A Generalized Asymmetric Student-t Distribution with Application to Financial Econometrics

Dongming Zhu
HSBC School of Business, Peking University

John W. Galbraith*
Department of Economics, McGill University

April 14, 2009

Abstract

This paper proposes a new class of asymmetric Student-t (AST) distributions, and investigates its properties, gives procedures for estimation, and indicates applications in financial econometrics. We derive analytical expressions for the cdf, quantile function, moments, and quantities useful in financial econometric applications such as the expected shortfall. A stochastic representation of the distribution is also given. Although the AST density does not satisfy the usual regularity conditions for maximum likelihood estimation, we establish consistency, asymptotic normality and efficiency of ML estimators and derive an explicit analytical expression for the asymptotic covariance matrix. A Monte Carlo study indicates generally good finite-sample conformity with these asymptotic properties.

JEL classification codes: C13, C16

Key words: asymmetric distribution, expected shortfall, maximum likelihood estimation.

1

1 Introduction

The Student-t distribution is commonly used in finance and risk management, particularly to model conditional asset returns for which the tails of the normal distribution are almost invariably found to be too thin. For example, Bollerslev (1987) used the Student-t to model the distribution of foreign exchange rate returns; Mittnik, Rachev and Paolella (1998) fitted a return distribution using a number of parametric distributions including Student-t, and found that the partially asymmetric Weibull, Student-t and the asymmetric stable distributions provide the best fit according to various measures. Recent applications include Alberg et al. (2008) and Franses et al. (2008).

Hansen (1994) was the first to consider a skewed Student's t distribution to model skewness in conditional distributions of financial returns. Since then, several skew extensions of the Student-t distribution have been proposed for financial and other applications; see for example 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.

All but two of these skew t-type distributions have two tails with identical polynomial rate of decay. The first of the exceptions is the skew extension of Jones and Faddy (2003), which has two tail parameters to control the left and right tail behavior, respectively, but does not embody a third that allows skewness to change independently of the tail parameters. The second is due to Aas and Haff (2006), who argued for a special case of the generalized hyperbolic (GH) distribution, called the GH Student-t distribution, in which one tail is determined by a polynomial rate, while the other has exponential behavior. For detailed descriptions of various skew Student-t type distributions, refer to the review in Aas and Haff (2006). However, in general, a skewness parameter mainly controls the asymmetry of the central part of a distribution. Therefore a class of generalized asymmetric Student-t (AST) distributions which has one skewness parameter and two tail parameters offers the potential to improve our ability to fit and forecast empirical data in the tail regions which are critical to risk management and other financial econometric applications. In this paper, we propose such a class of distributions, describe estimation methods and investigate properties of the distribution and of the estimators.

There are various methodologies for generation of a skewed Student-t distribution. One is the two-piece method; Hansen (1994) used this method to propose the first skew extension to the Student-t. More generally, Fer-

nandez and Steel (1998) introduced a skewness parameter γ to any univariate pdf which is unimodal and symmetric, resulting in a skewed version of the Student-t equivalent to that of Hansen (1994); Bauwens and Laurent (2002) generalized the procedure used in Fernandez and Steel (1998) to the multivariate case. A second methodology is the perturbation approach of Azzalini and Capitanio (2003), which can generate the multivariate skew elliptical distributions proposed by Branco and Dey (2001) and Sahu et al. (2003) using the conditioning method. [1] In this paper we will extend the two-piece method to allow the additional parameter.

Allowing an additional parameter offers the potential to fit more subtle features of the distribution than is possible with two-parameter versions, with the attendant potential for better descriptions of tail phenomena, and better predictions of quantitites such as expected shortfall which depend on the shape of the tail. Of course, relatively large sample sizes may be necessary in order to realize this potential: even if a three-parameter form provides in principle a better description of a given type of data, the two-parameter approximation may not be detectably poorer in a finite sample. However, we show here by simulation that the parameters can be distinguished in realistic sample sizes, and in a companion empirical study (Zhu and Galbraith 2009) we show that improved fit and forecast performance can be observed in financial return data.

The paper is organized as follows. Section 2 gives the definition of the AST distribution and section 3 provides an interpretation of parameters and gives some properties such as a stochastic representation and analytical expressions for the cdf, quantiles, moments, value at risk and expected shortfall. In section 4 we establish consistency and asymptotic normality of the MLE, and section 5 provides some finite-sample Monte Carlo results. Technical results and proofs are collected in the appendices.

2   Definition of the AST Distribution

The asymmetric Student-t (AST) distribution proposed in this paper is defined as follows. Its standard (location parameter is zero, scale parameter is

---

[1] By using the conditioning method, Branco and Dey (2001) and Sahu et al. (2003) construct two different classes of multivariate skew Student t distributions, which however coincide in the univariate case.

unity) probability density function has the form

$$
f_{AST}(y; \alpha, \upsilon_1, \upsilon_2) =
\begin{cases}
\frac{\alpha}{\alpha^*} K(\upsilon_1) \left[1 + \frac{1}{\upsilon_1}(\frac{y}{2\alpha^*})^2 \right]^{-\frac{\upsilon_1+1}{2}}, & y \le 0 \\
\\
\frac{1-\alpha}{1-\alpha^*} K(\upsilon_2) \left[1 + \frac{1}{\upsilon_2}(\frac{y}{2(1-\alpha^*)})^2 \right]^{-\frac{\upsilon_2+1}{2}}, & y > 0
\end{cases}
\tag{1}
$$

where $\alpha \in (0,1)$ is the skewness parameter, $\upsilon_1 > 0$ and $\upsilon_2 > 0$ are the left and right tail parameters respectively, $K(\upsilon) \equiv \Gamma((\upsilon+1)/2)/[\sqrt{\pi \upsilon}\Gamma(\upsilon/2)]$ (where $\Gamma(\cdot)$ is the gamma function), and α^* is defined as

$$
\alpha^* = \alpha K(\upsilon_1)/[\alpha K(\upsilon_1) + (1-\alpha)K(\upsilon_2)].
\tag{2}
$$

Denoting by μ and σ the location (center) and scale parameters, respectively, the general form of the AST density is expressed as $\frac{1}{\sigma}f_{AST}(\frac{y-\mu}{\sigma}; \alpha, \upsilon_1, \upsilon_2)$. Note that

$$
\frac{\alpha}{\alpha^*}K(\upsilon_1) = \frac{1-\alpha}{1-\alpha^*}K(\upsilon_2) = \alpha K(\upsilon_1) + (1-\alpha)K(\upsilon_2) \equiv B.
\tag{3}
$$

The AST density (1) is continuous and unimodal with mode at the center, $y = \mu = 0$, and is everywhere differentiable at least once. In the limit as α approaches either 0 or 1, the shape of the density resembles a Student-t truncated at the mode. The parameter α^* provides scale adjustments respectively to the left and right parts of the density so as to ensure continuity of the density under changes of shape parameters $(\alpha, \upsilon_1, \upsilon_2)$.

A new parameterization of a skewed Student-t (SST) distribution is given by letting $\upsilon_1 = \upsilon_2 = \upsilon$ (implying $\alpha^* = \alpha$) in the AST; the general form of its density is

$$
f_{SST}(y; \alpha, \upsilon, \mu, \sigma) =
\begin{cases}
\frac{1}{\sigma} K(\upsilon) \left[1 + \frac{1}{\upsilon}(\frac{y-\mu}{2\alpha\sigma})^2 \right]^{-(\upsilon+1)/2}, & y \le \mu; \\
\\
\frac{1}{\sigma} K(\upsilon) \left[1 + \frac{1}{\upsilon}(\frac{y-\mu}{2(1-\alpha)\sigma})^2 \right]^{-(\upsilon+1)/2}, & y > \mu.
\end{cases}
\tag{4}
$$

This parameterization of the SST is equivalent to those of Hansen (1994) and Fernandez and Steel (1998), but it will provide an interesting new interpretation of the skewness parameter in terms of L_p distances. By reparameterization with $\alpha = 1/(1+\gamma^2)$ and $\sigma = (\gamma+1/\gamma)\sigma'/2$, the SST (4) will become that of Fernandez and Steel (1998); letting $\alpha = (1-\lambda)/2$, $\sigma = \frac{1}{b}\sqrt{(\upsilon-2)/\upsilon}$ and $\mu = -a/b$, the density (4) will be that of Hansen (1994, eq. 10). With

$\alpha = 1/2$, the SST reduces to the general form of Student-t distribution. The skewed Cauchy and skewed normal distribution are special cases of the SST with $\upsilon = 1$ and $\upsilon = \infty$, respectively. By the skewness measure of Arnold and Groeneveld (1995), the SST density is skewed to the right for $\alpha < 1/2$ and to the left for $\alpha > 1/2$.

When one of the tail parameters goes to infinity, say, $\upsilon_2 \to \infty$, the AST behaves as a Student-t on the left side and as a Gaussian on the right side, implying one heavy tail and one exponential tail. This type of tail behavior is similar to that of the GH Student-t in Aas and Haff (2006). With these two tail parameters the AST can accommodate empirical distributions of daily returns of financial assets that are often skewed and have one heavy tail and one relatively thin tail. A potential disadvantage of the AST, compared with the GH Student's t and that of Jones and Faddy (2003), is the fact that the density function is differentiable only once at the mode μ; however, it is not an impediment in applications, because we can show that the usual $\sqrt{T}$ asymptotics of MLE still hold for the AST.

The definition in (1) above is useful in theoretical analysis, but it will sometimes be convenient to re-scale for computations and applications. We can give an alternative definition of the AST density as follows:

$$
f_{AST}(y;\theta) = \begin{cases} \frac{1}{\sigma}\left[1 + \frac{1}{\upsilon_1}\left(\frac{y-\mu}{2\alpha\sigma K(\upsilon_1)}\right)^2\right]^{-(\upsilon_1+1)/2}, & y \leq \mu; \\ \\ \frac{1}{\sigma}\left[1 + \frac{1}{\upsilon_2}\left(\frac{y-\mu}{2(1-\alpha)\sigma K(\upsilon_2)}\right)^2\right]^{-(\upsilon_2+1)/2}, & y > \mu, \end{cases}
\tag{5}
$$

where $\theta = (\alpha, \upsilon_1, \upsilon_2, \mu, \sigma)^T$ and μ and σ are the location and scale parameters respectively. From the rescaled AST density (5), we can clearly observe the effects of the shape parameters on the distribution. This also yields a simple closed-form expression for the information matrix of the maximum likelihood estimator (MLE).

3   Properties of the AST distribution

3.1   Stochastic representation, moments, and implications of parameters

Suppose that Y is a random variable with the standard AST density ($\mu = 0$, $\sigma = 1$). Define $a \wedge b \equiv \min\{a,b\}$, $a \vee b \equiv \max\{a,b\}$, by $F_t(\cdot;\upsilon)$ the cdf of standard Student-t with non-integer degrees of freedom υ, and by $F_t^{-1}(\cdot;\upsilon)$

the inverse function of $F_t(\cdot; \upsilon)$. The cdf and quantile function of the AST r.v. Y are given by a straightforward calculation as follows:

$$F_{AST}(y) = 2\alpha F_t\left(\frac{y \wedge 0}{2\alpha^*}; \upsilon_1\right) + 2(1-\alpha)\left[F_t\left(\frac{y \vee 0}{2(1-\alpha^*)}; \upsilon_2\right) - \frac{1}{2}\right] \quad (6)$$

and

$$F_{AST}^{-1}(p) = 2\alpha^* F_t^{-1}\left(\frac{p \wedge \alpha}{2\alpha}; \upsilon_1\right) + 2(1-\alpha^*)F_t^{-1}\left(\frac{p \vee \alpha + 1 - 2\alpha}{2(1-\alpha)}; \upsilon_2\right), \quad (7)$$

where α^* is defined in (2). Note that $F_t(0; \upsilon) = \frac{1}{2}$, which implies $F_{AST}(0) = \alpha$ and $F_{AST}^{-1}(\alpha) = 0$ (recall that α is the skewness parameter here). This means that the α-quantile of a standard AST r.v. is always zero. For a general AST with location μ and scale σ, the location μ corresponds to the α-quantile of the general AST random variable. This is the basic interpretation of the parameters α and μ.

A stochastic representation of the AST is useful in studying properties of the distribution, and in simulation studies. Denote by $T(\upsilon)$ a random variable having the standard Student-t distribution with υ degrees of freedom. Consider three independent random variables U, $T(\upsilon_1)$ and $T(\upsilon_2)$, where $U \sim U(0,1)$, the uniform distribution on $[0,1]$. Define

$$Y = \alpha^* |T(\upsilon_1)| \left[sign(U - \alpha) - 1\right] + (1-\alpha^*) |T(\upsilon_2)| \left[sign(U - \alpha) + 1\right], \quad (8)$$

where α^* is defined as in (2), $sign(x) = +1$ if $x > 0$, -1 if $x < 0$, and 0 if $x = 0$. Then it is easy to show that Y is a standard AST random variable that has the density (1) with the three shape parameters $(\alpha, \upsilon_1, \upsilon_2)$. The following moment properties of Y can be obtained using this stochastic representation.

Note that the absolute moment of $T(\upsilon)$, $E|T(\upsilon)|^r$, can be expressed as

$$E|T(\upsilon)|^r = \sqrt{\frac{\upsilon^r}{\pi}}\Gamma\left(\frac{r+1}{2}\right)\Gamma\left(\frac{\upsilon-r}{2}\right)/\Gamma\left(\frac{\upsilon}{2}\right), \quad -1 < r < \upsilon. \quad (9)$$

This formula can be found in Mittnik and Paolella (2003), or it can be derived from a straightforward calculation by using eq. II-10 in Farrell and Ross (1963, p55). We then obtain

$$E\left(|Y|^r \mid Y \leq 0\right) = [2\alpha^*]^r E|T(\upsilon_1)|^r, \quad r \in (-1, \upsilon_1) \quad (10)$$

$$E\left(|Y|^r \mid Y > 0\right) = [2(1-\alpha^*)]^r E|T(\upsilon_2)|^r, \quad r \in (-1, \upsilon_2) \quad (11)$$

6

and

$$\begin{aligned}
E\,|Y|^r &= \alpha E\left(|Y|^r \mid Y \le 0\right) + (1-\alpha)E\left(|Y|^r \mid Y > 0\right) \\
&= \alpha[2\alpha^*]^r E\,|T(\upsilon_1)|^r + (1-\alpha)[2(1-\alpha^*)]^r E\,|T(\upsilon_2)|^r, \quad (12)
\end{aligned}$$

where $r \in (-1, \upsilon_1 \wedge \upsilon_2)$, implying that the r-th absolute moment of the standard AST r.v. Y can be obtained by combining (9) and (12). Similarly, for any positive integer $k < \upsilon_1 \wedge \upsilon_2$, the k-th moment is given by

$$E(Y^k) = \alpha[-2\alpha^*]^k E\,|T(\upsilon_1)|^k + (1-\alpha)[2(1-\alpha^*)]^k E\,|T(\upsilon_2)|^k. \quad (13)$$

In particular, the mean and variance of a standard AST random variable are:

$$\begin{aligned}
E(Y) &= 4\left[-\alpha\alpha^* \frac{\upsilon_1 K(\upsilon_1)}{\upsilon_1 - 1} + (1-\alpha)(1-\alpha^*)\frac{\upsilon_2 K(\upsilon_2)}{\upsilon_2 - 1}\right] \\
&= 4B\left[-\alpha^{*2}\frac{\upsilon_1}{\upsilon_1 - 1} + (1-\alpha^*)^2\frac{\upsilon_2}{\upsilon_2 - 1}\right], \quad (14)
\end{aligned}$$

$$\begin{aligned}
Var(Y) &= 4\left[\alpha\alpha^{*2}\frac{\upsilon_1}{\upsilon_1 - 2} + (1-\alpha)(1-\alpha^*)^2\frac{\upsilon_2}{\upsilon_2 - 2}\right] \\
&\quad -16B^2\left[-\alpha^{*2}\frac{\upsilon_1}{\upsilon_1 - 1} + (1-\alpha^*)^2\frac{\upsilon_2}{\upsilon_2 - 1}\right]^2, \quad (15)
\end{aligned}$$

where $K(\cdot)$ and B are defined respectively in (1) and (3). We see that all the moments can be expressed simply and conveniently in terms of the Gamma function. For the skew Student-t, where $\upsilon_1 = \upsilon_2 = \upsilon$ and $\alpha^* = \alpha$, we can obtain simplified expressions for various moments:

$$E(Y^k) = (2\sqrt{\upsilon})^k \left[(-1)^k \alpha^{k+1} + (1-\alpha)^{k+1}\right] \frac{\Gamma(\frac{k+1}{2})\Gamma(\frac{\upsilon-k}{2})}{\sqrt{\pi}\Gamma(\frac{\upsilon}{2})}, \quad (16)$$

$$E(|Y|^r) = (2\sqrt{\upsilon})^r \left[\alpha^{r+1} + (1-\alpha)^{r+1}\right] \frac{\Gamma(\frac{r+1}{2})\Gamma(\frac{\upsilon-r}{2})}{\sqrt{\pi}\Gamma(\frac{\upsilon}{2})}, \quad (17)$$

where k is a non-negative integer less than υ, and $-1 < r < \upsilon$.

An interpretation of the parameters can be given by using the conditional L_r−norm deviations,

$$d_L(r) \equiv [E(|Y - \mu|^r \mid Y \le \mu)]^{1/r}, \quad d_R(r) \equiv [E(|Y - \mu|^r \mid Y > \mu)]^{1/r}, \quad (18)$$

7

where $r > 0$ is any given constant, μ is the location parameter and here $\mu = 0$ for the standard AST r.v. Y. Substituting (10) and (11) into (18) yields

$$d_L(r) = 2\alpha^* \left(E \left| T(\upsilon_1) \right|^r \right)^{1/r}, \quad d_R(r) = 2(1 - \alpha^*) \left(E \left| T(\upsilon_2) \right|^r \right)^{1/r}. \quad (19)$$

As we know from the alternative definition (5), the parameters υ_1 and υ_2 separately control the shapes of the left and right sides of the AST, so they can be referred to as the left and right shape parameters respectively. We can see this point also from the left and right "conditional generalized kurtosis" defined for every $r > 0$ as

$$kur_L(r) \equiv [d_L(2r)/d_L(r)]^{2r} = \frac{E \left| T(\upsilon_1) \right|^{2r}}{\left(E \left| T(\upsilon_1) \right|^r \right)^2}, \quad (20)$$

$$kur_R(r) \equiv [d_R(2r)/d_R(r)]^{2r} = \frac{E \left| T(\upsilon_2) \right|^{2r}}{\left(E \left| T(\upsilon_2) \right|^r \right)^2}, \quad (21)$$

where each depends on only one of the shape parameters υ_1 and υ_2. For the case in which $\upsilon_1 = \upsilon_2 = \upsilon$, the skewness parameter α has an interesting interpretation. Recall that $\alpha^* = \alpha$ when $\upsilon_1 = \upsilon_2$. It follows from (19) that

$$d_L(r) = 2\alpha \left(E \left| T(\upsilon) \right|^r \right)^{1/r}, \quad d_R(r) = 2(1 - \alpha) \left(E \left| T(\upsilon) \right|^r \right)^{1/r}. \quad (22)$$

This implies that the ratio of the probability (α) that Y occurs on the left side of μ to the probability $(1 - \alpha)$ that Y occurs on the right side of μ is equal to the ratio of the left deviation $d_L(r)$ to the right deviation $d_R(r)$, i.e., $\alpha/(1 - \alpha) = d_L(r)/d_R(r)$. Define $d(r) \equiv d_L(r) + d_R(r)$, the total conditional deviation; then for any $r > 0$,

$$\alpha = d_L(r)/d(r) = d_L(r)/[d_L(r) + d_R(r)], \quad (23)$$

implying that the skewness parameter α can also be interpreted as the ratio of the left deviation $d_L(r)$ to the total deviation $d_L(r) + d_R(r)$.

By substituting (9) into (20) and (21), the left and right (generalized) kurtosis are given as follows:

$$kur_L(r, \upsilon_1) = \frac{\sqrt{\pi} \Gamma(r + \frac{1}{2}) \Gamma(\frac{\upsilon_1}{2} - r) \Gamma(\frac{\upsilon_1}{2})}{\left(\Gamma(\frac{r+1}{2}) \Gamma(\frac{\upsilon_1 - r}{2}) \right)^2}, \forall r \in \left(-\frac{1}{2}, \frac{\upsilon_1}{2} \right), \ \upsilon_1 > 0 \quad (24)$$

and

$$kur_R(r, \upsilon_2) = \frac{\sqrt{\pi} \Gamma(r + \frac{1}{2}) \Gamma(\frac{\upsilon_2}{2} - r) \Gamma(\frac{\upsilon_2}{2})}{\left(\Gamma(\frac{r+1}{2}) \Gamma(\frac{\upsilon_2 - r}{2}) \right)^2}, \forall r \in \left(-\frac{1}{2}, \frac{\upsilon_2}{2} \right), \ \upsilon_2 > 0. \quad (25)$$

We can show that both $kur_L(r,v)$ and $kur_R(r,v)$ are strictly decreasing in v and strictly increasing in r (see Lemma 7 in Appendix A). From the expressions for $kur_L(r,v_1)$ and $kur_R(r,v_2)$ in (24) and (25), the heaviness of the left (or right) tail of the AST is controlled by only v_1(or v_2). If $v_1 < v_2$, then $kur_L(r,v_1) > kur_R(r,v_2)$, implying that the left tail is heavier than the right one; the smaller the value of v_1 (or v_2), the heavier the left (or the right) tail. If $v_i > 4$, then for $r = 2$ the left (or right) kurtosis has the simple expression $kur_L(2,v_1) = 3 + 6/(v_1 - 4)$ (or $kur_R(2,v_2) = 3 + 6/(v_2 - 4)$).

3.2   Value at Risk and Expected Shortfall

The Value at Risk (VaR) and the Expected Shortfall (ES) are two very widely used risk measures, defined for a standard AST random variable Y at a confidence level p or a point in the support of the distribution q as

$$VaR_{AST}(p) \equiv F_{AST}^{-1}(p), \quad ES_{AST}(q) \equiv E(Y \mid Y < q).$$

We will now show that each of these risk measures can be expressed in terms of the cdf and pdf of the standard Student-t, $F_t(\cdot;v)$ and $f_t(\cdot;v)$, with parameter values of v_1 and v_2. $VaR_{AST}(p)$ has been given in (7); we can express $ES_{AST}(q)$ in terms of $F_t(\cdot;v_i)$ and $ES_t(q;v_i)$ (see (40) in Lemma 8 of Appendix A), where $ES_t(q;v) \equiv E(T(v) \mid T(v) < q)$ is the expected shortfall of a standard Student-t r.v. $T(v)$ with degrees of freedom v, and $i = 1, 2$. Note that $ES_t(q;v)$ can be simply expressed as

$$ES_t(q;v) = -\left(\frac{v}{v-1}\right)\left[1 + \frac{1}{v}q^2\right]\frac{f_t(q;v)}{F_t(q;v)}. \tag{26}$$

Then, substituting $ES_t(\cdot;v)$ into (40), we obtain the expression for $ES_{AST}(q)$ as follows: $ES_{AST}(q) =$

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

$$\left.\cdot\left(\frac{v_2}{v_2-1}\right)\left(\left[1 + \frac{1}{v_2}\left(\frac{q \vee 0}{2(1-\alpha^*)}\right)^2\right]^{(1-v_2)/2} - 1\right)\right\}, \tag{27}$$

where again $a \wedge b = \min\{a,b\}$ and $a \vee b = \max\{a,b\}$, and B is defined as in (3).

For $q < 0$, v_2 vanishes from the expression and existence of the ES requires only $v_1 > 1$.

When considering the ES as a function of the confidence level p by taking $q = VaR_{AST}(p) = F_{AST}^{-1}(p)$, we obtain $ES_{AST}^*(p) \equiv ES_{AST}\left[F_{AST}^{-1}(p)\right] =$

$$
-\frac{4B}{p}\left\{\frac{v_1\alpha^{*2}}{v_1-1}\left[1+\frac{1}{v_1}\left(F_t^{-1}\left(\frac{p\wedge\alpha}{2\alpha};v_1\right)\right)^2\right]^{(1-v_1)/2} + \frac{v_2(1-\alpha^*)^2}{v_2-1}\cdot\right.
$$
$$
\left.\cdot\left(\left[1+\frac{1}{v_2}\left(F_t^{-1}\left(\frac{p\vee\alpha+1-2\alpha}{2(1-\alpha)};v_2\right)\right)^2\right]^{(1-v_2)/2} - 1\right)\right\}.
$$

4   Asymptotic Properties of the MLE

We now investigate asymptotic properties and finite-sample performance of ML estimators of the parameters of the AST distribution. In order to obtain a relatively straightforward form of information matrix of the MLE, we adopt the alternative definition of the AST density given in (5). This density is a parameter transformation (re-scaling) of the original one in (1). For any one-to-one parameter transformation, $\zeta = h(\theta)$, the information matrices of MLEs for ζ and θ, denoted by $J(\zeta)$ and $I(\theta)$, can be shown to have the following relationship:

$$
J^{-1}(\zeta) = \nabla_{\theta'}h(\theta)I^{-1}(\theta)\nabla_{\theta}h(\theta),
$$

where $\nabla_{\theta'}h(\theta) = (\nabla_{\theta}h(\theta))' = (\partial\zeta_i/\partial\theta_j)$ whose element in the i-th row and j-th column is $\partial\zeta_i/\partial\theta_j$, $i, j = 1, 2, ...5$.

Now consider the MLE of the parameters of the AST. Let $f(y;\theta)$ be the AST density (5). The true value of θ is denoted by $\theta_0 = (\alpha_0, v_{01}, v_{02}, \mu_0, \sigma_0)$. Suppose that $\theta_0 \in \Lambda \equiv \{\theta \mid \theta = (\alpha, v_1, v_2, \mu, \sigma), \sigma, v_1, v_2 > 0, \alpha \in (0,1), \mu \in (-\infty, +\infty)\}$, the parameter space. Given an i.i.d. sample $y = (y_1, y_2, ..., y_T)$ of size T we can write the log-likelihood function $l_T(\theta \mid y) \equiv \sum_{t=1}^{T}\ln f(y_t;\theta)$ as follows:

$$
l_T(\theta \mid y) = -T\ln\sigma - \frac{v_1+1}{2}\sum_{t=1}^{T}\ln\left[1+\frac{1}{v_1}\left(\frac{y_t-\mu}{2\alpha\sigma K(v_1)}\right)^2\right]1(y_t \le \mu)
$$
$$
-\frac{v_2+1}{2}\sum_{t=1}^{T}\ln\left[1+\frac{1}{v_2}\left(\frac{y_t-\mu}{2(1-\alpha)\sigma K(v_2)}\right)^2\right]1(y_t > \mu).
$$

Note that because the log-likelihood function is differentiable only once at $\mu,$, the AST distribution does not satisfy the usual regularity conditions under which the ML estimator has $\sqrt{T}$-asymptotics. In this case, however,

10

we can still establish the usual asymptotics by using Theorem 2.5 in Newey and McFadden (1994, p2131) and Theorem 3 as well as its corollary in Huber (1967). In addition, we obtain the closed-form expression for the Fisher information matrix $I(\theta)$. We use the notation $H(\theta)$ for the Hessian matrix.

Proposition 1 *The MLE $\widehat{\theta}_T$ of θ_0 is consistent, i.e., $\widehat{\theta}_T \to^p \theta_0$.*

Proof: See Appendix B.

Proposition 2 *The information matrix equality $I(\theta_0) = -H(\theta_0)$ holds. The elements of the Fisher information matrix, denoted by ϕ_{ij}, are*

$$\phi_{ij} \equiv E\left[\frac{\partial \ln f(y_t; \theta_0)}{\partial \theta_i} \cdot \frac{\partial \ln f(y_t; \theta_0)}{\partial \theta_j}\right], \qquad (28)$$

where $\phi_{ij} = \phi_{ji}$ and θ_j represents the jth element of the parameter vector $\theta = (\alpha, \upsilon_1, \upsilon_2, \mu, \sigma)^T$, and:

$$\phi_{11} = 3\left[\frac{\upsilon_1+1}{\alpha(\upsilon_1+3)} + \frac{\upsilon_2+1}{(1-\alpha)(\upsilon_2+3)}\right]; \qquad \phi_{12} = -\frac{1}{\upsilon_1+1} + \frac{\upsilon_1 D(\upsilon_1)}{\upsilon_1+3};$$

$$\phi_{13} = \frac{1}{\upsilon_2+1} - \frac{\upsilon_2 D(\upsilon_2)}{\upsilon_2+3}; \quad \phi_{14} = -\frac{2}{3\sigma}\phi_{11}; \quad \phi_{15} = \frac{2}{\sigma}\left[\frac{\upsilon_1}{\upsilon_1+3} - \frac{\upsilon_2}{\upsilon_2+3}\right];$$

$$\phi_{22} = \frac{\alpha}{2}\left\{\frac{\upsilon_1 D^2(\upsilon_1)}{\upsilon_1+3} - \frac{2D(\upsilon_1)}{\upsilon_1+1} - D'(\upsilon_1)\right\}; \quad \phi_{23} = 0; \quad \phi_{25} = \frac{\alpha}{\sigma}\phi_{12};$$

$$\phi_{24} = \frac{1}{\sigma}\left[\frac{1}{\upsilon_1+1} - \frac{\upsilon_1+1}{\upsilon_1+3}D(\upsilon_1)\right]; \quad \phi_{34} = -\frac{1}{\sigma}\left[\frac{1}{\upsilon_2+1} - \frac{\upsilon_2+1}{\upsilon_2+3}D(\upsilon_2)\right]; \qquad (29)$$

$$\phi_{33} = \frac{1-\alpha}{2}\left\{\frac{\upsilon_2 D^2(\upsilon_2)}{\upsilon_2+3} - \frac{2D(\upsilon_2)}{\upsilon_2+1} - D'(\upsilon_2)\right\}; \qquad \phi_{35} = -\frac{1-\alpha}{\sigma}\phi_{13};$$

$$\phi_{44} = \frac{1}{4\sigma^2}\left[\frac{\upsilon_1+1}{\alpha(\upsilon_1+3)}\frac{1}{K^2(\upsilon_1)} + \frac{\upsilon_2+1}{(1-\alpha)(\upsilon_2+3)}\frac{1}{K^2(\upsilon_2)}\right]; \quad \phi_{45} = -\frac{2}{3\sigma}\phi_{15};$$

$$\phi_{55} = \frac{2}{\sigma^2}\left[\alpha\frac{\upsilon_1}{\upsilon_1+3} + (1-\alpha)\frac{\upsilon_2}{\upsilon_2+3}\right];$$

where all the ϕ_{ij} are evaluated at the true values $(\alpha_0, \upsilon_{01}, \upsilon_{02}, \mu_0, \sigma_0)$, $K(\cdot)$ is defined in (1), $D(\upsilon) \equiv \Psi(\frac{\upsilon+1}{2}) - \Psi(\frac{\upsilon}{2})$, and $\Psi(\upsilon) \equiv \Gamma'(\upsilon)/\Gamma(\upsilon)$ is the digamma function.

Proof: See Appendix C.

Note that for the SST ($v_1 = v_2 = v$), its score component $\frac{\partial \ln f}{\partial v}$ is the sum of the AST score components $\frac{\partial \ln f}{\partial v_1}$ and $\frac{\partial \ln f}{\partial v_2}$. Thus, by combining the terms of ϕ_{ij} involving v_1 and v_2, i.e., $\phi_{12} + \phi_{13}$, $\phi_{22} + \phi_{33}$, $\phi_{24} + \phi_{34}$ and $\phi_{25} + \phi_{35}$, we can obtain the information matrix for the MLE of the SST parameters (α, v, μ, σ); the result appears in Gómez et al. (2007, Proposition 2.2).

Proposition 3 *The MLE $\widehat{\theta}_T$ of θ_0 is asymptotically normal,*

$$\sqrt{T}(\widehat{\theta}_T - \theta_0) \xrightarrow{D} N(0, I^{-1}(\theta_0)),$$

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

$$I(\theta_0) \equiv E[(\nabla_\theta \ln f(y_t; \theta_0))(\nabla_\theta \ln f(y_t; \theta_0))'];$$

provided by (29); it can be consistently estimated by $I(\widehat{\theta}_T)$.

Proof: See Appendix B.

From the proof of the Proposition, we can see that $I(\theta)$ is continuous in some neighborhood of θ_0, so it follows from the consistency of $\widehat{\theta}_T$ that $I(\widehat{\theta}_T)$ is a consistent estimator of $I(\theta_0)$.

5   Simulation performance of the MLE

To assess the asymptotic properties of the MLE in finite samples we report a numerical investigation of bias and variance of the estimators using sample sizes of $T = 1000$ and 5000. We choose $\mu_0 = 0$, $\sigma_0 = 1$ and various different true values of (α, v_1, v_2): $\alpha = 0.3$, 0.7 and $v_{1,2} = (0.7, 2.5), (2.0, 2.0)$, and $(2.0, 5.0)$; these cases are representative of a larger number of simulations producing qualitatively similar results. For each set of true values of the parameters and every sample size, $N = 10000$ simulated samples are drawn from the AST distribution with that set of parameter values, and then ML estimates $\widehat{\theta}^i$ $(i = 1, 2, ..., N)$ are obtained using these samples. We obtain the sample means and standard errors of the MLE's of the parameters on these 10 000 replications, denoted respectively by $M(\widehat{\theta})$ and $SE(\widehat{\theta})$,

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

and compare these standard errors with the theoretical standard deviations which are taken from the square root of the diagonal elements of the Cramèr-Rao bound (i.e., $I^{-1}(\theta)/T$). Simulation results are reported in Table 1(a/b)

and in Figure 1. All entries in Table 1a (labeled 'mean') report $M(\widehat{\theta})$, and those in Table 1b ('se') report $SE(\widehat{\theta})$.

To describe the ratios of simulated standard errors $SE(\widehat{\theta})$ to theoretical ones from $I^{-1}(\widehat{\theta})/T$, we report results graphically in Figure 1 for the larger set of sample sizes $T = \{1000, 2000, \ldots, 10000\}$; these results are conveniently viewed in graphical form since asymptotically the result will converge on unity, and we wish to see examples of the speed of convergence and the degree of finite-sample discrepancy.

Finally, note that all random samples from a standard AST are generated by using the stochastic representation (8) of the AST multiplied by $B \equiv \alpha K(v_1) + (1 - \alpha)K(v_2)$, $X \equiv BY$, which has the AST density (5) with shape parameters (α, v_1, v_2), location $\mu = 0$ and scale $\sigma = 1$.

From these simulations the estimates $\widehat{\theta}$ of all parameters appear asymptotically unbiased in each case and their variances appear to be approaching the Cramèr-Rao bound. However, ML estimates of the tail parameters have a slower convergence rate than those of the other parameters. In fact, skewness, scale and location parameters can be estimated well even for sample sizes smaller than 500; however, even for moderately large values of the tail parameters, such as 5.0, a sample size of 1000 or 2000 may not be large enough to give a good estimate. The highest variances that we observed arose for large values of v_1 where α was less than 0.5, and correspondingly for v_2 where α was greater than 0.5; the last lines of Table 1a/1b and panel D of Figure 1 illustrate such a case. Note in Figure 1 that the vertical scale in panel D differs from those of panels A-C.

Estimates of tail parameters depend crucially on the relatively sparse tail observations, suggesting that relative to the Student-t and the SST which have only a single tail parameter, approximately double the number of observations will be needed in order to obtain good tail parameter estimates in the AST, because the two tail parameters in the AST are distinct. For $\alpha = 0.3$ in our simulation studies there are fewer observations on the left side, so estimates of the left tail parameter should show poorer finite sample performance than those of the right tail parameter. Note that a smaller (larger) value of a tail parameter implies a heavier (thinner) tail, so that there are more (fewer) observations in the tail. As well, the shape of the distribution changes less with a one-unit change in the tail parameters when the value is large; that is, sensitivity of shape is greater at small values. These considerations suggest that we should observe lower standard errors for small tail parameter values than for large values.

6   Concluding remarks

Many processes display a relative frequency of extreme values which far exceeds what could be accounted for by a Gaussian distribution. This is true in particular for financial data, where the Student-t distribution has commonly been found valuable in modelling conditional returns. However, equality of the relative frequency of extreme returns in left and right tails (losses and gains) often seems violated in practice. Hence generalizations of the Student-t that allow asymmetry are potentially valuable in empirical modelling and forecasting.

The present study offers a three-parameter form which is more general than those available in the literature. The proposed distribution allows analytical computation of important quantities related to risk, and ML estimation of parameters with the usual $\sqrt{T}$ asymptotics. We show by simulation that finite-sample performance of ML estimation is reasonable, and also through empirical analysis that the potential of the more general form is realized both in better in-sample fits, and in better forecasts of tail-dependent quantities of interest such as the expected shortfall. This distribution therefore appears to offer a device for continuing to increase the subtlety of our understanding of financial returns and other heavy-tailed data.

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] Abramowitz, M. and Stegun, I. A. (1970). Handbook of Mathematical Functions, National Bureau of Standard, Applied Math. Series 55, U.S. Department of Commerce.

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

[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] Arnold, B.C. and Groeneveld, R.A. (1995). Measuring skewness with respect to the mode. *Am. Statistician*, 49, 34-38.

[6] Artin, E. (1964). *The Gamma Function*, Holt, Rinehart and Winston, Inc.

[7] 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.

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

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

[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] Farrell, O. J. and Ross, B. (1963). Solved Problems: Gamma and Beta Functions, Legendre Polynomials, Bessel Functions. The Macmillan Company, Now York.

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

[13] Franses, P.H., M. van der Leij and R. Paap (2008). A simple test for GARCH against a stochastic volatility model. *Journal of Finanical Econometrics* 6, 291-306.

[14] Gómez, H.W., F.J. Torres and H. Bolfarine (2007). Large-sample inference for the epsilon-skew-t distribution. *Communications in Statistics—Theory and Methods*, 36, 73-81.

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

[16] Huber, P. J. (1967). The behavior of Maximum Likelihood estimates under nonstandard conditions, in: L.M. LeCam and J. Neyman, eds., *Proceedings of the Fifth Berkeley Symposium on Mathematical Statistics and Probability, Berkeley: University of California Press.*

[17] 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.

[18] Kagan, A. M., Linnik, Yu. V. and Rao, C. R. (1973). *Characterization Problems in Mathematical Statistics*, Wiley, New York.

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

[20] 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.

[21] Mittnik, S., Rachev, T. and Paolella, M.S. (1998). Stable Paretian modelling in finance: Some empirical and theoretical aspects. In: Adler et al. (Eds.), A Practical Guide to Heavy Tails. Birkhäuser.

[22] Newey, W.K. and McFadden, D. (1994). Large sample estimation and hypothesis testing, in: R. Engle and D. McFadden, eds., *Handbook of Econometrics, Vol. 4,* Amsterdam: North-Holland.

[23] 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.

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

[25] Zhu, D. and J.W. Galbraith (2009). Forecasting expected shortfall with a generalized asymmetric Student-t distribution. Working paper: www.mcgill.ca/economics/papers .

7 Appendix A

Appendix A provides some lemmas that will be used in the proofs of properties of the AST.

We will use the Gamma function of a positive real variable only, and so for $x > 0$ we take $\Gamma(x) = \int_0^\infty t^{x-1} e^{-t} dt$ as our definition of the Gamma function. $\Gamma(x)$ has derivatives of arbitrarily high order:

$$\frac{\Gamma'(x)}{\Gamma(x)} = -C - \frac{1}{x} + \sum_{i=1}^\infty (\frac{1}{i} - \frac{1}{x+i}), \tag{30}$$

$$\frac{d^{k-1}}{dx^{k-1}} \left(\frac{\Gamma'(x)}{\Gamma(x)} \right) = \sum_{i=0}^\infty \frac{(-1)^k (k-1)}{(x+i)^k}, \text{ for } k \geq 2; \tag{31}$$

C is Euler's constant. See e.g. Artin, 1964, pp. 16 ff. for these and other properties. Let $\Psi(x) \equiv \Gamma'(x)/\Gamma(x)$; this is called the *digamma function*.

Lemma 4 *Let* $D(\upsilon) \equiv \Psi(\frac{\upsilon+1}{2}) - \Psi(\frac{\upsilon}{2})$ *for any* $\upsilon > 0$. *Then* $\Psi(\upsilon)$ *is strictly increasing while* $D(\upsilon)$ *is strictly decreasing, and the following equalities hold:*

$$\Psi(\upsilon + 1) = \frac{1}{\upsilon} + \Psi(\upsilon), \quad D(\upsilon + 2) = -\frac{2}{\upsilon(\upsilon+1)} + D(\upsilon). \tag{32}$$

Proof. From (31), taking $k = 2$, we get $\Psi'(x) = \sum_{i=0}^\infty 1/(x+i)^2 > 0$ for all x, implying that $\Psi(\upsilon)$ is a strictly increasing function. From the above expression for $\Psi'(x)$, we can also see that $\Psi'(x)$ is a strictly deceasing function for $x > 0$, implying $D'(\upsilon) = \Psi'(\frac{\upsilon+1}{2}) - \Psi'(\frac{\upsilon}{2}) < 0$. So $D(\upsilon)$ is strictly decreasing. If the first equality in (32) holds, then the second one is easily verified. Now we proceed to show the first equality. In fact, differentiating both sides of $\Gamma(x+1) = x\Gamma(x)$ leads to $\Gamma'(x+1)/\Gamma(x) = 1 + x\Psi(x)$, and then rewriting it does yield the first equality in (32). ∎

Lemma 5 *For any* $\upsilon > 0$, *recall* $K(\upsilon) \equiv \Gamma((\upsilon+1)/2)/[\sqrt{\pi\upsilon}\Gamma(\upsilon/2)]$. *Then the following equalities hold:*

$$\frac{1}{\upsilon} + 2\frac{K'(\upsilon)}{K(\upsilon)} = \Psi(\frac{\upsilon+1}{2}) - \Psi(\frac{\upsilon}{2}) = D(\upsilon), \tag{33}$$

$$\left(\frac{\upsilon}{\upsilon + 2j} \right)^{1/2} \frac{K(\upsilon)}{K(\upsilon + 2j)} = \begin{cases} 1, & j = 0; \\ \upsilon/(\upsilon+1), & j = 1; \\ \upsilon(\upsilon+2)/[(\upsilon+1)(\upsilon+3)], & j = 2. \end{cases} \tag{34}$$

Proof. The proofs are immediate. For equality (33), taking the log of the expression for $K(v)$ and then differentiating both sides, it follows that

$$
\begin{aligned}
\frac{K'(v)}{K(v)} &= \frac{d\ln K(v)}{dv} = d\left[\ln\Gamma((v+1)/2) - \ln\Gamma(v/2) - \frac{1}{2}\ln v - \ln\sqrt{\pi}\right]/dv \\
&= \frac{1}{2}\left[\Psi(\frac{v+1}{2}) - \Psi(\frac{v}{2}) - \frac{1}{v}\right].
\end{aligned}
$$

So we have shown that (33) holds. From the definition of $K(v)$, the left side of equality (34) is expressed as

$$
\left(\frac{v}{v+2j}\right)^{1/2}\frac{K(v)}{K(v+2j)} = \frac{\Gamma((v+1)/2)}{\Gamma(v/2)}\frac{\Gamma(v/2+j)}{\Gamma((v+1)/2+j)}.
$$

Using the fact $\Gamma(v+1) = v\Gamma(v)$, the proof of equality (34) is easily completed. ∎

Lemma 6 *For any $v > 0$, we have the following integral equalities:*

$$
\int_0^{+\infty}(1+z^2)^{-(v+1)/2}dz = \frac{1}{2\sqrt{v}K(v)} \tag{35}
$$

$$
\int_0^{+\infty}(1+z^2)^{-(v+1)/2}\ln(1+z^2)dz = \frac{D(v)}{2\sqrt{v}K(v)} \tag{36}
$$

$$
\int_0^{+\infty}(1+z^2)^{-(v+1)/2}\left[\ln(1+z^2)\right]^2 dz = \frac{D^2(v) - 2D'(v)}{2\sqrt{v}K(v)} \tag{37}
$$

where $K(\cdot)$ and $D(\cdot)$ are defined as the above, and $D'(v)$ is the derivative function of $D(v)$. Below we will continue to use $\Gamma(v)$, $\Psi(v)$, $K(v)$ and $D(v)$ to denote these functions defined earlier.

Proof. For the Student-t density $f_t(x) = K(v)(1 + x^2/v)^{-(v+1)/2}$, from $\int_{-\infty}^{+\infty}f_t(x)dx = 1$, we obtain equality (35) immediately. Differentiating both sides of equality (35) with respect to v, by Lemma 3.6 of Newey and McFadden (1994, p.2152) that ensures that the order of differentiation and integration can be interchanged, it follows that

$$
-\frac{1}{2}\int_0^{+\infty}(1+z^2)^{-(v+1)/2}\ln(1+z^2)dz = \frac{1}{2\sqrt{v}K(v)}\left[-\frac{1}{2v} - \frac{K'(v)}{K(v)}\right].
$$

Rewriting this equality and combining with equality (33) yields (36). Similarly, by differentiating both sides of (36) with respect to v and combining with (33), we obtain (37). ∎

18

Lemma 7 *The inequalities $\partial kur_L(r, \upsilon)/\partial \upsilon < 0$ and $\partial kur_L(r, \upsilon)/\partial r > 0$ hold for the AST.*

Proof. Taking the log of $kur_L(r, \upsilon)$, we have two partial derivatives as follows:

$$\frac{\partial \ln kur_L}{\partial \upsilon} = \frac{1}{2}\left[\Psi(\frac{\upsilon}{2}) - \Psi(\frac{\upsilon-r}{2})\right] - \frac{1}{2}\left[\Psi(\frac{\upsilon-r}{2}) - \Psi(\frac{\upsilon-2r}{2})\right] < 0 \quad (38)$$

and

$$\frac{\partial \ln kur_L}{\partial r} = \left[\Psi(\frac{2r+1}{2}) - \Psi(\frac{r+1}{2})\right] + \left[\Psi(\frac{\upsilon-r}{2}) - \Psi(\frac{\upsilon-2r}{2})\right] > 0. \quad (39)$$

The inequality (38) can be verified by using the mean value theorem and the fact that $\Psi'(x)$ is strictly decreasing for $x > 0$; then (39) follows immediately because $\Psi(x)$ is an increasing function. ■

Lemma 8 *The expected shortfall of the AST distribution, $ES_{AST}(q)$, can be expressed as*

$$\frac{1}{F_{AST}(q)}\left\{4\alpha\alpha^* ES_t\left(\frac{q \wedge 0}{2\alpha^*}; \upsilon_1\right) F_t\left(\frac{q \wedge 0}{2\alpha^*}; \upsilon_1\right) + 4(1-\alpha)(1-\alpha^*)\right.$$
$$\left. \cdot \left[ES_t\left(\frac{q \vee 0}{2(1-\alpha^*)}; \upsilon_2\right) F_t\left(\frac{q \vee 0}{2(1-\alpha^*)}; \upsilon_2\right) - \frac{1}{2}ES_t(0; \upsilon_2)\right]\right\}. \quad (40)$$

Proof. The expression for $ES_{AST}(q)$ in (40) is unified from the following two cases. When $q \leq 0$, by the definition of expected shortfall and the change of variable, $u = x/(2\alpha^*)$, we have

$$
\begin{aligned}
ES_{AST}(q) &\equiv E(Y \mid Y < q) \\
&= \left(\int_{-\infty}^{q} x f_{AST}(x; \alpha, \upsilon_1, \upsilon_2)dx\right) / F_{AST}(q) \\
&= \frac{\int_{-\infty}^{q} x \frac{\alpha}{\alpha^*} f_t(\frac{x}{2\alpha^*}; \upsilon_1)dx}{F_{AST}(q)} = 4\alpha\alpha^* \frac{\int_{-\infty}^{q/(2\alpha^*)} u f_t(u; \upsilon_1)du}{2\alpha F_t(\frac{q}{2\alpha^*}; \upsilon_1)} \\
&= 2\alpha^* ES_t(\frac{q}{2\alpha^*}; \upsilon_1).
\end{aligned}
$$

Similarly, when $q > 0$, using the change of variable $z = x/(2(1 - \alpha^*))$, and noting that

$$\int_{-\infty}^{0} x \frac{\alpha}{\alpha^*} f_t(\frac{x}{2\alpha^*}; \upsilon_1)dx = 4\alpha\alpha^* ES_t(0; \upsilon_1)F_t(0; \upsilon_1)$$

19

and

$$\int_{-\infty}^{q} z f_t(z; \upsilon) dz = ES_t(q; \upsilon) F_t(q; \upsilon), \ \forall q,$$

we obtain

$$
\begin{aligned}
ES_{AST}(q) &= \frac{\int_{-\infty}^{0} x \frac{\alpha}{\alpha^*} f_t(\frac{x}{2\alpha^*}; \upsilon_1) dx + \int_{0}^{q} x \frac{1-\alpha}{1-\alpha^*} f_t(\frac{x}{2(1-\alpha^*)}; \upsilon_2) dx}{F_{AST}(q)} \\
&= \frac{1}{F_{AST}(q)} \left\{ 4\alpha\alpha^* ES_t(0; \upsilon_1) F_t(0; \upsilon_1) + 4(1-\alpha)(1-\alpha^*) \cdot \right. \\
&\qquad \left. \cdot \left[\int_{-\infty}^{q/(2(1-\alpha^*))} z f_t(z; \upsilon_2) dz - \int_{-\infty}^{0} z f_t(z; \upsilon_2) dz \right] \right\} \\
&= \frac{1}{F_{AST}(q)} \left\{ 4\alpha\alpha^* ES_t(0; \upsilon_1) F_t(0; \upsilon_1) + 4(1-\alpha)(1-\alpha^*) \cdot \right. \\
&\qquad \left. \cdot \left[ES_t(\frac{q}{2(1-\alpha^*)}; \upsilon_2) F_t(\frac{q}{2(1-\alpha^*)}; \upsilon_2) - \frac{1}{2} ES_t(0; \upsilon_2) \right] \right\} . \blacksquare
\end{aligned}
$$

8   Appendix B

Appendix B is devoted to establishing consistency and asymptotic normality of the MLE of all parameters of the AST distribution.

Proof of Proposition 1 (consistency of MLE). The consistency of the MLE $\widehat{\theta}_T$ can be shown by verifying the conditions of Theorem 2.5 in Newey and McFadden (1994, p.2131), which holds under conditions that are primitive and also quite weak. Condition (ii) of Theorem 2.5, compactness of the parameter set, is ensured by considering a compact parameter set $\Theta \subset \Lambda$ such that it includes the true parameter θ_0 as an interior point. Condition (iii) of Theorem 2.5 requires that the log-likelihood $\ln f(y \mid \theta)$ be continuous at each $\theta \in \Theta$ with probability one. This condition holds by inspection. We only need to check the identification condition and dominance condition (corresponding to conditions (i) and (iv) of Theorem 2.5 respectively).

For the identification condition, it is sufficient to show that for any given $\theta \neq \theta_0$ and $\theta \in \Theta$,

$$\ln f(y \mid \theta) \neq \ln f(y \mid \theta_0), \text{ a.e.} \tag{41}$$

on a set of positive probability. The fact that the AST random variable Y has a positive probability on any interval will be used in the proof. If $\mu \neq \mu_0$, say, $\mu > \mu_0$, then on interval $(\mu_0, \mu]$ the log-density function $\ln f(y \mid \theta)$ is strictly increasing, but $\ln f(y \mid \theta_0)$ decreases strictly, so (41) holds for $y \in (\mu_0, \mu]$. Now suppose $\mu = \mu_0$. We can show that (41) is true on

$(-\infty, \mu_0]$ or $(\mu_0, +\infty)$ respectively if $v_1 \neq v_{01}$ or $v_2 \neq v_{02}$. In fact, assuming $v_2 \neq v_{02}$ and letting $C(\theta) = (v_2 + 1)/2$, for $y \in (\mu_0, +\infty)$, we have $\ln f(y \mid \theta) = -\ln \sigma - C(\theta) \ln R(y; \theta)$ with $\mu = \mu_0$, and $\ln f(y \mid \theta_0) = -\ln \sigma_0 - C(\theta_0) \ln R(y; \theta_0)$, where $R(y; \theta)$ is defined in (55). Note that $R(y; \theta)$ with $\mu = \mu_0$ and $R(y; \theta_0)$ are quadratic and strictly increasing on $(\mu_0, +\infty)$. Thus, both log-density functions intersect at no more than two points, so that (41) holds on $(\mu_0, +\infty)$. Similarly, for $\mu = \mu_0$, $v_1 = v_{01}$ and $v_2 = v_{02}$, it is easy to show that (41) holds if $\alpha \neq \alpha_0$ or $\sigma \neq \sigma_0$ (see Newey and McFadden, p. 2126).

The dominance condition of Theorem 2.5, $E[\sup_{\theta \in \Theta} |\ln f(Y \mid \theta)|] < \infty$, can be verified by the compactness of parameter set Θ and equalities (63) and (64). The parameter set Θ is assumed to be compact, so that any continuous function of θ is bounded on Θ, and it is easy to show that there exist certain positive constants K_j $(j = 1, ..., 5)$ such that

$$1 \leq L(y; \theta) \leq K_1 L(y; \theta_0), \quad 1 \leq R(y; \theta) \leq K_2 R(y; \theta_0), \quad \forall \theta \in \Theta \qquad (42)$$

and thus

$$|\ln f(y \mid \theta)| \leq K_3 + K_4 \left[\ln L(y; \theta_0)\right] 1(y < \mu_0) + K_5 \left[\ln R(y; \theta_0)\right] 1(y > \mu_0)$$

for all $\theta \in \Theta$. Using equalities (63) and (64), the dominance condition follows.∎

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

$$\lambda(\theta) = E\psi(y, \theta), \quad u(y, \theta, d) \equiv \sup_{\theta^* \in D^*} |\psi(y, \theta^*) - \psi(y, \theta)|, \qquad (43)$$

where $D^* \equiv \{\theta^* \mid |\theta^* - \theta| \leq d\}$ and all expectations are always taken with respect to the true underlying distribution $f(y; \theta_0)$ with $\theta_0 = (\alpha_0, v_{01}, v_{02}, \mu_0, \sigma_0)$. Similar to Example 1 of Huber (1967), the condition N-1 (i.e., for each fixed θ, $\psi(y, \theta)$ is measurable and separable: see Assumption (A-1) of Huber (1967)) is immediate; both conditions (N-2) and (N-4), i.e., $\lambda(\theta_0) = 0$ and $E[|\psi(y, \theta_0)|^2] < \infty$, hold immediately from (71) and the fact that ϕ_{ii} in (28) are finite. By the definition of the MLE $\widehat{\theta}$, we have $\sum_{t=1}^{T} \psi(y_t, \widehat{\theta}) = 0$, implying that Equation (27) of Huber (1967) holds. Since consistency has been proved, the remaining condition of Huber's (1967) Theorem 3 is the

21

condition (N-3): there are strictly positive numbers a, b, c, d_0 such that

$$
\begin{align}
|\lambda(\theta)| &\geq a\,|\theta - \theta_0|, \quad \text{for } |\theta - \theta_0| \leq d_0, \tag{44} \\
E\left[u(y, \theta, d)\right] &\leq bd, \quad \text{for } |\theta - \theta_0| + d \leq d_0,\ d \geq 0, \tag{45} \\
E[u(y, \theta, d)^2] &\leq cd, \quad \text{for } |\theta - \theta_0| + d \leq d_0,\ d \geq 0, \tag{46}
\end{align}
$$

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

To prove (44), it is sufficient to show

$$
\sup_{\theta \in \Theta} \left| tr \frac{\partial^2 \ln f(y, \theta)}{\partial \theta \partial \theta'} \right| < M < \infty, \tag{47}
$$

where M is some positive constant; it then follows by Lemma 3.6 of Newey and McFadden (1994, p2152) that $\lambda(\theta)$ is continuously differentiable in any neighborhood of θ_0, hence (44) can be verified by using the mean-value theorem and the fact that Hessian $H(\theta_0)$ is negative definite. Actually, (47) is immediate by (42) and the fact that the expressions for the diagonal elements of $\frac{\partial^2 \ln f(y,\theta)}{\partial\theta\partial\theta'}$ can always be written as

$$
\begin{align}
A(\tau) &+ \left[A_{11}(\tau) + A_{12}(\tau)L^{-1} + A_{13}(\tau)L^{-2}\right]\mathbf{1}(y < \mu) \\
&+ \left[A_{21}(\tau) + A_{22}(\tau)R^{-1} + A_{23}(\tau)R^{-2}\right]\mathbf{1}(y > \mu), \tag{48}
\end{align}
$$

where $A(\cdot)$ and $A_{ij}(\cdot)$ are some continuous functions of shape and scale parameters $\tau \equiv (\alpha, \upsilon_1, \upsilon_2, \sigma)$, and $L = L(y, \theta)$ and $R = R(y, \theta)$ are defined as in (54) and (55).

Now we check condition (45). Separate the location parameter from the other parameters, $\tau = (\alpha, \upsilon_1, \upsilon_2, \sigma)$, i.e. $\theta = (\tau, \mu)$ and $\theta^* = (\tau^*, \mu^*)$. Then

$$
u(y, \theta, d) \leq \sup_{\theta^* \in D^*} |\psi(y, \tau^*, \mu^*) - \psi(y, \tau^*, \mu)| + \sup_{|\tau^* - \tau| \leq d} |\psi(y, \tau^*, \mu) - \psi(y, \tau, \mu)|. \tag{49}
$$

The condition (45) is easily verified for the second part in (49), because the location μ is fixed and $\psi(y, \tau, \mu)$ as a function of τ is smooth enough. For the first part in (49), note from (56) to (60) that each element of $\psi(y, \tau, \mu)$ can be expressed in the following form:

$$
\begin{align}
C(\tau) &+ \left[\sum_{i=1}^{2} C_{1i}(\tau)(\ln L)^{i-1} + \sum_{i=3}^{4} C_{1i}(\tau)(\mu - y)^{i-3}L^{-1}\right]\mathbf{1}(y < \mu) \\
&+ \left[\sum_{i=1}^{2} C_{2i}(\tau)(\ln R)^{i-1} + \sum_{i=3}^{4} C_{2i}(\tau)(y - \mu)^{i-3}R^{-1}\right]\mathbf{1}(y > \mu) \tag{50}
\end{align}
$$

where $C(\cdot)$ and $C_{ij}(\cdot)$ are also certain continuous functions of $\tau = (\alpha, v_1, v_2, \sigma)$. Without loss of generality, we need just to prove that

$$E\left[\sup_{|\theta^* - \theta| \leq d} |1(y < \mu^*) - 1(y < \mu)| \, |\psi(y, \tau^*, \mu^*)|\right] \quad \leq \quad bd, \quad (51)$$

$$E\left[\sup_{|\theta^* - \theta| \leq d} |\ln L(y, \tau^*, \mu^*) - \ln L(y, \tau^*, \mu)| \, 1(y < \mu)\right] \quad \leq \quad bd, \quad (52)$$

and

$$E\left[\sup_{|\theta^* - \theta| \leq d} \left|(\mu^* - y)^k L(y, \tau^*, \mu^*)^{-1} - (\mu - y)^k L(y, \tau^*, \mu)^{-1}\right| 1(y < \mu)\right] \leq bd$$
$$(53)$$

where $k = 0, 1$. The similar inequalities for $R(y, \theta)$ can be proved in the same way. Equation (51) is immediate by (42) and the boundedness of $|\psi(y, \tau^*, \mu^*)| \, f(y; \theta_0)$. The other two equations (52) and (53) are easily verified by using the mean-value theorem. Finally, verification of condition (46) is similar. ∎

9   Appendix C

Appendix C is devoted to deriving a closed-form expression for the information matrix and to verifying the information matrix equality.

Suppose that y_t ($t = 1, 2, ... T$) are i.i.d. observations from the AST with density $f(y; \theta_0)$ defined in (5), where $\theta_0 = (\alpha_0, v_{01}, v_{02}, \mu_0, \sigma_0)$. Expectations are always taken with respect to the true underlying distribution $f(y; \theta_0)$. Let

$$L \equiv L(y_t; \theta) \equiv 1 + \frac{1}{v_1}\left(\frac{y_t - \mu}{2\alpha\sigma K(v_1)}\right)^2, \quad (54)$$

$$R \equiv R(y_t; \theta) \equiv 1 + \frac{1}{v_2}\left(\frac{y_t - \mu}{2(1 - \alpha)\sigma K(v_2)}\right)^2, \quad (55)$$

where $\theta = (\alpha, v_1, v_2, \mu, \sigma) \in \Lambda$, the parameter space. Then the log-density function with parameter θ is $\ln f(y_t; \theta) =$

$$- \ln \sigma - \frac{v_1 + 1}{2}\left[\ln L(y_t; \theta)\right] 1(y_t < \mu) - \frac{v_2 + 1}{2}\left[\ln R(y_t; \theta)\right] 1(y_t > \mu)$$

and the score vector for observation t, $\frac{\partial}{\partial\theta}\ln f(y_t;\theta)$, is given by

$$
\frac{\partial\ln f}{\partial\alpha} = \frac{v_1+1}{\alpha}\left[1-\frac{1}{L(y_t;\theta)}\right]1(y_t<\mu)
$$
$$
-\frac{v_2+1}{1-\alpha}\left[1-\frac{1}{R(y_t;\theta)}\right]1(y_t>\mu), \tag{56}
$$

$$
\frac{\partial\ln f}{\partial v_1} = \left\{-\frac{1}{2}\ln L(y_t;\theta)+\frac{v_1+1}{2}D(v_1)\frac{L(y_t;\theta)-1}{L(y_t;\theta)}\right\}1(y_t<\mu), \tag{57}
$$

$$
\frac{\partial\ln f}{\partial v_2} = \left\{-\frac{1}{2}\ln R(y_t;\theta)+\frac{v_2+1}{2}D(v_2)\frac{R(y_t;\theta)-1}{R(y_t;\theta)}\right\}1(y_t>\mu), \tag{58}
$$

$$
\frac{\partial\ln f}{\partial\mu} = \frac{v_1+1}{2}\frac{1}{L(y_t;\theta)}\frac{1}{v_1}\frac{2(y_t-\mu)}{[2\alpha\sigma K(v_1)]^2}1(y_t<\mu)
$$
$$
+\frac{v_2+1}{2}\frac{1}{R(y_t;\theta)}\frac{1}{v_2}\frac{2(y_t-\mu)}{[2(1-\alpha)\sigma K(v_2)]^2}1(y_t>\mu), \tag{59}
$$

$$
\frac{\partial\ln f}{\partial\sigma} = -\frac{1}{\sigma}+\frac{v_1+1}{\sigma}\left[1-\frac{1}{L(y_t;\theta)}\right]1(y_t<\mu)
$$
$$
+\frac{v_2+1}{\sigma}\left[1-\frac{1}{R(y_t;\theta)}\right]1(y_t>\mu), \tag{60}
$$

where we used equality (33) in the expressions for the components $\frac{\partial\ln f}{\partial v_1}$ and $\frac{\partial\ln f}{\partial v_2}$. To derive the information matrix $I(\theta_0)\equiv E[\frac{\partial}{\partial\theta}\ln f(y_t,\theta_0)\frac{\partial}{\partial\theta'}\ln f(y_t;\theta_0)]$ and the Hessian $H(\theta_0)\equiv E[\frac{\partial^2}{\partial\theta\partial\theta'}\ln f(y_t;\theta_0)]$ and to verify the information matrix equality $I(\theta_0)=-H(\theta_0)$, the following Lemma is needed.

Lemma 9 *For any $j,m=0,1,2,...$, the following moment equalities hold:*

$$
E\left\{\frac{1}{[L(y_t;\theta_0)]^j}1(y_t<\mu_0)\right\} = \alpha\left(\frac{v_1}{v_1+2j}\right)^{1/2}\frac{K(v_1)}{K(v_1+2j)} \tag{61}
$$
$$
= \begin{cases} \alpha, & j=0, \\ \alpha v_1/(v_1+1), & j=1, \\ \alpha\frac{v_1(v_1+2)}{(v_1+1)(v_1+3)}, & j=2; \end{cases}
$$

$$
E\left\{\frac{1}{[R(y_t;\theta_0)]^j}1(y_t>\mu_0)\right\} = (1-\alpha)\left(\frac{v_2}{v_2+2j}\right)^{1/2}\frac{K(v_2)}{K(v_2+2j)} \tag{62}
$$
$$
= \begin{cases} (1-\alpha), & j=0, \\ (1-\alpha)v_2/(v_2+1), & j=1, \\ (1-\alpha)\frac{v_2(v_2+2)}{(v_2+1)(v_2+3)}, & j=2; \end{cases}
$$

$$E\left\{\frac{\ln L(y_t;\theta_0)}{[L(y_t;\theta_0)]^j}1(y_t<\mu_0)\right\} = \alpha\left(\frac{v_1}{v_1+2j}\right)^{1/2}\frac{K(v_1)}{K(v_1+2j)}D(v_1+2j)$$

$$= \begin{cases} \alpha D(v_1), & j=0, \\ \alpha(\frac{v_1}{v_1+1})D(v_1+2), & j=1; \end{cases} \quad (63)$$

$$E\left\{\frac{\ln R(y_t;\theta_0)}{[R(y_t;\theta_0)]^j}1(y_t>\mu_0)\right\} = (1-\alpha)\left(\frac{v_2}{v_2+2j}\right)^{1/2}\frac{K(v_2)D(v_2+2j)}{K(v_2+2j)}$$

$$= \begin{cases} (1-\alpha)D(v_2), & j=0, \\ (1-\alpha)(\frac{v_2}{v_2+1})D(v_2+2), & j=1; \end{cases} \quad (64)$$

$$E\left\{\frac{(y_t-\mu_0)[\ln L(y_t;\theta_0)]^m}{[L(y_t;\theta_0)]^j}1(y_t<\mu_0)\right\} = -\frac{2^{m+1}m!v_1[2\alpha\sigma K(v_1)]^2}{2\sigma(v_1+2j-1)^{m+1}}, \quad (65)$$

$$E\left\{\frac{(y_t-\mu_0)[\ln R(y_t;\theta_0)]^m}{[R(y_t;\theta_0)]^j}1(y_t>\mu_0)\right\} = \frac{2^{m+1}m!v_2[2(1-\alpha)\sigma K(v_2)]^2}{2\sigma(v_2+2j-1)^{m+1}}, \quad (66)$$

$$E\left\{[\ln L(y_t;\theta_0)]^2 1(y_t<\mu_0)\right\} = \alpha[D^2(v_1)-2D'(v_1)], \quad (67)$$
$$E\left\{[\ln R(y_t;\theta_0)]^2 1(y_t>\mu_0)\right\} = (1-\alpha)[D^2(v_1)-2D'(v_1)], \quad (68)$$

where the right hand sides of all the equalities from (61) to (68) are evaluated at the true values $(\alpha_0, v_{01}, v_{02}, \mu_0, \sigma_0)$.

Proof.[2] We discuss equalities (61), (63), (65) and (67). Other equalities can be proved in the same manner. Note that, for any $j, m = 0, 1, 2, ...,$

$$EL_1(j,m) \equiv E\left\{[L(y_t;\theta)]^{-j}[\ln L(y_t;\theta)]^m 1(y_t<\mu)\right\}$$
$$= \int_{-\infty}^{\mu} [L(y;\theta)]^{-j}[\ln L(y;\theta)]^m \frac{1}{\sigma}[L(y;\theta)]^{-(v_1+1)/2}\}dy,$$

and that $L(y;\theta) \equiv 1 + \frac{1}{v_1}(\frac{y-\mu}{2\alpha\sigma K(v_1)})^2$. Then using the change of variable $z = -\frac{y-\mu}{2\alpha\sigma\sqrt{v_1}K(v_1)}$ yields

$$EL_1(j,m) = 2\alpha\sqrt{v_1}K(v_1)\int_0^{+\infty}(1+z^2)^{-(v_1+2j+1)/2}[\ln(1+z^2)]^m dz. \quad (69)$$

Setting $m = 0$, $m = 1$, and $(j,m) = (0,2)$ respectively, and correspondingly taking into account equality (35) with $v = v_1 + 2j$, equality (36) with $v =$

---

[2]For simplicity, we omit the subscript on the true parameters θ_0 in all the following proofs.

$v_1 + 2j$, and equality (37) with $v = v_1$, we obtain equalities (61), (63), and (67). These proofs use (34). Now consider equality (65). Denote by $EL_2(j, m)$ the expectation of the left side of equality (65), and note that the change of variable $z = \frac{1}{v_1}(\frac{y-\mu}{2\alpha\sigma K(v_1)})^2$ yields

$$EL_2(j, m) = -\frac{v_1[2\alpha\sigma K(v_1)]^2}{2\sigma} \int_0^{+\infty} (1+z)^{-(v_1+2j+1)/2} [\ln(1+z)]^m dz. \quad (70)$$

Subject to $v_1 + 2j - 1 > 0$, by integration by parts it follows that

$$EL_2(j, m) = m\left(\frac{2}{v_1 + 2j - 1}\right) EL_2(j, m-1) = m!\left(\frac{2}{v_1 + 2j - 1}\right)^m EL_2(j, 0).$$

A straightforward calculation for (70) gives $EL_2(j, 0) = -\frac{v_1[2\alpha\sigma K(v_1)]^2}{\sigma(v_1+2j-1)}$. ∎

Lemma 10 *The score vector for observation t, $\frac{\partial}{\partial\theta} \ln f(y_t; \theta)$, satisfies the equation*

$$E\left[\frac{\partial}{\partial\theta} \ln f(y_t; \theta_0)\right] = 0. \quad (71)$$

Proof. By using the equalities from (61) to (68), this Lemma is easily verified. In fact,

(i).

$$
\begin{aligned}
E\left[\frac{\partial \ln f}{\partial \alpha}\right] &= \frac{v_1 + 1}{\alpha} E\left\{\left[1 - \frac{1}{L(y_t; \theta)}\right] 1(y_t < \mu)\right\} \\
&\quad -\frac{v_2 + 1}{1 - \alpha} E\left\{\left[1 - \frac{1}{R(y_t; \theta)}\right] 1(y_t > \mu)\right\} \\
&= \frac{v_1 + 1}{\alpha}\alpha\left(1 - \frac{v_1}{v_1 + 1}\right) - \frac{v_2 + 1}{1 - \alpha}(1 - \alpha)\left(1 - \frac{v_2}{v_2 + 1}\right) = 0.
\end{aligned}
$$

(ii).

$$
\begin{aligned}
E\left[\frac{\partial \ln f}{\partial v_1}\right] &= E\left\{\left[-\frac{\ln L(y_t; \theta)}{2} + \frac{v_1 + 1}{2} D(v_1)\frac{L(y_t; \theta) - 1}{L(y_t; \theta)}\right] 1(y_t < \mu)\right\} \\
&= -\frac{1}{2}\alpha D(v_1) + \frac{v_1 + 1}{2} D(v_1)\alpha\left(1 - \frac{v_1}{v_1 + 1}\right) = 0.
\end{aligned}
$$

(iii). Similarly, we have

$$E\left[\frac{\partial \ln f}{\partial v_2}\right] = E\left\{\left[\frac{v_2 + 1}{2} D(v_2)\frac{R(y_t; \theta) - 1}{R(y_t; \theta)} - \frac{\ln R(y_t; \theta)}{2}\right] 1(y_t > \mu)\right\} = 0.$$

(iv).

$$
\begin{aligned}
E\left[\frac{\partial \ln f}{\partial \mu}\right] &= \frac{\upsilon_1 + 1}{2} E\left\{\frac{1}{L(y_t;\theta)}\frac{1}{\upsilon_1}\frac{2(y_t - \mu)}{[2\alpha\sigma K(\upsilon_1)]^2}1(y_t < \mu)\right\} \\
&\quad +\frac{\upsilon_2 + 1}{2} E\left\{\frac{1}{R(y_t;\theta)}\frac{1}{\upsilon_2}\frac{2(y_t - \mu)}{[2(1 - \alpha)\sigma K(\upsilon_2)]^2}1(y_t > \mu)\right\} \\
&= \frac{\upsilon_1 + 1}{2}\frac{-2}{\sigma(\upsilon_1 + 1)} + \frac{\upsilon_2 + 1}{2}\frac{2}{\sigma(\upsilon_2 + 1)} = 0.
\end{aligned}
$$

(v).

$$
\begin{aligned}
E\left[\frac{\partial \ln f}{\partial \sigma}\right] &= -\frac{1}{\sigma} + \frac{\upsilon_1 + 1}{\sigma}E\left\{\left[1 - \frac{1}{L(y_t;\theta)}\right]1(y_t < \mu)\right\} \\
&\quad +\frac{\upsilon_2 + 1}{\sigma}E\left\{\left[1 - \frac{1}{R(y_t;\theta)}\right]1(y_t > \mu)\right\} \\
&= -\frac{1}{\sigma} + \frac{\upsilon_1 + 1}{\sigma}\alpha\left(1 - \frac{\upsilon_1}{\upsilon_1 + 1}\right) + \frac{\upsilon_2 + 1}{\sigma}\frac{(1 - \alpha)}{\upsilon_2 + 1} \\
&= -\frac{1}{\sigma} + \frac{\alpha}{\sigma} + \frac{1 - \alpha}{\sigma} = 0.\blacksquare
\end{aligned}
$$

Proof of Proposition 2. We prove this by computing expectations on the both sides of the following equations and then verifying them,

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

In the proof, the fact that $1(y_t < \mu)1(y_t > \mu) = 0$ and the equalities (61)-(68) are used repeatedly. In addition, we use $E[\frac{\partial}{\partial\theta}\ln f(y_t;\theta_0)] = 0$ shown in (71) and $D(\upsilon + 2) = -\frac{2}{\upsilon(\upsilon+1)} + D(\upsilon)$ given in (32). Note that by the construction of the AST distribution, the left-tail parameter υ_1 and the right-tail parameter υ_2 have a symmetry property. Hence we do not consider the terms of the information matrix equality involved in the right-tail parameter υ_2.

(a)

$$E\left[\frac{\partial \ln f}{\partial \alpha}\right]^2 = \left(\frac{v_1+1}{\alpha}\right)^2 E\left\{\left(1 - \frac{1}{L(y_t;\theta)}\right)^2 1(y_t < \mu)\right\}$$

$$+ \left(\frac{v_2+1}{1-\alpha}\right)^2 E\left\{\left(1 - \frac{1}{R(y_t;\theta)}\right)^2 1(y_t > \mu)\right\}$$

$$= \left(\frac{v_1+1}{\alpha}\right)^2 \alpha \left[1 - 2\frac{v_1}{v_1+1} + \frac{v_1(v_1+2)}{(v_1+1)(v_1+3)}\right]$$

$$+ \left(\frac{v_2+1}{1-\alpha}\right)^2 (1-\alpha) \left[1 - \frac{2v_2}{v_2+1} + \frac{v_2(v_2+2)}{(v_2+1)(v_2+3)}\right]$$

$$= 3\left[\frac{v_1+1}{\alpha(v_1+3)} + \frac{v_2+1}{(1-\alpha)(v_2+3)}\right];$$

$$E\left[\frac{\partial^2 \ln f}{\partial \alpha^2}\right] = -\frac{v_1+1}{\alpha^2} E\left\{\left[1 + \frac{1}{L(y_t;\theta)} - \frac{2}{[L(y_t;\theta)]^2}\right] 1(y_t < \mu)\right\}$$

$$-\frac{v_2+1}{(1-\alpha)^2} E\left\{\left[1 + \frac{1}{R(y_t;\theta)} - \frac{2}{[R(y_t;\theta)]^2}\right] 1(y_t > \mu)\right\}$$

$$= -\frac{v_1+1}{\alpha^2}\alpha \left[1 + \frac{v_1}{v_1+1} - 2\frac{v_1(v_1+2)}{(v_1+1)(v_1+3)}\right]$$

$$-\frac{v_2+1}{(1-\alpha)^2}(1-\alpha) \left[1 + \frac{v_2}{v_2+1} - 2\frac{v_2(v_2+2)}{(v_2+1)(v_2+3)}\right]$$

$$= -3\left[\frac{v_1+1}{\alpha(v_1+3)} + \frac{v_2+1}{(1-\alpha)(v_2+3)}\right].$$

(**b**)

$$E\left[\frac{\partial \ln f}{\partial v_1}\right]^2$$

$$= E\left\{\left[\frac{[\ln L(y_t;\theta)]^2}{4} + \left(\frac{v_1+1}{2}\right)^2 D^2(v_1)\left(1 - \frac{1}{L(y_t;\theta)}\right)^2\right] 1(y_t < \mu)\right\}$$

$$- \left(\frac{v_1+1}{2}\right) D(v_1) E\left\{\left(1 - \frac{1}{L(y_t;\theta)}\right) \ln L(y_t;\theta) 1(y_t < \mu)\right\}$$

$$= \alpha\frac{D^2(v_1) - 2D'(v_1)}{4} + \left(\frac{v_1+1}{2}\right)^2 D^2(v_1)\frac{3\alpha}{(v_1+1)(v_1+3)}$$

$$- \left(\frac{v_1+1}{2}\right) D(v_1)\alpha\left[D(v_1) - \frac{v_1}{v_1+1}D(v_1+2)\right]$$

$$= \frac{\alpha}{2}\left\{\frac{v_1}{v_1+3}D^2(v_1) - \frac{2}{v_1+1}D(v_1) - D'(v_1)\right\};$$

$$E\left[\frac{\partial^2 \ln f}{\partial v_1^2}\right] = \left[D(v_1) + \frac{v_1+1}{2}D'(v_1)\right] E\left\{\left(1 - \frac{1}{L(y_t;\theta)}\right) 1(y_t < \mu)\right\}$$

$$- \frac{v_1+1}{2}D^2(v_1) E\left\{\left(\frac{1}{L(y_t;\theta)} - \frac{1}{[L(y_t;\theta)]^2}\right) 1(y_t < \mu)\right\}$$

$$= \left[D(v_1) + \frac{v_1+1}{2}D'(v_1)\right] \alpha\left[1 - \frac{v_1}{v_1+1}\right]$$

$$- \frac{v_1+1}{2}D^2(v_1)\alpha\left[\frac{v_1}{v_1+1} - \frac{v_1(v_1+2)}{(v_1+1)(v_1+3)}\right]$$

$$= -\frac{\alpha}{2}\left\{\frac{v_1}{v_1+3}D^2(v_1) - \frac{2}{v_1+1}D(v_1) - D'(v_1)\right\}.$$

(**c**).

$$
\begin{aligned}
E\left[\frac{\partial \ln f}{\partial \mu}\right]^2 &= \frac{1}{v_1}\left[\frac{v_1+1}{2\alpha\sigma K(v_1)}\right]^2 E\left\{\left(\frac{1}{L(y_t;\theta)} - \frac{1}{[L(y_t;\theta)]^2}\right)1(y_t < \mu)\right\} \\
&\quad + \frac{1}{v_2}\left[\frac{v_2+1}{2(1-\alpha)\sigma K(v_2)}\right]^2 E\left\{\left(\frac{1(y_t > \mu)}{R(y_t;\theta)} - \frac{1(y_t > \mu)}{[R(y_t;\theta)]^2}\right)\right\} \\
&= \left[\frac{v_1+1}{2\alpha\sigma K(v_1)}\right]^2 \frac{\alpha}{v_1}\left[\frac{v_1}{v_1+1} - \frac{v_1(v_1+2)}{(v_1+1)(v_1+3)}\right] + \\
&\quad + \left[\frac{v_2+1}{2(1-\alpha)\sigma K(v_2)}\right]^2 \frac{1-\alpha}{v_2}\left[\frac{v_2}{v_2+1} - \frac{v_2(v_2+2)}{(v_2+1)(v_2+3)}\right] \\
&= \frac{1}{4\sigma^2}\left[\frac{v_1+1}{\alpha(v_1+3)}\frac{1}{K^2(v_1)} + \frac{v_2+1}{(1-\alpha)(v_2+3)}\frac{1}{K^2(v_2)}\right].
\end{aligned}
$$

$$
\begin{aligned}
E\left[\frac{\partial^2 \ln f}{\partial \mu^2}\right] &= \frac{1}{v_1}\frac{v_1+1}{[2\alpha\sigma K(v_1)]^2}E\left\{\left(\frac{1}{L(y_t;\theta)} - \frac{2}{[L(y_t;\theta)]^2}\right)1(y_t < \mu)\right\} \\
&\quad + \frac{1}{v_2}\frac{v_2+1}{[2(1-\alpha)\sigma K(v_2)]^2}E\left\{\frac{1(y_t > \mu)}{R(y_t;\theta)} - 2\frac{1(y_t > \mu)}{[R(y_t;\theta)]^2}\right\} \\
&= \frac{v_1+1}{[2\alpha\sigma K(v_1)]^2}\frac{\alpha}{v_1}\left[\frac{v_1}{v_1+1} - \frac{2v_1(v_1+2)}{(v_1+1)(v_1+3)}\right] \\
&\quad + \frac{v_2+1}{[2(1-\alpha)\sigma K(v_2)]^2}\frac{1-\alpha}{v_2}\left[\frac{v_2}{v_2+1} - \frac{2v_2(v_2+2)}{(v_2+1)(v_2+3)}\right] \\
&= -\frac{1}{4\sigma^2}\left[\frac{v_1+1}{\alpha(v_1+3)}\frac{1}{K^2(v_1)} + \frac{v_2+1}{(1-\alpha)(v_2+3)}\frac{1}{K^2(v_2)}\right].
\end{aligned}
$$

(**d**).

$$
\begin{aligned}
E\left[\frac{\partial \ln f}{\partial \sigma}\right]^2 &= 0 - \frac{1}{\sigma^2} + \left(\frac{v_1+1}{\sigma}\right)^2 E\left\{\left(1 - \frac{1}{L(y_t;\theta)}\right)^2 1(y_t < \mu)\right\} \\
&\quad + \left(\frac{v_2+1}{\sigma}\right)^2 E\left\{\left(1 - \frac{1}{R(y_t;\theta)}\right)^2 1(y_t > \mu)\right\} \\
&= -\frac{1}{\sigma^2} + \left(\frac{v_1+1}{\sigma}\right)^2 \alpha\left[1 - 2\frac{v_1}{v_1+1} + \frac{v_1(v_1+2)}{(v_1+1)(v_1+3)}\right] \\
&\quad + \left(\frac{v_2+1}{\sigma}\right)^2 (1-\alpha)\left[1 - 2\frac{v_2}{v_2+1} + \frac{v_2(v_2+2)}{(v_2+1)(v_2+3)}\right] \\
&= \frac{2}{\sigma^2}\left[\alpha\frac{v_1}{v_1+3} + (1-\alpha)\frac{v_2}{v_2+3}\right];
\end{aligned}
$$

$$E\left[\frac{\partial^2 \ln f}{\partial \sigma^2}\right] = \frac{1}{\sigma^2} - \frac{v_1+1}{\sigma^2}E\left\{\left(1 - \frac{1}{L(y_t;\theta)}\right)\left(1 + \frac{2}{L(y_t;\theta)}\right)1(y_t < \mu)\right\}$$

$$- \frac{v_2+1}{\sigma^2}E\left\{\left(1 - \frac{1}{R(y_t;\theta)}\right)\left(1 + \frac{2}{R(y_t;\theta)}\right)1(y_t > \mu)\right\}$$

$$= \frac{1}{\sigma^2} - \frac{v_1+1}{\sigma^2}\alpha\left[1 + \frac{v_1}{v_1+1} - 2\frac{v_1(v_1+2)}{(v_1+1)(v_1+3)}\right]$$

$$- \frac{v_2+1}{\sigma^2}(1-\alpha)\left[1 + \frac{v_2}{v_2+1} - 2\frac{v_2(v_2+2)}{(v_2+1)(v_2+3)}\right]$$

$$= -\frac{2}{\sigma^2}\left[\alpha\frac{v_1}{v_1+3} + (1-\alpha)\frac{v_2}{v_2+3}\right].$$

(**e**).

$$E\left[\frac{\partial \ln f}{\partial \alpha}\frac{\partial \ln f}{\partial v_1}\right] = -\frac{v_1+1}{2\alpha}E\left\{\left(1 - \frac{1}{L(y_t;\theta)}\right)\ln L(y_t;\theta)1(y_t < \mu)\right\}$$

$$+ \frac{(v_1+1)^2}{2\alpha}D(v_1)E\left\{\left(1 - \frac{1}{L(y_t;\theta)}\right)^2 1(y_t < \mu)\right\}$$

$$= -\frac{v_1+1}{2\alpha}\alpha\left[D(v_1) - \frac{v_1}{v_1+1}D(v_1+2)\right]$$

$$+ \frac{(v_1+1)^2}{2\alpha}D(v_1)\alpha\left[1 - \frac{2v_1}{v_1+1} + \frac{v_1(v_1+2)}{(v_1+1)(v_1+3)}\right]$$

$$= -\frac{1}{v_1+1} + \frac{v_1}{v_1+3}D(v_1);$$

$$E\left[\frac{\partial^2 \ln f}{\partial \alpha \partial v_1}\right] = \frac{1}{\alpha}E\left\{\left(1 - \frac{1}{L(y_t;\theta)}\right)1(y_t < \mu)\right\}$$

$$- \frac{v_1+1}{\alpha}D(v_1)E\left\{\left(\frac{1}{L(y_t;\theta)} - \frac{1}{[L(y_t;\theta)]^2}\right)1(y_t < \mu)\right\}$$

$$= \frac{1}{\alpha}\alpha\left[1 - \frac{v_1}{v_1+1}\right] - \frac{v_1+1}{\alpha}D(v_1)\frac{\alpha v_1}{v_1+1}\left[1 - \frac{v_1+2}{v_1+3}\right]$$

$$= \frac{1}{v_1+1} - \frac{v_1}{v_1+3}D(v_1).$$

(**f**).

$$E\left[\frac{\partial \ln f}{\partial \alpha}\frac{\partial \ln f}{\partial \mu}\right]$$

$$= \frac{(v_1+1)^2}{2\alpha}E\left\{\left(\frac{1}{L(y_t;\theta)}-\frac{1}{[L(y_t;\theta)]^2}\right)\frac{1}{v_1}\frac{2(y_t-\mu)}{[2\alpha\sigma K(v_1)]^2}1(y_t<\mu)\right\}$$

$$-\frac{(v_2+1)^2}{2(1-\alpha)}E\left\{\left(\frac{1}{R(y_t;\theta)}-\frac{1}{[R(y_t;\theta)]^2}\right)\frac{1}{v_2}\frac{2(y_t-\mu)1(y_t>\mu)}{[2(1-\alpha)\sigma K(v_2)]^2}\right\}$$

$$= \frac{(v_1+1)^2}{2\alpha\sigma}\left[-\frac{2}{v_1+1}+\frac{2}{v_1+3}\right]-\frac{(v_2+1)^2}{2(1-\alpha)\sigma}\left[\frac{2}{v_2+1}-\frac{2}{v_2+3}\right]$$

$$= -\frac{2}{\sigma}\left[\frac{v_1+1}{\alpha(v_1+3)}+\frac{v_2+1}{(1-\alpha)(v_2+3)}\right];$$

$$E\left[\frac{\partial^2 \ln f}{\partial\alpha\partial\mu}\right] = -\frac{v_1+1}{\alpha}E\left\{\frac{1}{[L(y_t;\theta)]^2}\frac{1}{v_1}\frac{2(y_t-\mu)}{[2\alpha\sigma K(v_1)]^2}1(y_t<\mu)\right\}$$

$$+\frac{v_2+1}{1-\alpha}E\left\{\frac{1}{[R(y_t;\theta)]^2}\frac{1}{v_2}\frac{2(y_t-\mu)}{[2(1-\alpha)\sigma K(v_2)]^2}1(y_t>\mu)\right\}$$

$$= \frac{v_1+1}{\alpha}\frac{1}{\sigma}\frac{2}{v_1+3}+\frac{v_2+1}{1-\alpha}\frac{1}{\sigma}\frac{2}{v_2+3}$$

$$= \frac{2}{\sigma}\left[\frac{v_1+1}{\alpha(v_1+3)}+\frac{v_2+1}{(1-\alpha)(v_2+3)}\right].$$

(**g**).

$$E\left[\frac{\partial \ln f}{\partial \alpha}\frac{\partial \ln f}{\partial \sigma}\right] = -\frac{1}{\sigma}E[\frac{\partial \ln f}{\partial \alpha}]+\frac{(v_1+1)^2}{\alpha\sigma}E\left\{\left(1-\frac{1}{L(y_t;\theta)}\right)^2 1(y_t<\mu)\right\}$$

$$-\frac{(v_2+1)^2}{(1-\alpha)\sigma}E\left\{\left(1-\frac{1}{R(y_t;\theta)}\right)^2 1(y_t>\mu)\right\}$$

$$= 0+\frac{(v_1+1)^2}{\alpha\sigma}\alpha\left[1-\frac{2v_1}{v_1+1}+\frac{v_1(v_1+2)}{(v_1+1)(v_1+3)}\right]$$

$$-\frac{(v_2+1)^2}{(1-\alpha)\sigma}(1-\alpha)\left[1-\frac{2v_2}{v_2+1}+\frac{v_2(v_2+2)}{(v_2+1)(v_2+3)}\right]$$

$$= \frac{2}{\sigma}\left[\frac{v_1}{v_1+3}-\frac{v_2}{v_2+3}\right];$$

$$E\left[\frac{\partial^2 \ln f}{\partial \alpha \partial \sigma}\right] = -2\left(\frac{v_1+1}{\alpha\sigma}\right)E\left\{\left(\frac{1}{L(y_t;\theta)} - \frac{1}{[L(y_t;\theta)]^2}\right)1(y_t < \mu)\right\}$$

$$+2\left[\frac{v_2+1}{(1-\alpha)\sigma}\right]E\left\{\left(\frac{1}{R(y_t;\theta)} - \frac{1}{[R(y_t;\theta)]^2}\right)1(y_t > \mu)\right\}$$

$$= -2\left(\frac{v_1+1}{\alpha\sigma}\right)\alpha\left[\frac{v_1}{v_1+1} - \frac{v_1(v_1+2)}{(v_1+1)(v_1+3)}\right]$$

$$+2\left(\frac{v_2+1}{(1-\alpha)\sigma}\right)(1-\alpha)\left[\frac{v_2}{v_2+1} - \frac{v_2(v_2+2)}{(v_2+1)(v_2+3)}\right]$$

$$= -\frac{2}{\sigma}\left(\frac{v_1}{v_1+3} - \frac{v_2}{v_2+3}\right).$$

(**h**). Note that $\frac{\partial \ln f}{\partial v_1}\frac{\partial \ln f}{\partial v_2} = 0$ and $\frac{\partial^2 \ln f}{\partial v_1 \partial v_2} = 0$. Then we have

$$E\left(\frac{\partial \ln f}{\partial v_1}\frac{\partial \ln f}{\partial v_2}\right) = -E\left(\frac{\partial^2 \ln f}{\partial v_1 \partial v_2}\right) = 0.$$

(**i**).

$$E\left[\frac{\partial \ln f}{\partial v_1}\frac{\partial \ln f}{\partial \mu}\right]$$

$$= -\frac{v_1+1}{4}E\left\{\frac{1}{L(y_t;\theta)}\frac{1}{v_1}\frac{2(y_t-\mu)}{[2\alpha\sigma K(v_1)]^2}\ln L(y_t;\theta)1(y_t < \mu)\right\}$$

$$+\left(\frac{v_1+1}{2}\right)^2 D(v_1)E\left\{\frac{1}{L(y_t;\theta)}\frac{1}{v_1}\frac{2(y_t-\mu)}{[2\alpha\sigma K(v_1)]^2}1(y_t < \mu)\right\}$$

$$-\left(\frac{v_1+1}{2}\right)^2 D(v_1)E\left\{\frac{1}{[L(y_t;\theta)]^2}\frac{1}{v_1}\frac{2(y_t-\mu)}{[2\alpha\sigma K(v_1)]^2}1(y_t < \mu)\right\}$$

$$= \frac{v_1+1}{4\sigma}\left(\frac{2}{v_1+1}\right)^2 - \left(\frac{v_1+1}{2}\right)^2\frac{D(v_1)}{\sigma}\left[\frac{2}{v_2+1} - \frac{2}{v_2+3}\right]$$

$$= \frac{1}{\sigma}\left[\frac{1}{v_1+1} - \frac{v_1+1}{v_1+3}D(v_1)\right];$$

$$E\left[\frac{\partial^2 \ln f}{\partial v_1 \partial \mu}\right] = \frac{1}{2}E\left\{\frac{1}{L(y_t;\theta)}\frac{1}{v_1}\frac{2(y_t-\mu)}{[2\alpha\sigma K(v_1)]^2}1(y_t < \mu)\right\}$$

$$-\left(\frac{v_1+1}{2}\right)D(v_1)E\left\{\frac{1}{[L(y_t;\theta)]^2}\frac{1}{v_1}\frac{2(y_t-\mu)1(y_t<\mu)}{[2\alpha\sigma K(v_1)]^2}\right\}$$

$$= -\frac{1}{2}\frac{1}{\sigma}\left(\frac{2}{v_1+1}\right) + \left(\frac{v_1+1}{2}\right)D(v_1)\frac{1}{\sigma}\frac{2}{v_2+3}$$

$$= -\frac{1}{\sigma}\left[\frac{1}{v_1+1} - \frac{v_1+1}{v_1+3}D(v_1)\right].$$

33

(**j**).

$$E\left[\frac{\partial \ln f}{\partial v_1}\frac{\partial \ln f}{\partial \sigma}\right] = 0 - \frac{v_1+1}{2\sigma}E\left\{\left(1-\frac{1}{L(y_t;\theta)}\right)\ln L(y_t;\theta)1(y_t<\mu)\right\}$$

$$+\frac{(v_1+1)^2}{2\sigma}D(v_1)E\left\{\left(1-\frac{1}{L(y_t;\theta)}\right)^2 1(y_t<\mu)\right\}$$

$$= -\frac{v_1+1}{2\sigma}\alpha\left[D(v_1)-\frac{v_1}{v_1+1}D(v_1+2)\right]$$

$$+\frac{(v_1+1)^2}{2\sigma}D(v_1)\alpha\left[1-2\frac{v_1}{v_1+1}+\frac{v_1(v_1+2)}{(v_1+1)(v_1+3)}\right]$$

$$= \frac{\alpha}{\sigma}\left[-\frac{1}{v_1+1}+\frac{v_1}{v_1+3}D(v_1)\right];$$

$$E\left[\frac{\partial^2 \ln f}{\partial v_1\partial\sigma}\right] = \frac{1}{\sigma}E\left\{\left(1-\frac{1}{L(y_t;\theta)}\right)1(y_t<\mu)\right\}$$

$$-\frac{v_1+1}{\sigma}D(v_1)E\left\{\left(\frac{1}{L(y_t;\theta)}-\frac{1}{[L(y_t;\theta)]^2}\right)1(y_t<\mu)\right\}$$

$$= \frac{1}{\sigma}\alpha\left[1-\frac{v_1}{v_1+1}\right]-\frac{v_1+1}{\sigma}D(v_1)\alpha\frac{v_1}{v_1+1}\left[1-\frac{v_1+2}{v_1+3}\right]$$

$$= \frac{\alpha}{\sigma}\left[\frac{1}{v_1+1}-\frac{v_1}{v_1+3}D(v_1)\right].$$

(**k**).

$$E\left[\frac{\partial \ln f}{\partial\mu}\frac{\partial \ln f}{\partial\sigma}\right]$$

$$= \frac{(v_1+1)^2}{2\sigma}E\left\{\left(\frac{1}{L(y_t;\theta)}-\frac{1}{[L(y_t;\theta)]^2}\right)\frac{1}{v_1}\frac{2(y_t-\mu)}{[2\alpha\sigma K(v_1)]^2}1(y_t<\mu)\right\}$$

$$+\frac{(v_2+1)^2}{2\sigma}E\left\{\left(\frac{1}{R(y_t;\theta)}-\frac{1}{[R(y_t;\theta)]^2}\right)\frac{1}{v_2}\frac{2(y_t-\mu)1(y_t>\mu)}{[2(1-\alpha)\sigma K(v_1)]^2}\right\}$$

$$= \frac{(v_1+1)^2}{2\sigma^2}\left[-\frac{2}{v_1+1}+\frac{2}{v_1+3}\right]+\frac{(v_2+1)^2}{2\sigma^2}\left[\frac{2}{v_2+1}-\frac{2}{v_2+3}\right]$$

$$= \frac{2}{\sigma^2}\left[-\frac{v_1+1}{v_1+3}+\frac{v_2+1}{v_2+3}\right];$$

$$E\left[\frac{\partial^2 \ln f}{\partial \mu \partial \sigma}\right] = -\frac{v_1+1}{\sigma}E\left\{\frac{1}{[L(y_t;\theta)]^2}\frac{1}{v_1}\frac{2(y_t-\mu)}{[2\alpha\sigma K(v_1)]^2}1(y_t<\mu)\right\}$$

$$-\frac{v_2+1}{\sigma}E\left\{\frac{1}{[R(y_t;\theta)]^2}\frac{1}{v_2}\frac{2(y_t-\mu)}{[2(1-\alpha)\sigma K(v_1)]^2}1(y_t>\mu)\right\}$$

$$= \left(\frac{v_1+1}{\sigma}\right)\frac{1}{\sigma}\frac{2}{v_1+3} - \frac{v_2+1}{\sigma}\frac{1}{\sigma}\frac{2}{v_2+3}$$

$$= \frac{2}{\sigma^2}\left[\frac{v_1+1}{v_1+3} - \frac{v_2+1}{v_2+3}\right].$$

(1). By the symmetry property of v_1 and v_2, similarly, we have

$$E\left[\frac{\partial \ln f}{\partial v_2}\right]^2 = -E\left[\frac{\partial^2 \ln f}{\partial v_2^2}\right] = \frac{1-\alpha}{2}\left\{\frac{v_2 D^2(v_2)}{v_2+3} - \frac{2D(v_2)}{v_2+1} - D'(v_2)\right\};$$

$$E\left[\frac{\partial \ln f}{\partial v_2}\frac{\partial \ln f}{\partial \alpha}\right] = -E\left[\frac{\partial^2 \ln f}{\partial v_2 \partial \alpha}\right] = \frac{1}{v_2+1} - \frac{v_2}{v_2+3}D(v_2);$$

$$E\left[\frac{\partial \ln f}{\partial v_2}\frac{\partial \ln f}{\partial \mu}\right] = -E\left[\frac{\partial^2 \ln f}{\partial v_2 \partial \mu}\right] = -\frac{1}{\sigma}\left[\frac{1}{v_2+1} - \frac{v_2+1}{v_2+3}D(v_2)\right];$$

$$E\left[\frac{\partial \ln f}{\partial v_2}\frac{\partial \ln f}{\partial \sigma}\right] = -E\left[\frac{\partial^2 \ln f}{\partial v_2 \partial \sigma}\right] = \frac{1-\alpha}{\sigma}\left[-\frac{1}{v_2+1} + \frac{v_2 D(v_2)}{v_2+3}\right].\blacksquare$$

Table 1a
Means of simulated MLE's of AST parameters
$T = 1000,\ 5000;\ 10000$ replications; $\sigma = 1,\ \mu = 0$

$T:$	$\alpha, v_1, v_2:$	mean $\hat\alpha$	mean $\hat v_1$	mean $\hat v_2$	mean $\hat\sigma$	mean $\hat\mu$
1000	0.3,0.7,2.5	0.300	0.709	2.547	1.001	6×10^{-4}
5000	0.3,0.7,2.5	0.300	0.701	2.511	1.000	3×10^{-5}
1000	0.3,2.5,2.5	0.301	2.668	2.543	1.001	0.002
5000	0.3,2.5,2.5	0.300	2.534	2.510	1.000	5×10^{-4}
1000	0.3,2.0,5.0	0.301	2.083	5.311	1.001	0.002
5000	0.3,2.0,5.0	0.300	2.018	5.051	1.000	6×10^{-4}
1000	0.8,0.7,2.5	0.798	0.702	2.806	1.002	-0.002
5000	0.8,0.7,2.5	0.800	0.700	2.559	1.000	-3×10^{-4}
1000	0.8,2.5,2.5	0.798	2.533	2.826	1.001	-0.003
5000	0.8,2.5,2.5	0.800	2.506	2.543	1.000	-4×10^{-4}
1000	0.8,2.0,5.0	0.798	2.021	10.93	1.001	-0.003
5000	0.8,2.0,5.0	0.800	2.003	5.282	1.000	-7×10^{-4}

Table 1b
Simulation standard errors of MLE's of AST parameters
$T = 1000,\ 5000;\ 10000$ replications; $\sigma = 1,\ \mu = 0$

$T:$	$\alpha, v_1, v_2:$	se$(\hat\alpha)$	se$(\hat v_1)$	se$(\hat v_2)$	se$(\hat\sigma)$	se$(\hat\mu)$
1000	0.3,0.7,2.5	0.022	0.078	0.323	0.042	0.025
5000	0.3,0.7,2.5	0.010	0.033	0.133	0.019	0.011
1000	0.3,2.5,2.5	0.030	0.688	0.334	0.038	0.035
5000	0.3,2.5,2.5	0.013	0.244	0.139	0.017	0.015
1000	0.3,2.0,5.0	0.029	0.435	1.364	0.036	0.035
5000	0.3,2.0,5.0	0.013	0.171	0.471	0.016	0.015
1000	0.8,0.7,2.5	0.023	0.042	1.905	0.051	0.027
5000	0.8,0.7,2.5	0.010	0.019	1.435	0.023	0.012
1000	0.8,2.5,2.5	0.026	0.301	2.535	0.038	0.031
5000	0.8,2.5,2.5	0.011	0.128	0.313	0.017	0.013
1000	0.8,2.0,5.0	0.028	0.207	26.90	0.038	0.034
5000	0.8,2.0,5.0	0.012	0.089	1.228	0.017	0.015

Figure 1
Ratio of simulation standard error to Cramèr-Rao bound [3]
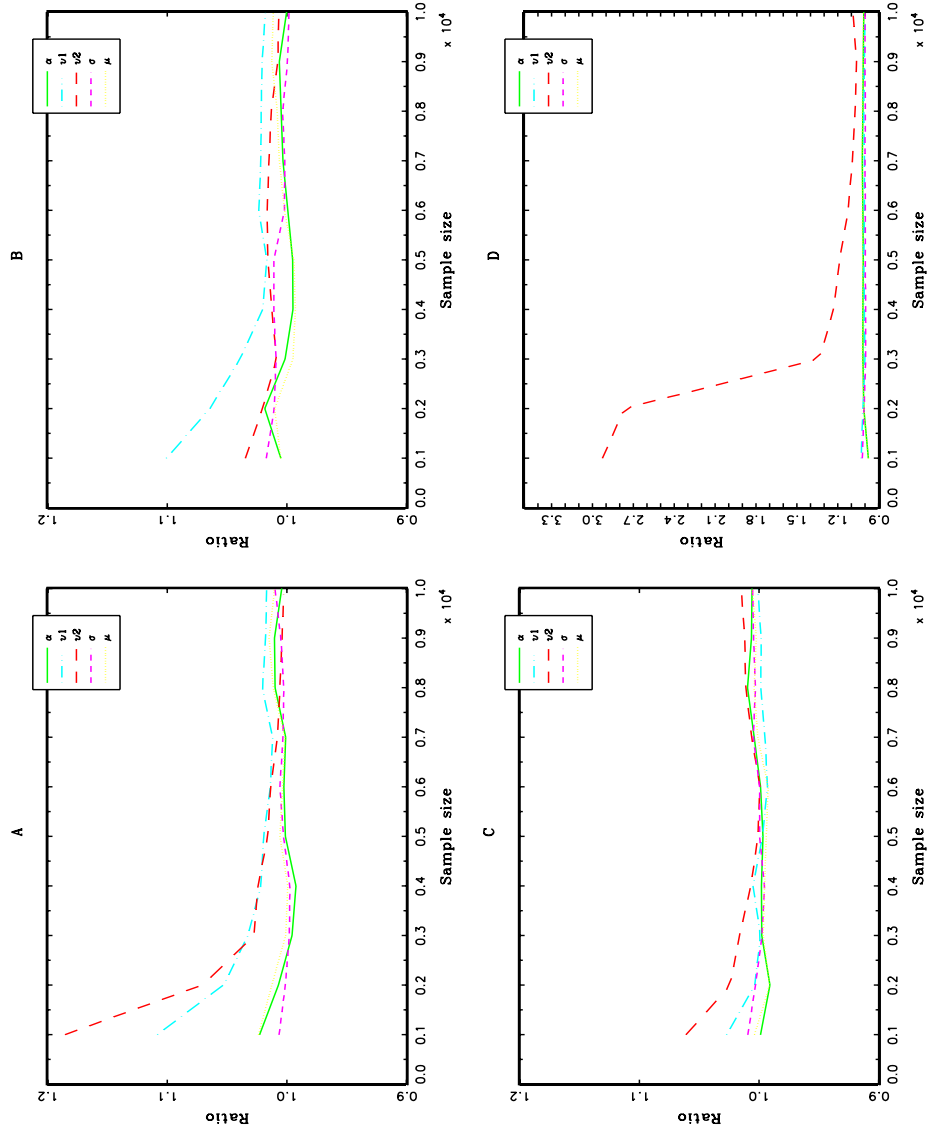
[3]Case A: $\alpha = 0.3, v_1 = 2.0, v_2 = 5.0, \sigma = 1, \mu = 0$; Case B: $\alpha = 0.3, v_1 = 2.0, v_2 = 2.0, \sigma = 1, \mu = 0$; Case C: $\alpha = 0.3, v_1 = 0.7, v_2 = 2.5, \sigma = 1, \mu = 0$; Case D: $\alpha = 0.8, v_1 = 2.0, v_2 = 5.0, \sigma = 1, \mu = 0$.

37