses the AST class rather than the AEPD.

3.2. Inference on conditional distribution specifications

As others have noted (e.g. Taylor, 2008), direct inference on expected shortfall forecasts leads to few significant distinctions because of the small number of values in the lower tails. However, we can test the conditional distribution specifications directly. The evidence in this section complements the likelihood-based inference just reported, which can test restrictions within the context of a model class, but cannot test the adequacy of the model class (here AEPD or AST) itself.

To test the applicability of the AEPD and AST distributions for modeling the standardized innovations, we apply the test of Bai (2003), which is designed precisely for models of the type used here. The test allows us to consider the null hypothesis H_0: the conditional distribution of r_t conditional on Ω_{t-1} is in the AEPD or AST family. Bai's test statistic is $T_n = sup_{0 \leq r \leq 1} \left| \widehat{W}_n(r) \right|$, where $\widehat{W}_n(r)$ is defined in Eq. (5) of Bai (2003); under some regularity conditions, $T_n \xrightarrow{d} sup_{0 \leq r \leq 1} |W(r)|$ under H_0, where $W(r)$ is a standard Brownian motion. Bai (2003) also provides critical values of the test procedure at significance levels 10%, 5%, and 1%, which are 1.94, 2.22, and 2.80, respectively. Details of the test implementation in this context are given in the Appendix A.

Table 5 provides test statistics for each of the data sets. On S&P 500 data, none of the more general specifications is rejected; in individual company data, however, there are numerous rejections, and indeed all AEPD class specifications are rejected on each company data series. More general AST specifications, by contrast, survive more tests. Nonetheless in data from Pfizer and J.P. Morgan all distributions are rejected, although less dramatically for the AST specifications. Results for these AST-class models are marginally significant (at 5%, not at 1%) for Bank of America data; given that little is known about the small-sample properties of the Bai test to outliers, we would interpret the marginal rejections very cautiously. AST and SST specifications are the only ones which are not rejected on Adobe and Starbucks data.

Table 5
Bai tests of conditional distribution.

Company or index	AEPD	SEPD	GED	AST	SST	ST
S&P 500, 1990–	0.93	1.32	3.10	1.70	1.46	3.11
S&P 500, 1965–	1.93	1.13	2.67	1.64	1.29	2.75
Adobe Systems	7.93	10.2	10.0	1.26	1.19	3.87
Bank of America	4.91	4.11	4.22	2.49	2.35	3.56
JP Morgan	5.40	5.51	3.83	4.12	3.99	6.13
Pfizer	9.22	9.94	5.54	5.59	5.46	8.22
Starbucks	4.74	4.58	3.98	1.08	0.94	3.64

Overall this evidence is compatible with that of Tables 2–4b, suggesting that AST-class models provide superior overall characterizations of the full conditional distribution. Again, however, we note that this does not imply that purely tail-related features of performance will be superior.

4. Expected shortfall prediction

4.1. ES forecast evaluation

To predict the downside risk in the period $t+j$ $(j=1, 2, 3, …)$ using the information available in period t, we must give a j-step-ahead forecast of the conditional distribution of returns r_{t+j}, $\widehat{F}_{t+j|t}(r_{t+j})$. Given the models specified in Eq. (1), the conditional distribution is time-varying only due to the time-varying conditional mean and variance. Therefore forecasting the conditional distribution amounts to estimating the parameters of the model using the data available at time t, and then forecasting the conditional mean (m_{t+j}) and variance (σ_{t+j}^2) of r_{t+j}.

The forecast of the conditional ES can be divided into two steps. First of all, we give the forecasts of the conditional mean and variance. Theoretically, the j-step-ahead forecast of σ_{t+j}^2, denoted by $\sigma_{t+j|t}^2$, is defined in the usual way: $\sigma_{t+j|t}^2 \equiv E(\sigma_{t+j}^2|\Omega_t)$;

$$\sigma_{t+1|t}^2 = b_0 + b_1\sigma_t^2 + b_2(r_t - m - c\sigma_t)^2, \tag{17}$$

$$\sigma_{t+j|t}^2 = b_0 + \left[b_1 + b_2\left(1 + c^2\right)\right]\sigma_{t+j-1|t}^2, \; j \geq 2. \tag{18}$$

Similarly, the j-step-ahead forecast of m_{t+j} is expressed as $m_{t+j|t} \equiv E(m_{t+j}|\Omega_t)$; for Models (1) $m_{t+j|t} = m.$[6] Note here that these forecasts of the conditional mean and variance do not depend on z_t's distribution specification.[7] We can therefore use the two-step method as in Komunjer (2007) to estimate m and the parameters of the NGARCH equation by using data up to time t and a Gaussian quasi-maximum likelihood estimator (QMLE), which are consistent and are expressed as $\left(\widehat{m}_t, \widehat{b}_{0t}, \widehat{b}_{1t}, \widehat{b}_{2t}, \widehat{c}_t\right)$. Substituting these estimates into the expressions for $m_{t+j|t}$ and $\sigma_{t+j|t}^2$ (see Eqs. (17) and (18)) yields $\widehat{m}_{t+j|t}$ and $\widehat{\sigma}_{t+j|t}^2$, the estimates of the j-step-ahead forecasts of m_{t+j} and σ_{t+j}^2,

$$\widehat{m}_{t+j|t} = \widehat{m}_t,$$

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

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

Based on these forecasts in time t, the return r_{t+j} can be expressed approximately as $\widehat{r}_{t+j} \equiv \widehat{m}_{t+j|t} + \widehat{\sigma}_{t+j|t}z_{t+j}$ that has an AEPD or AST-type conditional distribution with mean $\widehat{m}_{t+j|t}$ and standard deviation $\widehat{\sigma}_{t+j|t}$, and thus the j-step-ahead forecast of the conditional ES, $\widehat{ES}_{t+j|t}(q) \equiv E(\widehat{r}_{t+j}|\widehat{r}_{t+j}<q, \Omega_t)$, is expressed as

$$\widehat{ES}_{t+j|t}(q) = \widehat{m}_{t+j|t} + \widehat{\sigma}_{t+j|t}\left[\frac{ES_A\left(q_{t+j};\beta\right) - \omega(\beta)}{\delta(\beta)}\right], \tag{19}$$

where $ES_A(q;\beta)$ represents either $ES_{AEP}(q;\beta)$ or $ES_{AST}(q;\beta)$ given in Eqs. (14) and (15), $q_{t+j} = \omega(\beta) + \delta(\beta)(q - \widehat{m}_{t+j|t}) / \widehat{\sigma}_{t+j|t}$, $\omega(\beta)$ and $\delta(\beta)$ are defined in Eqs. (6)–(7) for the AEPD and in Eqs. (10)–(11) for the AST. Obviously, the conditional ES depends on all the specifications of m_t, σ_t^2 and z_t's distribution.

In the second step, we estimate the parameter vector β of the distribution of z_t specified in Eq. (1), which is necessary for evaluation of $\widehat{ES}_{t+j|t}(q)$ in Eq. (19). Note that if Model (1) is correctly specified, then the *ex post* innovations $\widehat{z}_i \equiv (r_i - \widehat{m}_t) / \widehat{\sigma}_i$

---

[6] For an ARMA specification of the conditional mean, the j-step-ahead forecast of m_{t+j} based on the information available at time t, $m_{t+j|t} \equiv E(m_{t+j}|\Omega_t)$, can also be evaluated; for example, if $m_t = m + a_1 r_{t-1}$, then $m_{t+j|t} = m(1 - a_1^j)/(1 - a_1) + a_1^j r_t$.

[7] However, for the TGARCH(1,1) specification, $\sigma_t^2 = b_0 + b_1\sigma_{t-1}^2 + \sigma_{t-1}^2 z_{t-1}^2[b_2 + c1(z_{t-1}<0)]$, $\sigma_{t+j|t}^2$ depends not only on the specifications of conditional mean and variance but also on the specification of z_t's distribution through the term $E[z_t^2 1(z_t<0)]$.

(i = 1, 2, ..., t) constructed by using the QMLE $\left(\widehat{m}_t, \widehat{b}_{0t}, \widehat{b}_{1t}, \widehat{b}_{2t}, \widehat{c}_t\right)$ from the first step (with $\widehat{\sigma}_i^2 = \widehat{b}_{0t} + \widehat{b}_{1t}\widehat{\sigma}_{i-1}^2 + \widehat{b}_{2t}\left(r_{i-1} - \widehat{m}_t - \widehat{c}_t\widehat{\sigma}_{i-1}\right)^2$), should approximately have a standardized AEPD or AST distribution. Therefore, a consistent MLE of β, denoted by $\widehat{\beta}_t$, can be obtained by fitting the standardized AEPD or AST distribution using the *ex post* innovations $\left\{\widehat{z}_i\right\}_{i=1}^t$. We evaluate $\widehat{ES}_{t+j|t}(q)$ by substituting $\widehat{\beta}_t$ into Eq. (19).

4.2. Empirical ES forecast results

Fig. 3 illustrates the one-step-ahead forecast expected shortfalls on all six data series using an initial sample for estimation of 1000, with recursive updating of parameter estimates thereafter, using the general AST distribution and a − 1% return threshold. Note that we are computing these expected shortfalls at a particular loss rather than at a particular tail probability, as is more
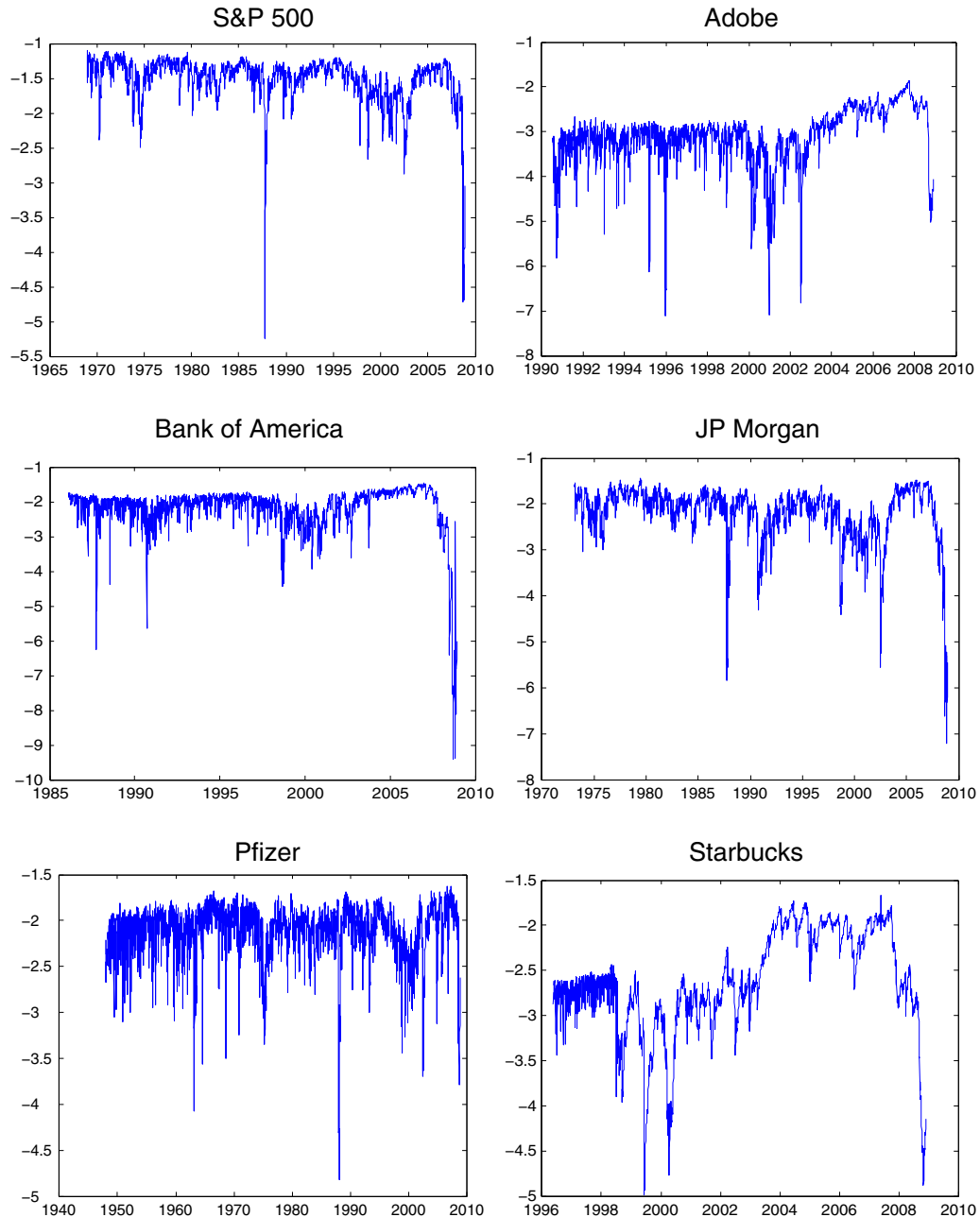
Vertical axes: percentage return; horizontal axes: date.

Fig. 3. One-day-ahead forecast expected shortfalls from the general AST: $E(r_t|r_t \leq -1\%)$, initial sample = 1000.

Table 6a
Predictive performance for expected shortfall risk [a] AEPD-NGARCH(1,1) models, S&P 500 daily returns, 02.01.1990–31.12.2008.

q	−1.2%	−1%	−0.8%	−0.6%
$j=1$	ME_1, MAE_1	ME_1, MAE_1	ME_1, MAE_1	ME_1, MAE_1
AEPD	− 0.062, **0.372**	− 0.030, **0.374**	− 0.012, **0.356**	0.018, **0.344**
AEPD, $\alpha = \frac{1}{2}$	− 0.043, 0.392	− 0.019, 0.391	− 0.010, 0.370	0.010, 0.359
SEPD	− 0.029, 0.400	− 0.007, 0.399	0.001, 0.376	0.020, 0.354
GED	0.005, 0.397	0.025, 0.397	0.030, 0.373	0.047, 0.355
$j=5$	ME_5, MAE_5	ME_5, MAE_5	ME_5, MAE_5	ME_5, MAE_5
AEPD	0.031, **0.331**	0.061, **0.325**	0.058, **0.312**	0.071, **0.299**
AEPD, $\alpha = \frac{1}{2}$	0.059, 0.349	0.080, 0.339	0.065, 0.320	0.066, 0.309
SEPD	0.073, 0.360	0.093, 0.349	0.076, 0.328	0.076, 0.304
GED	0.103, 0.364	0.121, 0.354	0.103, 0.330	0.101, 0.311

[a] The entries are the mean errors $ME_j(q) \equiv \widehat{ES}_j^M(q) - \widehat{ES}_j(q)$ and the mean absolute errors $MAE_j(q)$ defined in Eq. (22), multiplied by 100, for the threshold losses (negative returns) $q = -1.2\%, -1\%, -0.8\%, -0.6\%$ and $j = 1, 5$. The lowest MAE for each case is given in bold.

Table 6b
Predictive performance for expected shortfall risk [a] AST-NGARCH(1,1) models, S&P 500 daily returns, 02.01.1990–31.12.2008.

q	−1.2%	−1%	−0.8%	−0.6%
$j=1$	ME_1, MAE_1	ME_1, MAE_1	ME_1, MAE_1	ME_1, MAE_1
AST	− 0.034, **0.367**	0.001, **0.378**	0.018, **0.363**	0.043, **0.354**
AST, $\alpha = \frac{1}{2}$	− 0.034, 0.374	− 0.003, 0.382	0.012, 0.365	0.034, 0.362
SST	− 0.006, 0.391	0.020, 0.397	0.029, 0.376	0.048, **0.354**
ST	0.021, 0.391	0.045, 0.398	0.052, 0.377	0.068, 0.362
$j=5$	ME_5, MAE_5	ME_5, MAE_5	ME_5, MAE_5	ME_5, MAE_5
AST	0.057, **0.328**	0.092, **0.333**	0.088, **0.321**	0.097, 0.312
AST, $\alpha = \frac{1}{2}$	0.059, 0.334	0.090, 0.335	0.083, 0.322	0.088, 0.318
SST	0.089, 0.356	0.114, 0.353	0.101, 0.334	0.103, **0.310**
ST	0.113, 0.362	0.137, 0.360	0.123, 0.339	0.122, 0.321

[a] See note to Table 6a.

usual. There is considerable variation across securities, as well as through time, in the typical level of expected shortfall; note that vertical scales differ, and in particular that the two financial stocks have wide vertical scales to accommodate the high expected shortfalls observed recently.

To check predictive performance out-of-sample, we split the data samples. We use a larger initial sample of $N = 2000$ points for more precise initial estimation than was required for the graphical results depicted above; the remaining sample points (sub-samples of varying size across the data sets) are used for out-of-sample evaluation. We recursively evaluate $\widehat{ES}_{t+j|t}(q)$, $N = 2000 \leq t \leq T - j$, for one and five steps ahead ($j = 1, 5$), and we set the threshold (loss) returns $q = -1.2\%, -1\%, -0.8\%, -0.6\%$ in order to have a substantial number of sample points at which losses exceed q. For each of (j, q), if the model is correctly specified we expect the average of the observed r_{t+j}-values ($r_{N+j}, ..., r_T$) less than q should be approximately equal to the $\widehat{ES}_{t+j|t}(q)$ predicted by the model. If the observed expected shortfall

$$\widehat{ES}_j(q) \equiv \frac{1}{J} \sum_{t=N}^{T-j} r_{t+j} 1\{r_{t+j} < q\}, \text{ where } J = \sum_{t=N}^{T-j} 1\{r_{t+j} < q\} \tag{20}$$

is higher[8] than a model's average predictive ES,

$$\widehat{ES}_j^M(q) \equiv \frac{1}{J} \sum_{t=N}^{T-j} \widehat{ES}_{t+j|t}(q) 1\{r_{t+j} < q\}, \tag{21}$$

then the model tends to overestimate the risk. This is measured by the mean error, $ME_j(q) \equiv \widehat{ES}_j M(q) - \widehat{ES}_j(q)$; a negative value therefore represents overestimation of this measure of risk. Another important measure of predictive out-of-sample performance is the mean absolute error

$$MAE_j(q) \equiv \frac{1}{J} \sum_{t=N}^{T-j} \left| \widehat{ES}_{t+j|t}(q) - \widehat{ES}_j(q) \right| 1\{r_{t+j} < q\}. \tag{22}$$

---

[8] i.e. less negative: recall by Eq. (13) that we do not use a sign change in the definition of ES.

Tables 6a and 6b show the predictive performance for the expected shortfall risk on the S&P 500 data, 1990–; the entries in the table are the mean errors $100\text{ME}_j(q)$ and the mean absolute errors $100\text{MAE}_j(q)$ of the expected shortfall predictions for one and five steps ahead. From the mean errors ME_j, we see that both the AEPD-class and AST-class models tend to overestimate the risk (have a negative mean error) at horizon $j=1$ for larger loss thresholds ($q=-1.2\%, -1\%$), but show mean errors close to zero for the smaller thresholds ($q=-0.8\%, -0.6\%$). At the longer horizon $j=5$, all models underestimate risk on this data set. The AST-class tends to underestimate ES more than the AEPD-class. In the mean absolute errors MAE_j, the AST-class has better performance than the AEPD-class for larger loss thresholds ($q=-1.2\%$), but is in general worse for smaller loss thresholds; the more general distributions outperform the corresponding subclasses in almost all cases.

We do not record the information analogous to that in Tables 6a and 6b for each of the other data sets, but instead provide a summary of these results for each of the individual companies: the best-performing model by MAE at the smallest threshold, in 1-step and 5-step forecasts, is recorded in Tables 4a and 4b along with the model selection criteria. Results are presented separately for AEPD- and AST-class models; a '*' indicates which of these classes provided a superior results on each criterion.

With respect to model fit, as we have noted, each of the criteria tends clearly to favor the AST class rather than the AEPD; within each of these classes the most general variant is usually selected by AD and always by SIC, but there is more variability in within-class results from the AICC (which nonetheless invariably prefers the AST class to AEPD). While not unambiguous, therefore, the overall evidence tends to favor the most general AST specifications. Moreover, in likelihood ratio tests of the restriction to $\alpha = \frac{1}{2}$, the restriction is strongly rejected on company data, although not on the composite index. The estimated values of α are invariably below 0.5, suggesting a genuine empirical regularity. The restriction to one tail parameter rather than two is also rejected in general. Overall, both information criteria and statistical inference tend to suggest that the additional parameters in the most general AEPD or AST specifications are valuable, on the available sample sizes, in characterizing financial return volatility.

In the empirical prediction of expected shortfall, the AEPD class tends to outperform the AST on these data sets, and in most cases the model which provides the best ES predictions is the most general form of AEPD; however, there are exceptions. The best model for forecasting expected shortfall, whichever class it comes from, tends to be one of the fully-general models with α and both tail parameters allowed to vary. The empirical prediction results therefore tend to confirm the result just noted, that the additional parameters of the most general specifications can provide genuine improvements in our ability to model and forecast financial return data, relative to the more restrictive specifications available in the previous literature.

5. Concluding remarks

Useful measures of downside risk must reflect the potential for extreme outcomes, but measuring and forecasting such risk for financial assets are challenging because of the asymmetry and heavy-tailedness of return distributions. Nonetheless, a great deal of progress has recently been made in treating these features more realistically, with accompanying improvements in forecasting power for downside risk measures.

The present paper has attempted further progress on this research program. Using general, asymmetric exponential power and Student-*t* distributions with separate parameters to control skewness and the thickness of each tail, we have greater flexibility to use information in a large sample of data, and in particular to avoid constraining the left and right tails to have the same thickness. In doing so we can potentially obtain better estimates of the thickness of the left tail, with corresponding potential improvements in forecasting power for risk of loss, measured here by the expected shortfall.

The natural question to ask of any generalization of a model is whether the additional parameters provide discernible improvements, or whether the additional flexibility is unnecessary or is dominated at available sample sizes by the efficiency cost of estimating additional parameters. Here, evidence from model fit criteria, statistical inference and out-of-sample forecasts tends to suggest that the additional generality does in many cases produce improvements in both fit and forecasting power relative to those with more restricted specifications of the distribution of standardized innovations. Likelihood ratio tests of the restriction that sets the additional parameter α, controlling skewness independently of the tail parameters, to its central value of $\frac{1}{2}$ generally reject the restriction strongly, suggesting that the additional flexibility offered by the more general distributions has genuine value in describing financial returns data. The parameter is typically estimated to be less than $\frac{1}{2}$. Tail parameters are less precisely estimated, and tests of the restriction to equal tail parameters often fail to reject on these samples, although both goodness of fit and forecasting performance generally produce better results with the additional parameter.

The balance of evidence suggests the potential for genuine improvement in risk management models and forecasts relative to restricted variants of these models available in the prior literature, although there will certainly be cases in which more restricted versions will be appropriate. Whether improvements in fits and forecasts would also be visible in asset return series where asymmetry may be less important, such as some exchange rate series, remains to be seen.

Appendix A

Using the numerical method of Bai (2003, Appendix B), the test statistic can be computed as:

$$T_n \approx \sup_{1 \le j \le n} \left| \widehat{W}_n(v_j) \right| = \max_{1 \le j \le n} \sqrt{n} \left| \frac{j}{n} - \frac{1}{n} \sum_{k=1}^{j} [g'(v_k)]^T C_k^{-1} D_k (v_k - v_{k-1}) \right|,$$

where $D_k = \sum_{i=k}^{n} g'(v_i)$, $C_k = \sum_{i=k}^{n} g'(v_i)[g'(v_i)]^T(v_{i+1} - v_i)$, and where $v_1 < v_2 < ... < v_n$ are ordered values of $\widehat{U}_t = F_{AEP}(\omega(\beta) + \delta(\beta)\,\widehat{z}_t; \beta)$ or $F_{AST}(\omega(\beta) + \delta(\beta)\,\widehat{z}_t; \beta)$, where $F_{AEP}(\cdot)$ and $F_{AST}(\cdot)$ are the cdf's of standard AEPD and AST distributions. Expressions for these can be found in Zhu and Zinde-Walsh (2009) and Zhu and Galbraith (2010), respectively. In addition, we impose $v_0 = 0$ and $v_{n+1} = 1$. What remains is to derive the function $g(r)$ so that we can compute the derivative values $g'(v_i)$ of $g(\cdot)$; by Eq. (8) of Bai (2003),

$$g(r) = (g_1, g_2, g_3)^T = \left(r, f_0\left(F_0^{-1}(r)\right), f_0\left(F_0^{-1}(r)\right) \, F_0^{-1}(r)\right)^T,$$

where f_0 and F_0 are respectively the pdf and cdf of the standardized AEPD or AST, the derivative of $g(\cdot)$ is given by $g'_1(r) = 1$, $g'_2(r) = f'_0(F_0^{-1}(r))/f_0(F_0^{-1}(r))$, and $g'_3(r) = 1 + g'_2(r)F_0^{-1}(r)$. For the standardized AEPD, $F_0^{-1}(r) = [F_{AEP}^{-1}(r; \beta) - \omega(\beta)]/\delta(\beta)$,

$$g'_2(r) = \begin{cases} \delta(\beta)\left(\dfrac{1}{2\alpha^*}\right)^{p_1} \left|F_{AEP}^{-1}(r;\beta)\right|^{p_1-1}, & \text{if } r < \alpha, \\[2ex] -\delta(\beta)\left(\dfrac{1}{2(1-\alpha^*)}\right)^{p_2} \left|F_{AEP}^{-1}(r;\beta)\right|^{p_2-1}, & \text{if } r > \alpha. \end{cases}$$

For the standardized AST, $F_0^{-1}(r) = [F_{AST}^{-1}(r; \beta) - \omega(\beta)]/\delta(\beta)$,[9]

$$g'_2(r) = \begin{cases} -\delta(\beta)\left(\dfrac{v_1 + 1}{2\alpha^* v_1}\right) \dfrac{F_{AST}^{-1}(r;\beta)}{2\alpha^*} \left[1 + \dfrac{1}{v_1}\left(\dfrac{F_{AST}^{-1}(r;\beta)}{2\alpha^*}\right)^2\right]^{-1}, & \text{if } r < \alpha, \\[3ex] -\delta(\beta)\left(\dfrac{v_2 + 1}{2(1-\alpha^*)v_2}\right) \dfrac{F_{AST}^{-1}(r;\beta)}{2(1-\alpha^*)} \left[1 + \dfrac{1}{v_2}\left(\dfrac{F_{AST}^{-1}(r;\beta)}{2(1-\alpha^*)}\right)^2\right]^{-1}, & \text{if } r > \alpha. \end{cases}$$

When β is replaced with a consistent estimate $\widehat{\beta}$, for all C_k to be nonsingular, we use the statistic $T_n^* = [n/(n-l)]^{1/2}\sup_{1 \leq j \leq n-l}\left|\widehat{W}_n(v_j)\right|$ for a small number $l > 0$ (see Theorem 4 of Bai (2003)).

References

Aas, K., Haff, I.H., 2006. The generalized hyperbolic skew Student's t-distribution. J. Financ. Econom. 4 (2), 275–309.

Alberg, D., Shalit, H., Yosef, R., 2008. Estimating stock market volatility using asymmetric GARCH models. Appl. Financ. Econ. 18, 1201–1208.

Anderson, T., Darling, D., 1952. Asymptotic theory of certain "goodness of fit" criteria based on stochastic process. Ann. Math. Stat. 23, 193–212.

Artzner, P., Delbaen, F., Eber, J.M., Heath, D., 1997. Thinking coherently. Risk 10.

Artzner, P., Delbaen, F., Eber, J.M., Heath, D., 1999. Coherent measures of risk. Math. Financ. 9, 203–228.

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

Azzalini, A., Capitanio, A., 2003. Distributions generated by perturbation of symmetry with emphasis on a multivariate skew t distribution. J. R. Stat. Soc. B 65, 367–389.

Bai, J., 2003. Testing parametric conditional distributions of dynamic models. Rev. Econ. Stat. 85, 531–549.

Bauwens, L., Laurent, S., 2002. A new class of multivariate skew densities, with application to GARCH models. Comput. Econ. Financ. 5.

Bollerslev, T., 1986. Generalized autoregressive conditional hetero-skedasticity. J. Econom. 31, 307–327.

Bollerslev, T., 1987. A conditional heteroskedastic time series model for speculative prices and rates of return. Rev. Econ. Stat. 69, 542–547.

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

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

Branco, M.D., Dey, D.K., 2001. A general class of multivariate skew-elliptical distributions. J. Multivar. Anal. 79, 99–113.

Engle, R.F., 1982. Autoregressive conditional heteroskedasticity with estimates of the variance of U.K. inflation. Econometrica 50, 987–1008.

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

Fernandez, C., Steel, M.F.J., 1998. On Bayesian modeling of fat tails and skewness. J. Am. Stat. Assoc. 93, 359–371.

Fernandez, C., Osiewalski, J., Steel, M.F.J., 1995. Modeling and inference with $v-$distributions. J. Am. Stat. Assoc. 90, 1331–1340.

Hansen, B.E., 1994. Autoregressive conditional density estimation. Int. Econ. Rev. 35 (3), 705–730.

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

Jones, M.C., Faddy, M.J., 2003. A skew extension of the t distribution, with applications. J. R. Stat. Soc. B 65, 159–174.

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

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

Rombouts, J.V.K., Bouaddi, M., 2009. Mixed exponential power asymmetric conditional heteroskedasticity. Stud. Nonlinear Dynam. Econom. 13 (3) article 3.

Sahu, S.K., Dey, D.K., Branco, M.D., 2003. A new class of multivariate skew distributions with applications to Bayesian regression models. Can. J. Stat. 31, 129–150.

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

Stentoft, L., 2008. American option pricing using GARCH models and the Normal Inverse Gaussian distribution. J. Financ. Econom. 6, 540–582.

Taylor, J.W., 2008. Using exponentially weighted quantile regression to estimate value at risk and expected shortfall. J. Financ. Econom. 6, 382–406.

Theodossiou, P., 1998. Financial data and the skewed generalized t distribution. Manag. Sci. 44 (12–1), 1650–1661.

Zhu, D., Galbraith, J.W., 2010. A generalized asymmetric Student-t distribution with application to financial econometrics. J. Econom. 157, 297–305.

Zhu, D., Zinde-Walsh, V., 2009. Properties and estimation of asymmetric exponential power distribution. J. Econom. 148, 86–99.

---

[9] Here we need to evaluate only $g'(v_i)$ ($i = 1, 2, ..., n$). In fact, for the standardized AEPD or ASTD, $F_0^{-1}(v_i) = \widehat{z}_{(i)}$, $F_{AEP}^{-1}(v_i; \beta)$ or $F_{AST}^{-1}(v_i; \beta) = \omega(\beta) + \delta(\beta)\widehat{z}_{(i)}$ with the corresponding $\omega(\beta)$ and $\delta(\beta)$.