2009s-24

Forecasting Expected Shortfall with a Generalized Asymmetric Student-t Distribution

Dongming Zhu, John W. Galbraith

Série Scientifique
Scientific Series

Montréal
Mai 2009

CIRANO
Allier savoir et décision

Centre interuniversitaire de recherche en analyse des organisations

Forecasting Expected Shortfall with a Generalized Asymmetric Student-t Distribution[*]

Dongming Zhu[†], John W. Galbraith[‡]

Résumé / *Abstract*

De façon générale, les rendements financiers sont caractérisés par des queues épaisses et une certaine asymétrie. Ainsi, les modèles à variance conditionnelle dotés de ces caractéristiques donnent de meilleurs résultats que les modèles plus limités. La différence dans les résultats obtenus peut être particulièrement importante lorsqu'il s'agit d'évaluer des quantités qui dépendent des caractéristiques des queues, y compris les mesures du risque, tel que le manque à gagner prévu. Dans le cas actuel, en recourant à une généralisation récente de la distribution asymétrique suivant la loi *t* de Student, de sorte que des paramètres distincts limitent l'asymétrie et l'épaisseur de chaque queue, nous intégrons les rendements financiers quotidiens et estimons le manque à gagner prévu dans le cas de l'indice S&P 500 et de certaines actions de compagnies individuelles. La distribution généralisée est utilisée pour les innovations normalisées contenues dans un modèle asymétrique non linéaire de type GARCH. Les résultats démontrent de façon empirique l'utilité de la distribution généralisée pour améliorer les prévisions au sujet du risque de perte en cas de baisse du marché des actifs financiers.

Mots clés : distribution asymétrique, manque à gagner prévu, modèle NGARCH (**N**onlinear **G**eneralized **A**uto**R**egressive **C**onditional **H**eteroscedasticity)

Financial returns typically display heavy tails and some 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 a recent generalization of the asymmetric Student-t distribution to allow separate parameters to control skewness and the thickness of each tail, we fit daily financial returns and forecast expected shortfall for the S&P 500 index and a number of individual company stocks; the generalized distribution is used for the standardized innovations in a nonlinear, asymmetric GARCH-type model. The results provide empirical evidence for the usefulness of the generalized distribution in improving prediction of downside market risk of financial assets.

Keywords: *asymmetric distribution, expected shortfall, NGARCH model.*

Codes JEL : C16, G10

---

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 alone.

Subsequent literature has explored the use of skewed versions of the Student-t for the conditional distribution. 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. The availability of relatively long data sets has allowed the estimation and exploitation of these effects, and empirical investigation has found that such skew extensions have value in characterizing the pattern of financial returns; see for example Mittnik and Paolella (2003) and Alberg et al. (2008).

The distributions just mentioned use two parameters which together control skewness and thickness of the two tails. Zhu and Galbraith (2009) propose a further extension, in which two parameters control thickness of the two tails and a third allows skewness to change independently of the tail parameters. Because there is the potential that tail thickness differs non-negligibly between the left and right tails, this extension 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, this extension offers the potential for further 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 this distribution for conditional asset returns, and in particular the application to forecasting downside risk through the expected shortfall, a measure which is sensitive to losses in the extreme tail of the distribution of returns. The generalized asymmetric Student-t (AST) distribution proposed by Zhu and Galbraith (2009)

is to model the standardized innovations from the nonlinear asymmetric NGARCH model of Engle and Ng (1993). We find evidence of both improved fits to financial return data, and improved forecasts of the expected shortfall.

The next section of the paper gives the form of the AST distribution and summarizes results on associated risk measures. Section 3 describes the financial return models and the financial return data that will be used. Section 4 gives results on the core problem of predicting downside risk, in the form of the expected shortfall. A final section concludes.

2 The generalized asymmetric Student-t distribution

We begin with a brief review, from Zhu and Galbraith (2009), of results that characterize the distribution to be used below in modeling standardized innovations.

2.1 The AST density

The AST distribution is defined such that its standard probability density function (location and scale parameters are 0 and 1 respectively) is

$$
f_{AST}(y; \alpha, v_1, v_2) =
\begin{cases}
\frac{\alpha}{\alpha^*} K(v_1) \left[1 + \frac{1}{v_1} \left(\frac{y}{2\alpha^*} \right)^2 \right]^{-\frac{v_1+1}{2}}, & y \leq 0 \\[3mm]
\frac{1-\alpha}{1-\alpha^*} K(v_2) \left[1 + \frac{1}{v_2} \left(\frac{y}{2(1-\alpha^*)} \right)^2 \right]^{-\frac{v_2+1}{2}}, & y > 0
\end{cases}
\tag{1}
$$

where $\alpha \in (0,1)$ is the skewness parameter, $v_1 > 0$ and $v_2 > 0$ are the left and right tail parameters respectively, $K(v) \equiv \Gamma((v+1)/2)/[\sqrt{\pi v}\Gamma(v/2)]$, [1] and α^* is defined as $\alpha^* = \alpha K(v_1)/[\alpha K(v_1) + (1-\alpha)K(v_2)]$; as well,

$$
\frac{\alpha}{\alpha^*} K(v_1) = \frac{1-\alpha}{1-\alpha^*} K(v_2) = \alpha K(v_1) + (1-\alpha)K(v_2) \equiv B.
\tag{2}
$$

We can write a general form of the AST density, with location and scale parameters μ and σ, as $\frac{1}{\sigma} f_{AST}(\frac{y-\mu}{\sigma}; \alpha, v_1, v_2)$. As well, a re-scaled version is

---

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

useful for computational convenience:

$$
f_{AST}(y;\theta) = \begin{cases} \frac{1}{\sigma}\left[1 + \frac{1}{v_1}\left(\frac{y-\mu}{2\alpha\sigma K(v_1)}\right)^2\right]^{-(v_1+1)/2}, & y \le \mu; \\[2em] \frac{1}{\sigma}\left[1 + \frac{1}{v_2}\left(\frac{y-\mu}{2(1-\alpha)\sigma K(v_2)}\right)^2\right]^{-(v_2+1)/2}, & y > \mu, \end{cases} \tag{3}
$$

where $\theta = (\alpha, v_1, v_2, \mu, \sigma)^T$.

Figure 1 illustrates AST densities using (3) with various values for α, v_1, v_2; in each case the location parameter μ is 0 and the scale parameter σ is 1. In left-hand-side panels, v_1 is held constant at 2 while v_2 varies; in right-hand-side panels, $v_2 = 2$ while v_1 varies. The central panels illustrate the $\alpha = 0.5$ case; these densities are of course symmetric in the $v_1 = v_2$ cases which are among those depicted.

2.2 Downside risk measurement in the 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 sensitive to extreme negative returns. [2]

Zhu and Galbraith (2009) show that the expected shortfall for a standard AST random variable Y at a point in the support of the distribution q, i.e.

$$
ES_{AST}(q) \equiv E(Y \mid Y < q),
$$

can be expressed as $ES_{AST}(q) =$

$$
\frac{-4B}{F_{AST}(q)}\left\{\alpha^{*2}\left(\frac{v_1}{v_1 - 1}\right)\left[1 + \frac{1}{v_1}\left(\frac{q \wedge 0}{2\alpha^*}\right)^2\right]^{(1-v_1)/2} + (1 - \alpha^*)^2 \cdot\right.
$$

$$
\left.\cdot\left(\frac{v_2}{v_2 - 1}\right)\left(\left[1 + \frac{1}{v_2}\left(\frac{q \vee 0}{2(1 - \alpha^*)}\right)^2\right]^{(1-v_2)/2} - 1\right)\right\}, \tag{4}
$$

with $a \wedge b = \min\{a, b\}$ and $a \vee b = \max\{a, b\}$, and B is defined as in (2).

---

[2] An expression for Value at Risk in the AST is given in Zhu and Galbraith 2009, but is not used in the present study.

Now consider estimation of the j-step-ahead forecast of the conditional ES; $z_{t+j} = (r_{t+j} - m)/\sigma_{t+j}$ is assumed to be an AST random variable with zero mean and unit variance. Then based on the predicted values $\widehat{m}_{t+j|t}$, $\widehat{\sigma}_{t+j|t}$, $\widehat{\omega}_t = \omega(\widehat{\beta}_t)$ and $\widehat{\delta}_t = \delta(\widehat{\beta}_t)$, where $\omega(\cdot)$ and $\delta(\cdot)$ are defined in (10), we expect at time t that, conditional on the information available in period t,

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

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

$$ES_{t+j|t}(q) \equiv E_t(r_{t+j} \mid r_{t+j} < q) = \widehat{m}_{t+j|t} + \widehat{\sigma}_{t+j|t} \left[\frac{ES_{AST}(q_{t+j}; \widehat{\beta}_t) - \widehat{\omega}_t}{\widehat{\delta}_t} \right] \quad (5)$$

where

$$q_{t+j} = \widehat{\omega}_t + \widehat{\delta}_t \left[\frac{q - \widehat{m}_{t+j|t}}{\widehat{\sigma}_{t+j|t}} \right].$$

Note that, in the calculation of the conditional ES, we also need to give the expressions for $ES_{AST}(q; \beta)$. That is, the downside risk is determined not only by the specification of the conditional mean and NGARCH equations, but also by the distributional choice for the innovations.

3 Models and data

In this section, we examine forecasts of the expected shortfall with GARCH-class models and compare performance with error distributions given by the usual Student-t (ST), the SST (skewed Student-t in which $v_1 = v_2$ is imposed in the AST), the AST with α restricted to $1/2$, and finally the general AST. Empirical evidence has indicated that daily (for example) financial return data continue to exhibit conditional tail-fatness even after allowing for the GARCH effect (see Bollerslev et al., 1992).

GARCH-class models have been widely and successfully used to model financial asset returns. In general, a return process $r = \{r_t\}$ is modeled as[3]

$$r_t = m_t + \sigma_t z_t, \quad (6)$$

---

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

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

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

$$
\begin{aligned}
r_t &= m + \sigma_t z_t, \quad z_t \sim i.i.d.\,AST(0,1), \\
\sigma_t^2 &= b_0 + b_1 \sigma_{t-1}^2 + b_2 \sigma_{t-1}^2 (z_{t-1} - c)^2 \\
&= b_0 + b_1 \sigma_{t-1}^2 + b_2 (r_{t-1} - m - c\sigma_{t-1})^2.
\end{aligned}
\tag{7}
$$

The parameter c in the NGARCH equation (7) captures the leverage effect; that is, a positive value of c gives rise to a negative correlation between the innovations in the asset return and its conditional volatility. The j-step-ahead forecast of σ_{t+j}^2, denoted by $\sigma_{t+j|t}^2$, is defined as $\sigma_{t+j|t}^2 \equiv E_t\left(\sigma_{t+j}^2\right)$:

$$
\begin{aligned}
\sigma_{t+1|t}^2 &= b_0 + b_1 \sigma_t^2 + b_2 (r_t - m - c\sigma_t)^2, \tag{8} \\
\sigma_{t+j|t}^2 &= b_0 + \left[b_1 + b_2(1 + c^2)\right] \sigma_{t+j-1|t}^2, \ j \geq 2. \tag{9}
\end{aligned}
$$

We consider daily returns on the S&P500 composite index and several individual company stocks (Adobe Systems, Bank of America, JP Morgan, Johnson & Johnson, Merck, and Starbucks). The data sets end at December 31 2008; the beginning dates and numbers of observations are given in Table 5. [4]

3.1 Estimation and goodness of fit

The maximum likelihood estimate of the parameter vector ϕ, where $\phi = (m, b_0, b_1, b_2, c, \alpha, \upsilon_1, \upsilon_2)$, is obtained by maximizing the log-likelihood function

$$
l_T(\phi; r) = \sum_{t=1}^{T} \left\{ \log \delta - \log \sigma_t + \log f_Y(\omega + \delta \frac{r_t - m}{\sigma_t}; \beta) \right\}, \tag{10}
$$

---

[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 level of the index at time t.

where $f_Y(\cdot; \beta)$ is the standard AST density function with the distributional parameters $\beta = (\alpha, v_1, v_2)'$, $\omega \equiv \omega(\beta)$ and $\delta \equiv \delta(\beta)$ denote the mean and standard deviation of $f_Y(\cdot; \beta)$ respectively; they are given by

$$
\begin{aligned}
E(Y) &= 4\left[-\alpha\alpha^*\frac{v_1 K(v_1)}{v_1 - 1} + (1-\alpha)(1-\alpha^*)\frac{v_2 K(v_2)}{v_2 - 1}\right] \\
&= 4B\left[-\alpha^{*2}\frac{v_1}{v_1 - 1} + (1-\alpha^*)^2\frac{v_2}{v_2 - 1}\right],
\end{aligned} \tag{11}
$$

$$
\begin{aligned}
Var(Y) &= 4\left[\alpha\alpha^{*2}\frac{v_1}{v_1 - 2} + (1-\alpha)(1-\alpha^*)^2\frac{v_2}{v_2 - 2}\right] \\
&\quad - 16B^2\left[-\alpha^{*2}\frac{v_1}{v_1 - 1} + (1-\alpha^*)^2\frac{v_2}{v_2 - 1}\right]^2,
\end{aligned} \tag{12}
$$

where $K(\cdot)$ and B are defined in (1) and (2). [5]

To examine the significance of asymmetric behavior in the tails, we consider the AST and the nested distribution classes mentioned above: the AST with $\alpha = 1/2$ to represent asymmetry arising only from different tail behavior, the SST (i.e. AST with $v_1 = v_2$), and the ST (i.e. AST with $\alpha = 1/2$ and $v_1 = v_2$). The ML estimates of the parameters and their standard deviations for the S&P 500 index data are displayed in Table 1; for individual company stocks we report below a more limited set of results on fit and forecast performance.

Table 1: Parameter estimates for AST-NGARCH(1,1) models

	m	b_0	b_1	b_2	c	α	v_1	v_2
AST	.0214	.0096	.8763	.0592	1.011	.4989	6.36	15.98
	(.0115)	(.0019)	(.0127)	(.0061)	(.1016)	(.0169)	(.9154)	(5.622)
AST, α=0.5	.0212	.0096	.8762	.0592	1.011		6.40	15.73
	(.0113)	(.0021)	(.0125)	(.0065)	(.0957)		(.6765)	(3.899)
SST	.0220	.0100	.8729	.0599	1.023	.532		8.50
	(.0109)	(.0021)	(.0136)	(.0061)	(.0968)	(.0094)		(.9329)
ST	.0318	.0092	.8755	.0589	1.015			8.27
	(.0108)	(.0020)	(.0132)	(.0064)	(.1002)			(.9262)

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

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

$$AICC = -2L + \frac{2T(k+1)}{T-k-2}, \quad SBC = -2L + \frac{k\log(T)}{T}; \qquad (13)$$

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

$$AD = \sup_{-\infty < x < +\infty} \sqrt{T} \frac{\left| F_T(x) - \widehat{F}(x) \right|}{\sqrt{\widehat{F}(x)(1 - \widehat{F}(x))}}, \qquad (14)$$

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

The AD statistic is a reasonable measure of the discrepancy or "distance" between the empirical cdf $F_T(x)$ and the hypothetical distribution $F(x)$; this statistic gives appropriate weight to the tails of the distribution so that it can be used to measure goodness of fit in the tails. In our applications, since the innovations are assumed to have zero mean and unit variance, the estimated cdf of the innovations, $\widehat{F}(x)$ in (14), can be expressed as $\widehat{F}(x) = F_Y(\widehat{\omega} + \widehat{\delta}x; \widehat{\beta})$, where $F_Y(\cdot; \beta)$ is the cdf of the standard AST with $\beta = (\alpha, \upsilon_1, \upsilon_2)^T$, $\widehat{\beta}$ is the ML estimate of β, and $\widehat{\omega}$ and $\widehat{\delta}$ are given by $\widehat{\omega} = \omega(\widehat{\beta})$ and $\widehat{\delta} = \delta(\widehat{\beta})$, which are, respectively, the estimated mean and standard deviation of $F_Y(\cdot; \beta)$. For simplicity we compute the AD statistic as follows:

$$AD = \max_j AD_j, \quad AD_j = \sqrt{T} \frac{\left| F_T(\widehat{z}_{j,T}) - \widehat{F}(\widehat{z}_{j,T}) \right|}{\sqrt{\widehat{F}(\widehat{z}_{j,T})(1 - \widehat{F}(\widehat{z}_{j,T}))}}, \qquad (15)$$

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

8

Table 2: Goodness-of-fit measures for the AST-NGARCH(1,1) models

	L	AICC	SBC	AD
AST	-6208.6	12435	12417	2.68
AST, $\alpha = 0.5$	-6208.6	12433	12417	2.71
SST	-6212.2	12440	12424	4.97
ST	-6217.2	12448	12434	5.95

Table 2 displays the four measures of goodness-of-fit for the estimated AST-NGARCH(1,1) models on S&P 500 index data; for individual companies, Table 4 below reports the model ranked best by each measure. For the S&P 500, the AD statistic ranks the distribution with full asymmetry as the best; by SBIC full asymmetry and the restriction to $\alpha = 1/2$ are equivalent to five significant digits; the AICC selects the restricted version. No criterion selects the SST or ST. The LR test, $LR = 2(L^u - L^r)$, where L^u and L^r are respectively unrestricted and restricted log-likelihood values, does not reject $\alpha = 1/2$, but does reject the restriction to one tail parameter rather than two.

For individual company data, the general AST again tends to be selected, but there are two cases in which the AICC is the same (to four significant digits) for general AST and SST, and there are two cases in which SST is selected by one of the other criteria. While $\alpha = 1/2$ is not rejected on the S&P 500 data, the restriction is typically rejected in the individual company data (for brevity we do not report all results for individual company data, but we give a summary in Table 4).

4　Prediction performance for downside risk

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}, $\widehat{F}_{t+j|t}(r_{t+j})$. We can express the expected shortfall measure of predictive downside risk as follows: conditional on a return lower than q after j periods, the expected return will be $ES_{t+j|t}(q)$. Given models specified in (6) and (7), 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^2_{t+j}) of r_{t+j}. Denote the time-t ML estimates of these parameters by $(\widehat{m}_t, \widehat{b}_{0t}, \widehat{b}_{1t}, \widehat{b}_{2t}, \widehat{c}_t, \widehat{\beta}'_t)$ and the estimates of the j-step-ahead forecasts of m_{t+j} and σ^2_{t+j} by $\widehat{m}_{t+j|t}$ and $\widehat{\sigma}^2_{t+j|t}$, respectively. Then, $\widehat{m}_{t+j|t} = \widehat{m}_t$ for any j, and $\widehat{\sigma}^2_{t+j|t}$ is obtained by substituting the estimated parameters into (8) and (9),

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

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

To check predictive performance out-of-sample, we split the sample of S&P index data, retaining in this case $N = 2000$ points for out-of-sample evaluation.[6] We then recursively evaluate $ES_{t+j|t}(q)$, $N = 2000 \leq t \leq T - j$, for one and five steps ahead: $j = 1, 5$. We set the threshold (loss) returns $q = -1.2\%, -1\%, -0.8\%, -0.6\%$. 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 $ES_{t+j|t}(q)$ predicted by the model. If the observed expected shortfall

$$\widehat{\mathbf{ES}}_j(q) \equiv \frac{1}{J}\sum_{t=N}^{T-j} r_{t+j}\mathbf{1}\{r_{t+j} < q\}, \text{ where } J = \sum_{t=N}^{T-j}\mathbf{1}\{r_{t+j} < q\} \quad (16)$$

is lower (higher) than the average predictive ES,

$$\widehat{ES}_j^M(q) \equiv \frac{1}{J}\sum_{t=N}^{T-j} ES_{t+j|t}(q)\mathbf{1}\{r_{t+j} < q\}, \quad (17)$$

then the model tends to underestimate (overestimate) the risk. This is measured by the mean error, $ME_j(q) \equiv \widehat{ES}_j^M(q) - \widehat{\mathbf{ES}}_j(q)$. Another important measure of predictive out-of-sample performance is the mean absolute error

$$MAE_j(q) \equiv \frac{1}{J}\sum_{t=N}^{T-j} \left| ES_{t+j|t}(q) - \widehat{\mathbf{ES}}_j(q)\right|\mathbf{1}\{r_{t+j} < q\}. \quad (18)$$

---

[6]For individual company data, we use half of the sample for out-of-sample evaluation to a maximum of 3000 data points, so that N is the minimum of $T/2$ and 3000.

Table 3 shows the predictive performance for the expected shortfall risk on the S&P 500 data; the entries in the table are the mean errors $100ME_j(q)$ and the mean absolute errors $100MAE_j(q)$ of the expected shortfall predictions for one and five steps ahead.

From the mean errors ME_j, we see that the AST-type models tend to overestimate the risk (have a negative mean error) for larger loss thresholds ($q = -1.2\%, -1\%$), but show mean errors close to zero for the smaller thresholds ($q = -0.8\%, -0.6\%$). The SST- and ST- type models tend to overestimate risk for larger threshold losses, and to underestimate for small thresholds. By the mean absolute errors MAE_j, the AST models are uniformly and significantly better than the SST and ST models; the full AST specification is the best choice in each case.

Table 3: Predictive performance for expected shortfall risk[7]

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	-.0869, **.3673**	-.0466,**.3785**	-.0233,**.3675**	.0069, **.3595**
AST, α=0.5	-.0835, **.3766**	-.0476,**.3851**	-.0279,**.3720**	-.0006,**.3612**
SST	-.0464, **.3949**	-.0177,**.4006**	-.0041,**.3823**	.0178, **.3690**
ST	-.0150, **.3933**	.0119, **.3997**	.0229, **.3805**	.0422, **.3671**
$j = 5$	ME_5,MAE_5	ME_5,MAE_5	ME_5,MAE_5	ME_5,MAE_5
AST	-.0798, **.3623**	-.0394,**.3696**	-.0215,**.3564**	.0047, **.3494**
AST, α=0.5	-.0760, **.3717**	-.0402,**.3761**	-.0262,**.3604**	-.0030,**.3511**
SST	-.0376, **.3905**	-.0091,**.3913**	.-0014,**.3707**	.0163, **.3584**
ST	-.0056, **.3896**	.0211,**.3912**	.0264,**.3688**	.0416, **.3561**

For each of the six individual companies, the best-performing model in 1-step and 5-step forecasts is recorded in Table 4 (the results are shown as the

---

[7]Note: 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 (18), multiplied by 100, for the threshold losses (negative returns) $q = -1.2\%, -1\%, -0.8\%, -0.6\%$ and $j = 1, 5$.

1% threshold, but give the same best-performing model at each of the other thresholds examined as well). The general AST gives the best performance in four of six cases, whereas the plain Student-t performs best in two of six.

Thus both by measures of fit and by forecasting performance for this important measure of downside risk, the AST class of distributions appears to offer the potential to provide useful improvements in model performance.

5   Concluding remarks

Useful measures of downside risk must reflect the potential for extreme outcomes, but measuring and forecasting such risk for financial assets is 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 attendant improvements in forecasting power for downside risk measures.

The present paper has attempted further progress on this research program. Using an asymmetric Student-t distribution 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 obtain better estimates of the thickness of the left tail, with attendant potential improvements in forecasting power for risk of loss, measured here by the expected shortfall.

This potential seems to be realizable on available samples of financial return data. Nonlinear, asymmetric GARCH models in which standardized innovations are modeled as arising from the generalized asymmetric Student-t distribution often, though not always, show improvements in both fit and forecasting power relative to those with more restricted specifications of the distribution of standardized innovations. Whether these improvements would also be visible in asset return series where asymmetry seems to be less important, such as some exchange rate series, remains to be seen.

References

[1] Aas, K. and Haff, I.H. (2006). The generalized hyperbolic skew Student's t-distribution. *Journal of Financial Econometrics*, 4(2), 275-309.

[2] Alberg, D., H. Shalit and R. Yosef (2008). Estimating stock market volatility using asymmetric GARCH models. *Applied Financial Economics* 18, 1201-1208.

[3] Andersen, T.G., T. Bollerslev, P. F. Christoffersen and F. X. Diebold (2005). Volatility and correlation forecasting, in: *Handbook of Economic Forecasting,* edited by Graham Elliott, Clive W.J. Granger, and Allan Timmermann. Amsterdam: North Holland.

[4] Anderson, T. and Darling, D. (1952). Asymptotic theory of certain "goodness of fit" criteria based on stochastic process. *The Annals of Mathematical Statistics* 23, 193-212.

[5] Azzalini, A. and A. Capitanio (2003). Distributions generated by perturbation of symmetry with emphasis on a multivariate skew t distribution. *Journal of the Royal Statistical Society B* 65, 367-389.

[6] Bauwens, L. and Laurent, S. (2002). A new class of multivariate skew densities, with application to GARCH models. *Journal of Business and Economic Statistics.*

[7] Bollerslev, T. (1986) Generalized autoregressive conditional heteroskedasticity. *Journal of Econometrics* 31, 307-327.

[8] Bollerslev, T. (1987). A conditional heteroskedastic time series model for speculative prices and rates of return. *Review of Economics and Statistics* 69, 542-547.

[9] Bollerslev, T., R.Y. Chou and K.F. Kroner (1992). ARCH modeling in finance: A selective review of the theory and empirical evidence. *Journal of Econometrics* 52, 5-59.

[10] Branco, M.D. and Dey, D.K. (2001). A general class of multivariate skew-elliptical distributions. *Journal of Multivariate Analysis* 79, 99-113.

[11] Engle, R.F. (1982) Autoregressive conditional heteroskedasticity with estimates of the variance of U.K. inflation. *Econometrica* 50, 987-1008.

[12] Engle, R.F. and V.K. Ng (1993). Measuring and testing the impact of news on volatility. *Journal of Finance* 48, 1749-1778.

[13] Fernandez, C. and Steel, M.F.J. (1998). On Bayesian modeling of fat tails and skewness, *Journal of the American Statistical Association*, 93, 359-371.

[14] Hansen, B. E. (1994). Autoregressive conditional density estimation. *International Economic Review,* 35 (3), 705-730.

[15] Hurvich, C. and Tsai, C. (1989). Regression and time series model selection in small samples. *Biometrika* 76, 297-307.

[16] Jones, M.C. and Faddy, M.J. (2003). A skew extension of the t distribution, with applications. *Journal of the Royal Statistical Society, Series B*, 65, 159-174.

[17] Mittnik, S. and Paolella, M. S. (2003). Prediction of financial downside-risk with heavy-tailed conditional distributions. *Handbook of Heavy Tailed Distributions in Finance*, edited by S. T. Rachev.

[18] Sahu, S.K., D.K. Dey and M.D. Branco (2003). A new class of multivariate skew distributions with applications to Bayesian regression models. *The Canadian Journal of Statistics* 31, 129-150.

[19] Theodossiou, P. (1998). Financial data and the skewed generalized t Distribution. *Management Science*, 44 (12-1), 1650-1661.

[20] Zhu, D. and J.W. Galbraith (2009) A generalized asymmetric Student-t distribution with application to financial econometrics. Under revision.

Table 4: Preferred model by various criteria

Data series	AICC	SBC	AD	$MAE_{-1\%,j=1}$	$MAE_{-1\%,j=5}$
S&P 500 composite	B	A	A	A	A
Adobe Systems	A	A	A	A	A
Bank of America	A	A	A	D	D
Johnson&Johnson	A/C	A	C	A	A
JP Morgan	A/C	A	A	A	A
Merck	A	A	A	D	D
Starbucks	A	C	A	A	A

Model labels:

A: General AST errors
B: AST errors with $\alpha = 0.5$
C: skewed Student-t (SST) errors
D: standard Student-t (ST) errors

Table 5: List of data samples

Company or index	beginning date	end date	no. obs.
S&P 500 composite	1990.01.02	2008.12.31	4791
Adobe Systems	1986.08.14	2008.12.31	5646
Bank of America	1982.03.19	2008.12.31	6760
Johnson&Johnson	1961.01.03	2008.12.31	12082
JP Morgan	1969.03.06	2008.12.31	10052
Merck	1949.05.04	2008.12.31	15950
Starbucks	1992.06.29	2008.12.31	4161

Figure 1
AST densities for various parameter values
$\mu = 0,\ \sigma = 1$

$\alpha = 0.3,\ \upsilon_1 = 2$

$\alpha = 0.3,\ \upsilon_2 = 2$

$\alpha = 0.5,\ \upsilon_1 = 2$

$\alpha = 0.5,\ \upsilon_2 = 2$

$\alpha = 0.7,\ \upsilon_1 = 2$

$\alpha = 0.7,\ \upsilon_2 = 2$