Skew-t Information Matrix: Evaluation and Use

Chindhanai Uthaisaad

A thesis submitted in partial fulfillment of the

requirements for the degree of

Master of Science

University of Washington

2018

Committee:

Douglas Martin

Aleksandr Aravkin

Program Authorized to Offer Degree:

Applied Math

University of Washington

Abstract

Skew-t Information Matrix: Evaluation and Use

Chindhanai Uthaisaad

Chair of the Supervisory Committee:

Professor Douglas R. Martin

Department of Applied Mathematics

Azzalini's skew-t distributions, described in detail in Azzalini [2013], have become very popular because of their practical usefulness and the complete `R` package `sn` for skew-normal distributions that include extensive support for fitting and analysis of skew-t distributions. A major difficulty is that the skew-t distribution expected information matrix has no analytical form. This thesis develops an `R` package `skewtInfo` to compute the expected information matrix through numerical integrations and investigates the accuracy of the resulting matrix and its use for computing finite-sample approximations for skew-t parameter estimates.

1 Introduction

The main motivation for this Master's Thesis study is the well-known fact that stock returns have non-normal distributions that exhibit skewness - a measure of asymmetry in the data distribution - and excess kurtosis to various degrees, where excess kurtosis is associated with fat-tailed deviations of the returns distributions from normality. See for example Amaya et al. [2015] and Masud [2012] for further analysis on skewness and kurtosis of stock returns.

Symmetric t-distributions are often used to model stock returns . While capturing the fat-tailedness of the data in the polynomial form, the symmetric t-distributions do not have a parameter that governs the skewness. Therefore we need fat-tailed skewed distributions to take care of both skewness and kurtosis as evidenced in the stock returns.

There exist several alternative definitions of skewed t-distributions and we briefly summarize these in Appendix A. With a comprehensive `R` package `sn` that supports a wide range of its fitting and analysis, Azzalini's skew-t distribution, introduced by Azzalini and Capitanio [2009] and thoroughly discussed in the book by Azzalini [2013], is a particularly attractive choice for our purposes.

Nonetheless the main problem arisen from the Azzalini's skew-t distribution due to its complication in its Hessian is that it has no closed form formula for the expected Fisher information matrix. We therefore evaluate such a matrix numerically via numerical integrations. This approach provides us with a tool to compute finite sample standard errors (S.E.'s) of skew-t MPLE's using their asymptotic variance expression based on parameter estimates. We also envision using the result to: (a) study the inefficiency of skew-t MPLE's when the true distribution is a symmetric-t, and (b) study the behavior of expected shortfall (ES) based on skew-t distributions.

The remainder of this paper is organized as follows. Section 2 introduces the Azzalini's skew-t distribution family, and the relationship between the skew-t distribution parameters and the usual moment-based mean, standard deviation, skewness and excess kurtosis. Section 3 explains the method of skew-t distribution maximum penalized likelihood estimation (MPLE) used to fit the data. Section 4 defines the expected information matrix of the skew-t distribution and discusses the use of the numerical integration to compute the expected information matrix. We then show a Monte Carlo study of the information matrix based on finite-sample standard errors and the inferences of the MPLE's. We finally wrap up with the discussion conclusions and remarks for the future research in Section 5.

2 Azzalini Skew-t Distributions

It is well-known that a random variable X that has a (symmetric) t-distribution with mean zero, unit variance and ν degrees of freedom, has the stochastic representation

$$X = \frac{Z}{\sqrt{\chi^2_\nu/\nu}} \tag{1}$$

where Z has a standard normal distribution and χ^2_ν has a chi-squared distribution with ν degrees of freedom. Azzalini obtained a standard skew-t random variable $Y_{0,1,\,\alpha,\,\nu} \sim \mathbf{ST}(0,1,\ \alpha,\ \nu)$ with slant (skewness) parameter α and degrees of freedom ν by replacing Z with a standard skew-normal random variable $Z_{0,1,\alpha} \sim \mathbf{SN}(0,1,\alpha)$, whose probability density function is given by

$$f_{SN}(x;\alpha) = 2\ \phi(x) \cdot \Phi(\alpha x). \tag{2}$$

The resulting standard skew-t random variable

$$Y_{0,1,\alpha,\nu} = \frac{Z_{0,1,\alpha}}{\sqrt{\chi^2_\nu/\nu}} \tag{3}$$

has the probability density

$$f_{ST}(x;\alpha,\nu) = 2\, t_\nu(x) \cdot T_{\nu+1}\left(\alpha x \cdot \sqrt{\frac{\nu+1}{\nu+x^2}}\right) \tag{4}$$

where $t_\nu(x)$ is a standard t-distribution probability density function with degrees of freedom ν, and $T_\nu(x)$ is the corresponding cumulative distribution function with degrees of freedom ν.

A skew-t random variable $Y_{\xi,\omega,\alpha,\nu} \sim \mathbf{ST}(\xi,\omega^2,\alpha,\nu)$ with location parameter ξ and scale parameter ω has the representation

$$Y_{\xi,\,\omega,\,\alpha,\,\nu} = \xi + \omega Y_{0,1,\,\alpha,\,\nu}. \tag{5}$$

With

$$z = \frac{y-\xi}{\omega}, \qquad \tau(z) = \sqrt{\frac{\nu+1}{\nu+z^2}} \tag{6}$$

the probability density function of $Y_{\xi,\,\omega,\,\alpha,\,\nu}$ is then given by

$$f(y;\xi,\omega^2,\alpha,\nu) = \frac{2}{\omega} \cdot t(z) \cdot T_{\nu+1}\left(\alpha z \tau(z)\right) \tag{7}$$

Let

$$\delta = \frac{\alpha}{\sqrt{1+\alpha^2}}, \qquad b_\nu = \sqrt{\frac{\nu}{\pi}} \cdot \frac{\Gamma((\nu-1)/2)}{\Gamma(\nu/2)}, \qquad \sigma_z^2 = \mathbf{var}(z) = \frac{\nu}{\nu-2} - (b_\nu\delta)^2. \tag{8}$$

Then the values of the the mean μ, standard deviation σ, skewness γ, and excess kurtosis κ

are given in terms of the parameters of general skew-t distribution as follows[1]

$$
\begin{aligned}
\mu &= \xi + \omega b_\nu \delta \\
\sigma^2 &= \omega^2 \sigma_z^2 \quad \text{for } \nu > 2 \\
\gamma &= \frac{b_\nu \delta}{\sigma_z^{3/2}} \left[\frac{\nu(3-\delta^2)}{\nu-3} - \frac{3\nu}{\nu-2} + 2(b_\nu\delta)^2 \right] \quad \text{for } \nu > 3 \\
\kappa &= \frac{1}{\sigma_z^4} \left[\frac{3\nu^2}{(\nu-2)(\nu-4)} - \frac{4(b_\nu\delta)^2\nu(3-\delta^2)}{\nu-3} + \frac{6\nu(b_\nu\delta)^2}{\nu-2} - 3(b_\nu\delta)^4 \right] - 3 \quad \text{for } \nu > 4.
\end{aligned}
\tag{9}
$$

3 Fitting Skew-t Distributions with the MPLE Method

The values of the skewness parameter α can get very large as the skewness is farther from zero. In order to overcome such an occurrence, Azzalini and Arellano-Valle [2012], hence AAV for short, showed that, with non-negligible probability, the MLE of the skewness parameter α diverges, and proposed a maximum penalized likelihood estimator (MPLE) to solve this problem. Please find how the penalization works in detail in Appendix A.

Let $\boldsymbol{\theta} = (\xi,\ \omega,\ \alpha,\ \nu)$ and consider the skew-t $\mathbf{ST}(\xi,\ \omega^2,\ \alpha,\ \nu)$ log-likelihood function:

$$
\begin{aligned}
\ell(\boldsymbol{\theta};\ y) = \text{constant} &- \frac{1}{2}\log\nu + \log\Gamma(\frac{1}{2}(\nu+1)) - \log\Gamma(\frac{\nu}{2}) \\
&- \frac{1}{2}(\nu+1)\log\left(1+\frac{z^2}{\nu}\right) + \log T_{\nu+1}\left(\alpha z \tau(z)\right).
\end{aligned}
\tag{10}
$$

Let $\hat{\boldsymbol{\theta}} = (\hat{\xi},\ \hat{\omega},\ \hat{\alpha},\ \hat{\nu})$ denote a maximum-likelihood estimate (MLE) for an implicit finite sample size n.

Consider the penalized log-likelihood (PL) function

$$
\ell_p(\boldsymbol{\theta}) = \ell(\boldsymbol{\theta}) - Q(\boldsymbol{\theta}) \tag{11}
$$

[1]The skewness and excess kurtosis are given by $\mathbf{E}\left[\left(\frac{Y-\mu}{\sigma}\right)^3\right]$ and $\mathbf{E}\left[\left(\frac{Y-\mu}{\sigma}\right)^4\right] - 3$, respectively.

where $Q(\boldsymbol{\theta}) = Q(\alpha, \nu)$ represents a nonnegative quantity penalizing the divergence of α by virtue of satisfying the following conditions for each fixed positive and finite ν:

$$Q(\boldsymbol{\theta}) \geq 0, \quad Q(0, \nu) = 0, \quad \lim_{\alpha \to \pm\infty} Q(\alpha, \nu) = \infty. \tag{12}$$

Section 3 of AAV introduces a formula for $Q(\alpha, \nu)$ that depends on the complicated equations (23)-(24) therein. The equation (23) requires integration for each value of the degrees of freedom parameter ν, which greatly complicates numerical maximization of the PL function. Consequently AAV derive an approximation for $Q(\alpha, \nu)$ that does not require numerical integration for each ν. Evidently this is the version of the penalty function used in the `sn` package function `st.mple` that we use throughout the remainder of this paper to fit skew-t distributions.

Let $\tilde{\boldsymbol{\theta}} = (\tilde{\xi},\ \tilde{\omega},\ \tilde{\alpha},\ \tilde{\nu})$ denote a maximum-penalized likelihood estimate (MPLE) using the above penalty function, for an implicit sample size n. For purposes of the next section, we assume that a skew-t distribution MPLE converges to the same limit distribution as a skew-t distribution MLE. The justification for this assumption is the argument in Section 2.1 of AAV in support of the equivalence of the asymptotic distribution of an MLE and an MPLE for the case of a skewed normal distribution, and that the argument also applies in the case of a skew-t distribution.

A Stock Returns Example

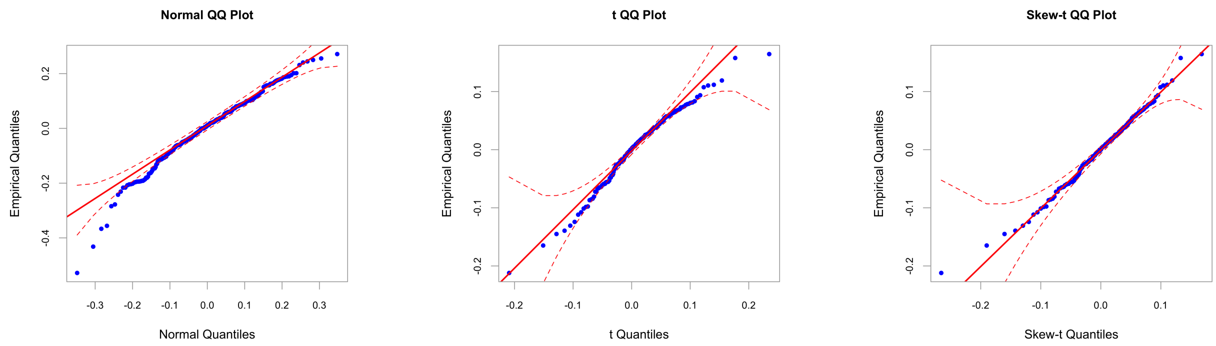
Martin and Uthaisaad (2017) studied the distributions of skew-t MPLE parameters for the monthly returns of a cross section of 300 stocks from December 1991 to September 2015. They found that the median values of the skew-t parameters (to 2 significant digits) are $\xi = 0.03,\ \omega = 0.08,\ \alpha = -0.36,\ \nu = 5.50$. The following code uses the `sn` package function `st.mple` to compute a skew-t distribution MPLE fit to the stock in that group that has the ticker D.

```
# Simulate the returns parametrically using the
# median of the parameters
library(sn)
returns <- largecap[, "D"]
fitD <- st.mple(y = returns, penalty = "Qpenalty")

# Show the parameter estimates
fitD$dp
```

```
##          xi       omega       alpha          nu
##  0.03976989  0.04384282 -1.05281814  4.24263265
```

Figure 1a, 1b, and 1c show a normal qqplot with $\mu = 0.009$, $\sigma = 0.049$, a symmetric-t qqplot with parameter estimates $\mu = 0.013$, $s = 0.036$, $\nu = 4.07$, and a skew-t qqplot with MPLE parameter estimates $\xi = 0.040$, $\omega = 0.044$, $\gamma = -1.05$, $\nu = 4.24$ corresponding to the D returns. The normal qqplot reveals that the returns distribution is clearly long-tailed to the left and perhaps very slightly short-tailed to the right. The skew-t qqplot reveals that the skew-t MPLE fit is quite good, and it better fits the D returns compared to symmetric-t distribution.



(a) (b) (c)

Figure 1: (a) Normal qqplot (b) t qqplot (c) Skew-t qqplot

Figure 2 shows kernel density estimate (KDE) of the D returns, overlaid with the density function of the fitted skew-t distribution with parameters $\xi = 0.09, \ \omega = 0.13, \ \gamma = -1.30, \ \nu = 10.24$[2].

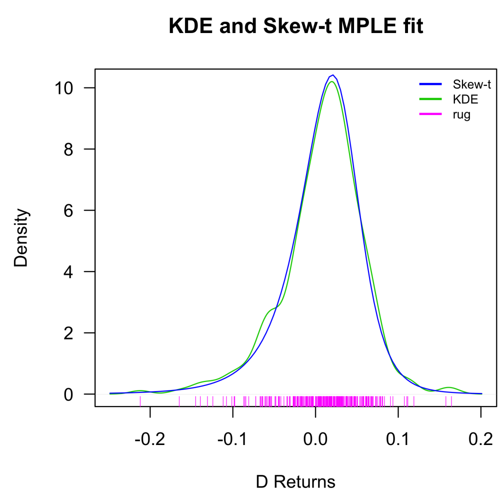


Figure 2: Skew-t fit on the D returns compared to its KDE

Skew-t Parametric vs Non-Parametric Skewness and Kurtosis We will compared the skew-t parametric and non-parametric skewness and kurtosis together with the mean

[2]It is always a good idea to make such a plot to confirm that an MPLE fit is a good one, i.e. it is not due to a local maximum of the penalized likelihood that is far from a global maximum.

and standard deviation. The mean, volatility, skewness and excess kurtosis values of the D returns are obtained from the skew-t parameter estimates via the equations (9). Table 1 compares skew-t parametric and non-parametric mean, volatility, skewness and excess kurtosis of the D return.

	Parametric	Nonparametric
μ	0.009	0.008
σ	0.049	0.052
γ	-0.737	-2.12
κ	2.17	42.56

Table 1: Parametric vs Nonparametric moment based estimates

We can see that the MPLE estimates the mean and volatility pretty accurately, but it overestimates the skewness and dramatically overestimates excess kurtosis.

4 The Skew-t Information Matrix

It is well-known that there are two equivalent forms of an information matrix, one of which involves the product of score functions that we define below, and the other of which involves the partial derivatives of the score functions (second partial derivatives of the log-likelihood), see Casella and Berger [2001]. Ciccio and Monti [2011] provided analytic expression for both forms, the first which is provided just below, and the second is provided in Appendix B. The information matrix in the form of the product of score functions is much more viable for the purposes of the numerical integration required, and its form with $\boldsymbol{\theta} = (\xi,\ \omega,\ \alpha,\ \nu)$ is

$$\mathbf{I}_{\boldsymbol{\theta}} = \begin{bmatrix} \mathbb{E}[S_\xi(y)S_\xi(y)] & \mathbb{E}[S_\xi(y)S_\omega(y)] & \mathbb{E}[S_\xi(y)S_\alpha(y)] & \mathbb{E}[S_\xi(y)S_\nu(y)] \\ \mathbb{E}[S_\xi(y)S_\omega(y)] & \mathbb{E}[S_\omega(y)S_\omega(y)] & \mathbb{E}[S_\omega(y)S_\alpha(y)] & \mathbb{E}[S_\omega(y)S_\nu(y)] \\ \mathbb{E}[S_\xi(y)S_\alpha(y)] & \mathbb{E}[S_\omega(y)S_\alpha(y)] & \mathbb{E}[S_\alpha(y)S_\alpha(y)] & \mathbb{E}[S_\alpha(y)S_\nu(y)] \\ \mathbb{E}[S_\xi(y)S_\nu(y)] & \mathbb{E}[S_\omega(y)S_\nu(y)] & \mathbb{E}[S_\alpha(y)S_\nu(y)] & \mathbb{E}[S_\nu(y)S_\nu(y)] \end{bmatrix} \tag{13}$$

where the score functions $S_\xi(y),\ S_\omega(y),\ S_\alpha(y),\ S_\nu(y)$ are defined in Section 4.1.

The importance of an information matrix $\mathbf{I}_{\boldsymbol{\theta}}$ is that its inverse is the asymptotic covariance matrix $\mathbf{V}_{\boldsymbol{\theta}} = \mathbf{I}_{\boldsymbol{\theta}}^{-1}$ of a consistent and asymptotically normal maximum-likelihood estimator (MLE). As discussed in the previous section, $\mathbf{V}_{\boldsymbol{\theta}} = \mathbf{I}_{\boldsymbol{\theta}}^{-1}$ will also be the asymptotic covariance matrix of a skew-t MPLE $\tilde{\theta} = \left(\tilde{\xi},\, \tilde{\omega},\, \tilde{\alpha},\, \tilde{\nu}\right)$. If $\mathbf{I}_{\boldsymbol{\theta}}$ is known, then $\mathbf{V}_{\boldsymbol{\theta}}$ can be used to compute finite-sample standard errors in the usual way[3]. In this section we focus on a method of evaluating $\mathbf{I}_{\boldsymbol{\theta}}$, and in the next section we focus on computing finite-sample standard errors based on $\mathbf{I}_{\tilde{\boldsymbol{\theta}}}$.

4.1 Skew-t Vector Score Function and its Components

The score function for a single skew-t return y as derived by Ciccio and Monti [2011] is of the form

$$S(y; \xi, \omega, \alpha, \nu) = \nabla \ell(\boldsymbol{\theta};\ y) = [S_\xi(y)\ S_\omega(y)\ S_\alpha(y)\ S_\nu(y)]'.$$

Let

$$w(z) = \frac{t_{\nu+1}(\alpha z \cdot \tau(z))}{T_{\nu+1}(\alpha z \cdot \tau(z))}, \qquad \psi(x) = \frac{\partial}{\partial x} \log(\Gamma(x))$$

where $\psi(x)$ is the digamma function. Then the components of the score function are the following derivatives of the skew-t log-likelihood given in (10) with respect to the individual

[3]The finite-sample S.E. approximation is obtained by extracting the diagonal elements of $\mathbf{V}_{\tilde{\boldsymbol{\theta}}}$, taking the square root of the result and then dividing by $\sqrt{n}$.

parameters:

$$S_\xi(y) = \frac{z \cdot \tau(z)^2}{\omega} - \frac{\alpha\nu \cdot \tau(z)}{\omega(\nu + z^2)} \cdot w(z) \tag{14}$$

$$S_\omega(y) = -\frac{1}{\omega} + \frac{z^2 \cdot \tau(z)^2}{\omega} - \frac{\alpha z\nu \cdot \tau(z)}{\omega(\nu + z^2)} \cdot w(z) \tag{15}$$

$$S_\alpha(y) = z \cdot w(z) \cdot \tau(z) \tag{16}$$

$$S_\nu(y) = \varphi(\nu) - \log\left(1 + \frac{z^2}{\nu}\right) + \frac{z^2 \cdot \tau(z)^2}{\nu} + \frac{\alpha z(z^2 - 1)}{(\nu + z^2)^2 \tau(z)} \cdot w(z) + \frac{\zeta(z)}{T_{\nu+1}(\alpha z \cdot \tau(z))} \tag{17}$$

where

$$\varphi(\nu) = \frac{1}{2}\left[\psi\left(\frac{\nu}{2} + 1\right) - \psi\left(\frac{\nu}{2}\right)\right] - \frac{2\nu + 1}{\nu(\nu + 1)} \tag{18}$$

$$\zeta(z) = \int_{-\infty}^{\alpha z \cdot \tau(z)} \left\{\frac{(\nu + 1)u^2}{(\nu + 1)(\nu + 1 + u^2)} - \log\left(1 + \frac{u^2}{\nu + 1}\right)\right\} t(u; \nu + 1)\, du. \tag{19}$$

In this case, we implement a function to compute the expected information matrix.

4.2 Evaluation of the Skew-t Information Matrix

Due to its complexity, there is no closed form formula for the information matrix of skew-t distribution. We have developed the package `skewtInfo`, located at the Github site https://github.com/chindhanai/skewtInfo, for numerically evaluating Equation (13). The package contains a function `stInfoMat` with the argument `type` indicating the type of required information matrix - the choice `type = ‘‘expected’’` is for computing an expected information matrix, while `type = ‘‘observed’’` is for computing a finite-sample observed information matrix and is described and used in Section 4.3.

The package contains about 400 lines of `R` code in total and uses numerical quadrature as implemented in the function `integrate` from a fundamental `R` package `base`, with the choice of tolerance of 10^{-9}. Technically, the `integrate` function maps the infinite range of parameters onto finite subintervals, and then uses globally adaptive interval subdivision in conjunction with Wynn's Epsilon algorithm extrapolation, with the basic step being Gauss-Kronrod quadrature. Please see Piessens et al. [1983] and the documentation of the function `integrate` for more information.

Note that the information matrix computation involves the double integrations over general regions. Indeed, they appear in expectations of the pairwise products of (17) with (14)-(16). For example, as a part of the expectation of $S_{\alpha\nu}$, we need to evaluate[4]

$$\begin{aligned}\mathbf{E}\left(\frac{zw\tau\zeta}{T_{\nu+1}(\alpha z\tau)}\right) &= \int\left(\frac{zw\tau\zeta}{T_{\nu+1}(\alpha z\tau)}\right) f_{ST}(y)\,dy \\ &= \int_{\mathbb{R}}\left(\int_{-\infty}^{\alpha z\tau_z}\left\{\frac{(\nu+1)u^2}{(\nu+1)(\nu+1+u^2)} - \log\left(1+\frac{u^2}{\nu+1}\right)\right\} t(u;\nu+1)\,du\right)\frac{zw\tau\zeta\cdot f_{ST}(y)}{T_{\nu+1}(\alpha z\tau)}\,dy\end{aligned}$$

To compute such an integral in `R`, we do the following

1. Define the innermost integrand of 19 as a function of u and ν
2. Integrate the innermost integrand with respect to u and define this inner integral as a function of z and ν
3. Define the inner integrand to be the product of integral from Step 2 and $\frac{z\cdot w(z)\cdot\tau(z)\cdot f_{ST}(y)}{T_{\nu+1}(\alpha z\cdot\tau(z))}$
4. Integrate the vectorized integral in Step 3 with respect to y by using `sapply` on the y range over the real line

[4]We suppress the variable z dependency for w, τ and ζ for a nicer expression of the integrals.

Usage Example

We used the median skew-t parameter values[5] from the Martin and Uthaisaad (2017) study, mentioned in the example at the end of Section 3, namely

$$\boldsymbol{\theta}_{med} = (0.03,\ 0.08, -0.36,\ 5.50) \tag{20}$$

to compute the information matrix (13) with the function `stInfoMat` in the `skewTInfo` package. The code and results are shown below.

```
library(devtools)
install_github("chindhanai/skewtInfo")
library(skewtInfo)
med = c(0.03018407, 0.07550424, -0.36373528, 5.49905975)
stInfoMat(dp = med, type = "expected")$stInfoMat
```

```
##              [,1]        [,2]         [,3]         [,4]
## [1,] 144.65797434 -28.4160660  7.945534658  0.027423845
## [2,] -28.41606604 232.6453731  1.610349244 -0.490044089
## [3,]   7.94553466   1.6103492  0.501303409 -0.004357227
## [4,]   0.02742385  -0.4900441 -0.004357227  0.002231010
```

Thus, rounding the numbers for convenience, the information matrix is:

$$\mathbf{I}_{\boldsymbol{\theta}_{med}} = \begin{bmatrix} 145 & -28.4 & 7.95 & 0.0274 \\ -28.4 & 232.6 & 1.61 & -0.490 \\ 7.95 & 1.61 & 0.501 & -0.00436 \\ 0.0274 & -0.490 & -0.00436 & 0.00223 \end{bmatrix}. \tag{21}$$

The corresponding eigenvalues are 252, 119, 0.0200, 0.00110, which show that the matrix is not well-conditional, but nonetheless indicate that $\mathbf{I}_{\boldsymbol{\theta}_{med}}$ is positive definite and its inverse exists.

[5] In practice, we will use the skew-t MPLE's, rather than the set of median values (20), to compute the information matrix.

4.3 Expected Information Matrix Accuracy Verification

Recall that the log-likelihood function corresponding to independently and identically distributed random variables $\mathbf{y} = (y_1,\ y_2, \dots)$ is given by

$$\ell_n(\boldsymbol{\theta}; \mathbf{y}) = \sum_{i=1}^{n} \log\ f_{\boldsymbol{\theta}}(y_i)$$

where $f_{\boldsymbol{\theta}}$ is the probability density function corresponding to $\boldsymbol{\theta}$. With

$$\nabla^2 \ell_n(\boldsymbol{\theta}; \mathbf{y}) = \begin{bmatrix} \sum_{i=1}^n S_{\xi\xi}(y_i) & \sum_{i=1}^n S_{\xi\omega}(y_i) & \sum_{i=1}^n S_{\xi\alpha}(y_i) & \sum_{i=1}^n S_{\xi\nu}(y_i) \\ \sum_{i=1}^n S_{\xi\omega}(y_i) & \sum_{i=1}^n S_{\omega\omega}(y_i) & \sum_{i=1}^n S_{\omega\alpha}(y_i) & \sum_{i=1}^n S_{\omega\nu}(y_i) \\ \sum_{i=1}^n S_{\xi\alpha}(y_i) & \sum_{i=1}^n S_{\omega\alpha}(y_i) & \sum_{i=1}^n S_{\alpha\alpha}(y_i) & \sum_{i=1}^n S_{\alpha\nu}(y_i) \\ \sum_{i=1}^n S_{\xi\nu}(y_i) & \sum_{i=1}^n S_{\omega\nu}(y_i) & \sum_{i=1}^n S_{\alpha\nu}(y_i) & \sum_{i=1}^n S_{\nu\nu}(y_i) \end{bmatrix} \tag{22}$$

and by the law of large numbers

$$-\frac{1}{n}\nabla^2 \ell_n(\boldsymbol{\theta}; \mathbf{y}) \rightarrow^P \mathbf{I}_{\boldsymbol{\theta}}. \tag{23}$$

In other words, the finite-sample observed information matrices converge in probability to the expected information matrix corresponding to the same parameter $\boldsymbol{\theta}$.

In this study, we use large sample size n to establish the convergence of the finite-sample information matrix to the expected information matrix to confirm the accuracy of the expected information matrix computation from the function `stInfoMat` with `type = ‘‘expected’’`. We are concerned about the standard errors that may be infinitesimally small when n gets too large, so we show both the element-wise absolute differences between the two matrices and the standard errors for comparison of how large the individual absolute differences are compared to their corresponding standard errors.We have tried several sample sizes n, and found $n = 10,000$ quite adequate in terms of the standard errors shown in Table 2.

We now compute the finite-sample observed information matrix, given by the left-hand side of (23) and computed with the function `stInfoMat` with `type = ‘‘observed’’`, and compare it to the expected information matrix corresponding to $\boldsymbol{\theta}_{med} = (0.03,\ 0.08,\ -0.36,\ 5.50)$ computed with the function `stInfoMat` with `type = ‘‘expected’’`. With the sample size of $n = 10,000$, the elements of the finite-sample observed information matrix and the expected information matrix are given in Table 2.

Diagonal	Expected	Observed	AD	SE	AD/SE
$\mathbf{I}_{11}$	144.7	144.5	0.136	0.760	0.179
$\mathbf{I}_{22}$	232.6	233.2	0.604	3.89	0.155
$\mathbf{I}_{33}$	0.501	0.503	0.0016	0.0052	0.308
$\mathbf{I}_{44}$	0.00223	0.00216	0.000074	0.000066	1.121

Off-diagonal	Expected	Observed	AD	SE	AD/SE
$\mathbf{I}_{12}$	−28.4	−27.6	0.807	2.33	0.346
$\mathbf{I}_{13}$	7.95	7.95	0.0035	0.044	0.795
$\mathbf{I}_{14}$	0.0274	0.0267	0.0008	0.0050	0.160
$\mathbf{I}_{23}$	1.610	1.629	0.0188	0.0670	0.281
$\mathbf{I}_{24}$	−0.490	−0.503	0.0125	0.0135	0.926
$\mathbf{I}_{34}$	−0.00439	−0.00458	0.00023	0.00027	0.852

Table 2: Absolute difference (AD) and standard error (SE) comparisons of diagonal elements (Top) and off-diagonal elements (Bottom) of the expected information matrix and the finite-sample observed information matrix based on $10,000$ realizations of y_i.

This tells us that the finite-sample observed information matrix with large sample size n is very close to the expected information matrix, so the convergence (23) is established here thereby showing that our expected information matrix computation is reliable.

4.4 Observed Information Matrix Calculation Comparison

The function `stInfoMat` with `type = ‘‘observed’’` in `skewtInfo` package evaluates the finite-sample observed information matrix, i.e. the left-hand-side of (23). It inputs n i.i.d. realizations of y_i simulated from a skew-t distribution with the assumed parameter $\boldsymbol{\theta}$. To compute the observed information matrix using (22), we have to use involved closed-form formulas of the second partial derivatives of the skew-t information matrix as expressed in Appendix B. Therefore, the code can be very error-prone and therefore needs to be appropriately verified to ensure its computational accuracy. In this section, we compare the observed information matrix computed with our function `stInfoMat` with `type = "observed"` to that computed with the function `st.infoUv` in the `sn` package for the case of the median skew-t parameter vector (20).

With $n = 50,\ 100,\ 200$, our finite-sample observed information matrix computation agrees with the finite-sample observed information matrix implemented in the `sn` package. The element-wise absolute differences of the expected information matrices from our computation and the `sn` package are of order $O(10^{-9})$. For example, in the case of $n = 50$, the finite-sample observed information matrix from the function `stInfoMat` is

$$\begin{bmatrix} 130 & -26.3 & 7.610 & 0.103 \\ -26.3 & 306 & 2.09 & -0.660 \\ 7.610 & 2.09 & 0.594 & -0.0025 \\ 0.103 & -0.660 & -0.0025 & 0.0017 \end{bmatrix}$$

with the element-wise absolute differences compared to that from the function `st.infoUv` of order $O(10^{-12})$. This ensures our package's computational accuracy.

Table 3 shows the comparison between the standard errors of the skew-t MPLE's from the `sn` package to our package[6].

[6]The finite-sample standard errors of the MPLE's are given by $\sqrt{\text{diag}\left(\mathbf{V}_{\boldsymbol{\theta}_{med}}\right)/n}$ where $\mathbf{V}_{\boldsymbol{\theta}_{med}} = \mathbf{I}^{-1}_{\boldsymbol{\theta}_{med}}$.

S.E.	$n = 50$		$n = 100$		$n = 200$	
	`sn`	`info`	`sn`	`info`	`sn`	`info`
ξ	0.0301	0.0592	0.0260	0.0419	0.0273	0.0296
ω	0.0205	0.0202	0.0115	0.0143	0.0106	0.0101
α	0.4395	1.002	0.4159	0.7087	0.4607	0.5011
ν	8.845	4.167	2.974	2.947	2.052	2.084

Table 3: Standard errors of the skew-t MPLE's from the `sn` package and `skewtInfo` package

Most of the standard errors obtained from the function `st.infoUv` are lower than and getting closer to finite-sample standard errors as the sample size increases. However, the standard errors of ξ and α from the function `st.infoUv` decrease from $n = 50$ to $n = 100$ but increase from $n = 100$ to $n = 200$.

5 Concluding Discussion

Due to the fact that there is no analytic formula for the skew-t expected information matrix, this Master's Thesis has examined the skew-t information matrices and provided the way to numerically evaluate the skew-t expected information matrix via the function `stInfoMat` in an `R` package `skewtInfo` with `type = ‘‘expected’’`. We verify the information matrix computation by using the well-known fact in statistics that the observed information matrices, computed using the function `stInfoMat` with `type = ‘‘observed’’`, converge to the negative of the expected information matrix. The results have shown the convergence of the finite-sample observed information matrices to the expected information matrix with the sample size of $n = 10,000$, indicating that the numerical integrations performed pretty accurately in the expected information matrix computations.

Once we have accurate skew-t expected information matrices, we can now extend our study to explore the finite sample properties of the skew-t MPLE's. We will then take a step further to evaluate loss of statistical efficiency, i.e. increase in variance, of skew-t MPLE's relative to symmetric-t MLE's when returns do in fact have a symmetric-t distribution. Finally, we will study the performance of parametric expected shortfall (ES) for skew-t distributions, in comparison with ES for symmetric t-distributions - going a step forward from Martin and Zhang (2017) study of ES MLE for symmetric t-distributions.

Appendix A: The MPLE Penalty Function

The actual choice of Q stems from Sartori [2006] by replacing the first derivative (with respect to α) likelihood function $\ell'(\alpha) = 0$ by the modified form

$$\ell'(\alpha) + M(\alpha) = 0, \tag{24}$$

where $M(\alpha)$ is the correction term equal to the first derivative of the penalty function Q with respect to α.

In the centered skew-normal case $Z \sim SN(0,\ 1,\ \alpha)$, Satori asserted that the correction term $M(\alpha)$ is of the form:

$$M(\alpha) = -\frac{\alpha}{2} \cdot \frac{a_4(\alpha)}{a_2(\alpha)}, \quad a_p(\alpha) = \mathbb{E}\left\{Z^p \zeta_1(\alpha Z)^2\right\}$$

and obtained that $a_2(\alpha)/a_4(\alpha)$ is a linear function in α^2 . That is,

$$\frac{a_2(\alpha)}{a_4(\alpha)} \approx e_1 + e_2 \alpha^2.$$

With $X \sim \mathbf{N}(0,1)$ and $\zeta_1(x) = \zeta_0'(x)$ the inverse Mill's ratio, and by matching the behavior when $\alpha^2 = 0$ and $\alpha^2 \to \infty$, we obtain the expressions of e_1 and e_2 as follows:

$$\begin{aligned} e_1 &= \frac{a_2(0)}{a_4(0)} = \frac{\mathbf{E}\left[X^2\right]}{\mathbf{E}\left[X^4\right]} = \frac{1}{3}, \\ e_2 &= \lim_{\alpha^2 \to \infty} \frac{1}{\alpha^2} \left(\frac{a_2(\alpha)}{a_4(\alpha)} - e_1 \right) = \frac{\mathbf{E}\left[X^2 \zeta_1\left(X\right)\right]}{\mathbf{E}\left[X^4 \zeta_1\left(X\right)\right]} \approx 0.28542. \end{aligned}$$

Hence the penalty function is given by

$$Q(\alpha) \approx \frac{1}{4e_2} \log\left(1 + \frac{e_2}{e_1} u^2\right).$$

AAV then extended this to the general skew-normal case $Z \sim \mathbf{SN}(\xi,\ \omega^2,\ \alpha)$ by using the same penalty function and solve for the MPLEs from equation (11). They obtain very close parameter values to the skew-normal MPLEs obtained from Satori's two-step approach - fixing the location and scale parameters to be MLEs $\widehat{\xi}$ and $\widehat{\omega}$ and solve for $\widetilde{\alpha}$ from (24).

In the skew-t case, AAV also noticed that in the centered skew-t random variable $Z \sim \mathbf{ST}(0, 1,\ \alpha,\ \nu)$:

$$\frac{a_2(\alpha)}{a_4(\alpha)} \approx e_{1\nu} + e_{2\nu}\alpha^2.$$

By the same approach of two-side behavior matching $\alpha^2 = 0$ and $\alpha^2 \to \infty$, fixed ν and $X_k \sim t_{\nu+k}$ the symmetric-t random variable with $\nu + k$ degrees of freedom, we have

$$\begin{aligned} e_{1\nu} &= \frac{(\nu+2)(\nu+3)}{3(\nu+1)^2}, \\ e_{2\nu} &= \left(\frac{(\nu+2)(\nu+3)}{(\nu+1)^2}\right)^2 \frac{\mathbb{E}\left[X_1^2\zeta_1(X_1;\nu+1)\right]}{\mathbb{E}\left[X_3^4\zeta_1\left(X_3\sqrt{(\nu+1)/(\nu+3)};\nu+1\right)\right]}. \end{aligned}$$

The correction term $M(\alpha)$ is therefore given by

$$M(\alpha) \approx -\frac{\alpha}{2\left(e_{1\nu} + e_{2\nu}\alpha^2\right)},$$

so that

$$Q(\alpha) \approx \frac{1}{4e_{2\nu}} \log\left(1 + \frac{e_{2\nu}}{e_{1\nu}}\alpha^2\right). \tag{25}$$

Note that the form of the penalty functions in the centered skew-normal and skew-t cases are of the form

$$Q(\alpha) = c_1 \log\left(1 + c_2\alpha^2\right),$$

where $(c_1,\ c_2) = (1/4e_2,\ e_2/e_1)$ in the skew-normal case and $(c_1,\ c_2) = (1/4e_{2\nu},\ e_{2\nu}/e_{1\nu})$ in the skew-t case.

AAV applied the analogous approach as in the skew-normal case - using the same penalty function (25) and solve for the MPLE's from Equation (11). However, this approach posed a high computational cost since, when ν is not fixed, the numerical optimization used to solve (11) will look for different ν and then numerically integrate the two integrals in $e_{2\nu}$. AAV then gave the approximation:

$$e_{2\nu} \approx e_2 \left(1 + \frac{4}{\nu + \gamma}\right),$$

where $\gamma \approx 0.57721$ is the Euler's constant - the asymptotic difference between the harmonic series and the natural logarithm.

Appendix B: The Information Matrix Alternative Form

As is well-known, the skew-t expected information matrix can be expressed in the form

$$\mathbf{I}_{\boldsymbol{\theta}} = -\begin{bmatrix} \mathbb{E}[S_{\xi\xi}(y)] & \mathbb{E}[S_{\xi\omega}(y)] & \mathbb{E}[S_{\xi\alpha}(y)] & \mathbb{E}[S_{\xi\nu}(y)] \\ \mathbb{E}[S_{\xi\omega}(y)] & \mathbb{E}[S_{\omega\omega}(y)] & \mathbb{E}[S_{\omega\alpha}(y)] & \mathbb{E}[S_{\omega\nu}(y)] \\ \mathbb{E}[S_{\xi\alpha}(y)] & \mathbb{E}[S_{\omega\alpha}(y)] & \mathbb{E}[S_{\alpha\alpha}(y)] & \mathbb{E}[S_{\alpha\nu}(y)] \\ \mathbb{E}[S_{\xi\nu}(y)] & \mathbb{E}[S_{\omega\nu}(y)] & \mathbb{E}[S_{\alpha\nu}(y)] & \mathbb{E}[S_{\nu\nu}(y)] \end{bmatrix}.$$

Define[7]

$$\begin{aligned}
w &= \frac{t_{\nu+1}(\alpha z\tau)}{T_{\nu+1}(\alpha z\tau)}, \\
w_z &= -\frac{\nu(\nu+2)\alpha^2 z w}{(\nu+z^2+\alpha^2 z^2)(\nu+z^2)} - \frac{\nu\alpha\tau w^2}{\nu+z^2}, \\
w_\alpha &= -\frac{(\nu+2)\alpha z^2 w}{(\nu+z^2+\alpha^2 z^2)} - z\tau w^2, \\
w_\nu &= \frac{w}{2}\left\{\frac{(\nu+2)\alpha^2 z^2}{(\nu+z^2+\alpha^2 z^2)(\nu+z^2)} - \log\left(1+\frac{\alpha^2 z^2}{\nu+z^2}\right) + \frac{\eta}{T(\alpha z\tau;\nu+1)}\right\} \\
&\quad + \frac{\alpha z(1-z^2)w^2}{2\tau(\nu+z^2)^2}, \\
\beta &= \int_{-\infty}^{\alpha z\tau}\left\{\frac{(\nu+2)\alpha^2 u^2}{(\nu+1)(\nu+1+\alpha^2 u^2)} - \log\left(1+\frac{\alpha^2 u^2}{\nu+1}\right)\right\}^2 t(u;\nu+1)\,du, \\
\delta &= \int_{-\infty}^{\alpha z\tau}\frac{(\nu u^2-2\nu-2)u^2}{(\nu+1)^2(\nu+1+u^2)^2}\,t(u;\nu+1)du, \\
\eta &= \int_{-\infty}^{\alpha z\tau}\left\{\frac{(\nu+2)\alpha^2 u^2}{(\nu+1)(\nu+1+\alpha^2 u^2)} - \log\left(1+\frac{\alpha^2 u^2}{\nu+1}\right)\right\} t(u;\nu+1)\,du.
\end{aligned}$$

According to Ciccio and Monti [2011], the second partial derivatives of the skew-t log-likelihood function in the alternative formula of expected information matrix above can be expressed as follows:

[7]For a clean look of the involved terms, we suppress the dependency of all of the terms to the variable z. For example, we write τ instead of $\tau(z)$ to make the expression cleaner in complicated formulas.

$$
\begin{aligned}
S_{\xi\xi}(y) &= \frac{1}{\omega^2}\left(\frac{2\tau^2 z^2}{\nu+z^2} - \tau_z^2 - \frac{3\alpha\tau\nu z w}{(\nu+z^2)^2} + \frac{\alpha\tau\nu w_z}{\nu+z^2}\right) \\
S_{\xi\omega}(y) &= \frac{z}{\omega^2}\left(\frac{2\tau^2 z^2}{\nu+z^2} - \tau^2 - \frac{3\alpha\tau\nu z w}{(\nu+z^2)^2} + \frac{\alpha\tau\nu w_z}{\nu+z^2}\right) + \frac{1}{\omega^2}\left(\tau^2 z + \frac{\alpha\tau\nu w}{\nu+z^2}\right) \\
S_{\xi\alpha}(y) &= -\frac{1}{\omega}\left(\frac{\nu\tau(w+\alpha w_\alpha)}{\nu+z^2}\right) \\
S_{\xi\nu}(y) &= -\frac{1}{\omega}\left(\frac{z(1-z^2)}{(\nu+z^2)^2} + \frac{\alpha\{\nu(3z^2-1)+2z^2\}w}{2\tau(\nu+z^2)^3} + \frac{\alpha\tau\nu w_\nu}{\nu+z^2}\right) \\
S_{\omega\omega}(y) &= \frac{1}{\omega^2} + \frac{z^2}{\omega^2}\left(\frac{2\tau^2 z^2}{\nu+z^2} - \tau^2 - \frac{3\alpha\tau\nu z w}{(\nu+z^2)^2} + \frac{\alpha\tau\nu w_z}{\nu+z^2}\right) + \frac{2z}{\omega^2}\left(\tau^2 z + \frac{\alpha\tau\nu w}{\nu+z^2}\right), \\
S_{\omega\alpha}(y) &= -\frac{z}{\omega}\left(\frac{\nu\tau(w+\alpha w_\alpha)}{\nu+z^2}\right), \\
S_{\omega\nu}(y) &= -\frac{z}{\omega}\left(\frac{z(1-z^2)}{(\nu+z^2)^2} + \frac{\alpha\{\nu(3z^2-1)+2z^2\}w}{2\tau(\nu+z^2)^3} + \frac{\alpha\tau\nu w_\nu}{\nu+z^2}\right), \\
S_{\alpha\alpha}(y) &= z\tau w_\alpha, \\
S_{\alpha\nu}(y) &= \frac{z(z^2-1)w}{2\tau(\nu+z^2)^2} + z\tau w_\nu, \\
S_{\nu\nu}(y) &= \frac{1}{4}\left\{\psi_1\left(\frac{\nu}{2}+1\right) - \psi_1\left(\frac{\nu}{2}\right)\right\} + \frac{2\nu^2+2\nu+1}{2\nu^2(\nu+1)^2} + \frac{z^2}{2\nu(\nu+z^2)} - \frac{z^2(\nu^2+2\nu+z^2)}{2\nu^2(\nu+z^2)^2} \\
&\quad - \frac{\alpha z(z^2-1)(z^2+4\nu+3)w}{4\tau(\nu+1)(\nu+z^2)^3} + \frac{\alpha z(1-\tau^2)w_\nu}{2\tau(\nu+z^2)} - \frac{\eta^2}{4T(\alpha z\tau;\nu+1)^2} \\
&\quad + \frac{2\delta+\beta}{4T(\alpha z\tau;\nu+1)} - \frac{\alpha z(z^2-1)w\eta}{4T(\alpha z\tau;\nu+1)\tau(\nu+z^2)^2} \\
&\quad + \frac{\alpha z(z^2-1)w}{4\tau(\nu+z^2)^2}\left\{\frac{(\nu+2)\alpha^2 z^2}{(\nu+1)(\nu+z^2+\alpha^2 z^2)} - \log\left(1+\frac{\alpha^2 z^2}{\nu+z^2}\right)\right\}.
\end{aligned}
$$

References

Diego Amaya, Peter Christoffersen, Kris Jacobs, and Aurelio Vasquez. Does realized skewness predict the cross-section of equity returns? *Journal of Financial Economics*, 118:135–167, 2015.

Adelchi Azzalini. *The Skew-Normal and Related Families*. Institute of Mathematical Statistics Monographs, 2013.

Adelchi Azzalini and Reinaldo B. Arellano-Valle. Maximum penalized likelihood estimation for skew-normal and skew-t distributions. 2012.

Adelchi Azzalini and Antonella Capitanio. Distributions generated by perturbation of symmetry with emphasis on a multivariate skew t distribution. 2009. doi: 10.1111/1467-9868.00391.

George Casella and Roger Berger. *Statistical Inference*. Duxbury Resource Center, June 2001. ISBN 0534243126.

Thomas J. Di Ciccio and Anna Clara Monti. Inferential aspects of the skew t-distribution. *Quaderni di Statistica*, 13, 2011.

Abdullah Al Masud. The effect of kurtosis on the cross-section ofstock returns. *All Graduate Plan B and other Reports*, 180, 2012.

Robert Piessens, Elise de Doncker-Kapenga, Christoph W. Uberhuber, and David K. Kahaner. *Quadpack - A Subroutine Package for Automatic Integration*. Springer Series in Computational Mathematics, 1983.

Nicola Sartori. Bias prevention of maximum likelihood estimates for scalar skew normal and skew t distributions. *Journal of Statistical Planning and Inference*, 136(12):4259 – 4275, 2006. ISSN 0378-3758. doi: https://doi.org/10.1016/j.jspi.2005.08.043. URL `http://www.sciencedirect.com/science/article/pii/S0378375805002405`.