Quaderni di Statistica
Vol. 13, 2011

Inferential aspects of the skew *t*-distribution

Thomas J. Di Ciccio
Department of Social Statistics, Cornell University
E-mail: tjd9@cornell.edu

Anna Clara Monti
Department of Social, Juridical and Economic Systems, University of Sannio
E-mail:acmonti@unisannio.it

Summary: This paper concerns likelihood inference for the skew t-distribution, which includes both the skew normal and the normal distributions as important special cases that occur when the degrees of freedom is infinite. Inference based on the skew t-model becomes problematic in these special cases for two reasons: the expected information matrix is singular; and the parameter corresponding to the degrees of freedom takes a value occurring at the boundary of its parameter space. For each of the special cases, a reparameterization is introduced that copes with these difficulties, thereby producing consistent estimators with known asymptotic properties. Inference for multiple linear regression models based on the skew t-distribution is also considered.

Keywords: Asymptotic distribution; Boundary-value parameter; Flexible parametric model; Likelihood inference; Linear regression; Non-standard asymptotics; Skew normal distribution; Skew t-distribution; Singular information matrix.

1. Introduction

The univariate skew t-distribution, with location parameter ξ, scale parameter ω, skewness parameter α, and degrees of freedom ν, has density function

$$f\big(y;\xi,\omega,\alpha,\nu\big) = 2\omega^{-1}t(z;\nu)T(\alpha z\tau;\nu+1), \quad -\infty < y < \infty, \tag{1.1}$$

where $z = (y-\xi)/\omega$ and $\tau = \big\{(\nu+1)/\big(z^2+\nu\big)\big\}^{1/2}$. In formula (1.1)

$$t(z;\nu) = \frac{\Gamma\big\{(\nu+1)/2\big\}}{(\pi\nu)^{1/2}\Gamma(\nu/2)}\big(1+z^2/\nu\big)^{-(\nu+1)/2}, \quad -\infty < z < \infty,$$

and $T(z;\nu) = \int_{-\infty}^{z} t(u;\nu)\, du$ are the density and cumulative distribution functions, respectively, of Student's t-distribution with ν degrees of freedom. The skew t-distribution is denoted by $St(\xi, \omega, \alpha, \nu)$. The skew t-distribution was introduced by Branco and Dey (2001) and Azzalini and Capitanio (2003). As the skewness parameter α and the degrees of freedom ν vary, this model can accommodate both skewness and heavy tails. Thus, it provides considerable flexibility for fitting data that exhibit deviations from normality.

Special cases of the skew t-distribution are the location-scale Student's t-distribution, obtained when $\alpha = 0$, and the skew normal (SN) distribution, obtained as $\nu \to \infty$. The $SN(\xi, \omega, \alpha)$ distribution (Azzalini, 1985; Azzalini and Capitanio, 1999) has density

$$f_{SN}(y;\xi,\omega,\alpha) = 2\omega^{-1}\phi(z)\Phi(\alpha z), \quad z = (y-\xi)/\omega,\ -\infty < y < \infty,$$

where $\phi(z)$ and $\Phi(z)$ are the density and cumulative distribution functions, respectively, of the standard normal distribution. When both $\alpha = 0$ and $\nu \to \infty$, the $St(\xi, \omega, \alpha, \nu)$ distribution tends to the normal with mean ξ and standard deviation ω.

When the distribution is either skew normal or normal, inference based on the skew t-model becomes problematic for two reasons: the expected information matrix is singular; and the parameter corresponding to the degrees of freedom takes a value occurring at the boundary of its parameter space. For each of these special cases, the present paper proposes a reparameterization which copes with these difficulties, thereby producing consistent estimators with known asymptotic properties.

In the next section, the score function is derived in the general case; furthermore, the behaviour of the score function and the information matrix is investigated in the special cases where the distribution is Student's t, skew normal or normal. Problems arising in likelihood inference are illustrated in Section 3. Sections 4 and 5 provide the reparameterizations to be adopted when the distribution is skew normal or normal, respectively. Inference for the multiple linear regression model is considered in Section 6. Technical details are confined to the Appendices.

2. Score function and information matrix

Let $S(y) = \big\{S_\xi(y), S_\omega(y), S_\alpha(y), S_\nu(y)\big\}'$ be the score function of the St model based on a single observation y; thus, $S_\xi(y) = \partial \ln f\big(y;\xi,\omega,\alpha,\nu\big)/\partial\xi$, and so forth. The components of $S(y)$ are

$$S_\xi(y) = \frac{z\tau^2}{\omega} - \frac{\alpha\tau\nu}{\omega(\nu+z^2)}w, \quad S_\omega(y) = -\frac{1}{\omega} + \frac{z^2\tau^2}{\omega} - \frac{\varsigma\nu}{\omega(\nu+z^2)}w, \quad S_\alpha(y) = z\tau w,$$

$$\begin{aligned} S_\nu(y) = \frac{1}{2}\bigg\{ &\Psi\left(\frac{\nu}{2}+1\right) - \Psi\left(\frac{\nu}{2}\right) - \frac{2\nu+1}{\nu(\nu+1)} - \ln\left(1+\frac{z^2}{\nu}\right) + \frac{z^2\tau^2}{\nu}, \\ &+ \frac{\alpha z\big(z^2-1\big)}{\big(\nu+z^2\big)^2\tau}w + \frac{\gamma}{T(\varsigma;\nu+1)}\bigg\}, \end{aligned} \tag{2.1}$$

where $\varsigma = \alpha z \tau$, $w = t(\varsigma; \nu+1)/T(\varsigma; \nu+1)$,

$$\gamma = \int_{-\infty}^{\varsigma} \left\{ \frac{(\nu+2)u^2}{(\nu+1)(\nu+1+u^2)} - \ln\left(1 + \frac{u^2}{\nu+1}\right) \right\} t(u; \nu+1) du,$$

and $\Psi(x) = \partial \ln\{\Gamma(x)\}/\partial x$. Since $S_\nu(y)$ is of order $O(\nu^{-2})$, it vanishes as $\nu \to \infty$.

The second derivatives $S_{\xi\xi}(y)$, where $S_{\xi\xi}(y) = \partial^2 \ln f(y; \xi, \omega, \alpha, \nu)/\partial\xi\partial\xi$, and so forth, which are useful for computing the observed information matrix for $(\xi, \omega, \alpha, \nu)$ based on a single observation y, are derived in Appendix A.1. Since these second derivatives are bounded and continuous functions for fixed ν, their expectations are finite. Thus, the expected information matrix I, whose components are $I_{\xi\xi} = -E\{S_{\xi\xi}(y)\}$, and so forth, always exists.

Azzalini and Genton (2008) discussed likelihood inference for the St model. They showed that the profile likelihood function for the skewness parameter α does not have an inflection point when $\alpha = 0$, as it does for the SN model, and that its shape is closer to quadratic for the St model than it is for other flexible models such as the SN and the Skew Exponential Power (Azzalini, 1986; DiCiccio and Monti, 2004). They also investigated the finite-sample performance of maximum likelihood estimators in the St model.

It is illuminating to consider the score function and the expected information matrix for the two special cases of the location-scale Student's t-distribution ($\alpha = 0$) and the SN distribution ($\nu \to \infty$). Calculations summarized in Appendix A.2 show that, when $\alpha = 0$, the components of the score function become

$$S_\xi^t(y) = \frac{z\tau^2}{\omega}, \quad S_\omega^t(y) = \frac{z^2\tau^2 - 1}{\omega}, \quad S_\alpha^t(y) = 2z\tau t(0; \nu+1),$$

$$S_\nu^t(y) = \frac{1}{2}\left\{ \Psi\left(\frac{\nu}{2} + \frac{1}{2}\right) - \Psi\left(\frac{\nu}{2}\right) - \ln\left(1 + \frac{z^2}{\nu}\right) + \frac{z^2 - 1}{\nu + z^2} \right\}. \tag{2.2}$$

Observe that the components for ξ, ω, and ν are the same as those for the location-scale Student's t-model. The non-zero components of the information matrix are shown in Appendix A.2 to be

$$I_{\xi\xi}^t = \frac{\nu+1}{\omega^2(\nu+3)}, \qquad I_{\xi\alpha}^t = \left(\frac{\nu}{\pi}\right)^{1/2} \frac{\Gamma(\nu/2 + 3/2)}{\omega\Gamma(\nu/2+2)}, \qquad I_{\omega\omega}^t = \frac{2\nu}{\omega^2(\nu+3)},$$

$$I_{\omega\nu}^t = -\frac{2}{\omega(\nu+1)(\nu+3)}, \qquad I_{\alpha\alpha}^t = 4\frac{\Gamma(\nu/2+1)^2}{\pi(\nu+1)\Gamma(\nu/2+1/2)^2}, \tag{2.3}$$

$$I_{\nu\nu}^t = \frac{1}{4}\left\{ \Psi_1\left(\frac{\nu}{2}\right) - \Psi_1\left(\frac{\nu}{2} + \frac{1}{2}\right) \right\} - \frac{\nu+5}{2\nu(\nu+1)(\nu+3)},$$

where $\Psi_1(x) = \partial\Psi(x)/\partial x$. Since $I_{\xi\omega}^t = I_{\xi\nu}^t = I_{\omega\alpha}^t = I_{\alpha\nu}^t = 0$, it follows that the maximum likelihood estimators $(\hat{\xi}, \hat{\alpha})$ and $(\hat{\omega}, \hat{\nu})$ are asymptotically independent. The maximum likelihood estimators of the parameters ω and ν have the same asymptotic variances in the St model as they have in the location-scale Student's t-model; estimation of the additional parameter α in the St model only affects the asymptotic properties of the maximum likelihood estimator of ξ. Furthermore, Azzalini and Genton (2008) remarked that, for finite ν, the expected information matrix for $(\xi, \omega, \alpha, \nu)$ in the St model is invertible when $\alpha = 0$, which is in contrast to the case of the SN model, where the information matrix for (ξ, ω, α) is singular when $\alpha = 0$.

As $\nu \to \infty$, the St distribution tends to the SN, and the components of the score function become

$$S_\xi^{SN}(y) = \frac{z}{\omega} - \frac{\alpha}{\omega} w_\phi, \quad S_\omega^{SN}(y) = -\frac{1}{\omega} + \frac{z^2}{\omega} - \frac{\alpha z}{\omega} w_\phi,$$

$$S_\alpha^{SN}(y) = z w_\phi, \quad S_\nu^{SN}(y) = 0,$$

where $w_\phi = \phi(\alpha z)/\Phi(\alpha z)$. Since $S_\nu^{SN}(y) = 0$, the information matrix for $(\xi, \omega, \alpha, \nu)$ in the St model becomes singular as $\nu \to \infty$. This singularity is understandable, for, if the parameters ξ, ω, and α are all known in the St model and only ν is estimated, then, under the SN distribution, the parameter ν is infinite, and thus the variance of the estimator is infinite.

The score components $S_\xi^{SN}(y)$, $S_\omega^{SN}(y)$, and $S_\alpha^{SN}(y)$ are the same as the components for ξ, ω, and α under the SN model, so the portion of the expected information matrix relating to these parameters in the St model coincides with the information matrix for (ξ, ω, α) under the SN model given by Azzalini (1985); thus,

$$I_{\xi\xi}^{SN} = \omega^{-2}\left(1 + \alpha^2 d_0\right), \qquad I_{\xi\omega}^{SN} = \omega^{-2}\left\{b\delta\left(1+\delta^2\right) + \alpha^2 d_1\right\},$$
$$I_{\xi\alpha}^{SN} = \omega^{-1}\left\{b/\left(1+\alpha^2\right)^{3/2} - \alpha d_1\right\}, \qquad I_{\omega\omega}^{SN} = \omega^{-2}\left(2 + \alpha^2 d_2\right),$$
$$I_{\omega\alpha}^{SN} = -\omega^{-1}\alpha d_2, \qquad I_{\alpha\alpha}^{SN} = d_2,$$

where $b = (2/\pi)^{1/2}$, $\delta = \alpha/\left(1+\alpha^2\right)^{1/2}$, and $d_r = E(Z^r W_\phi^2)$ $(r = 0, 1, \ldots)$. Moreover, $I_{\xi\nu}^{SN} = I_{\omega\nu}^{SN} = I_{\alpha\nu}^{SN} = I_{\nu\nu}^{SN} = 0$. The singularity of the expected information matrix for $(\xi, \omega, \alpha, \nu)$ in the St model as $\nu \to \infty$, i.e., under the SN distribution, can be alleviated by a reparameterization discussed in Section 4.

Finally, for the normal distribution, obtained when $\alpha = 0$ and $\nu \to \infty$, the components of the score function become

$$S_\xi^N(y) = \frac{z}{\omega}, \qquad S_\omega^N(y) = \frac{z^2 - 1}{\omega}, \qquad S_\alpha^N(y) = z\left(\frac{2}{\pi}\right)^{1/2}, \qquad S_\nu^N(y) = 0.$$

In addition to the score component for ν being 0, as occurs in the case of the SN distribution, the component for ξ is a multiple of the one for α. Thus, the expected information matrix is singular with rank 2 and takes the form

$$I^N = \begin{pmatrix} \frac{1}{\omega^2} & 0 & \frac{1}{\omega}\left(\frac{2}{\pi}\right)^{1/2} & 0 \\ 0 & \frac{2}{\omega^2} & 0 & 0 \\ \frac{1}{\omega}\left(\frac{2}{\pi}\right)^{1/2} & 0 & \frac{2}{\pi} & 0 \\ 0 & 0 & 0 & 0 \end{pmatrix}.$$

Singular information matrices also arise in three related situations: in the location-scale Student's t-model, the expected information matrix for (ξ, ω, ν) becomes singular as $\nu \to \infty$; in the SN model, the observed and expected information matrices for (ξ, ω, α) become singular as $\alpha \to \pm\infty$; and in the St model, both the observed and expected information matrices for $(\xi, \omega, \alpha, \nu)$ become singular as $\alpha \to \pm\infty$. The singularity of the information matrices for the SN and St models as $\alpha \to \pm\infty$ is shown in Appendix A.3.

3. Problems arising in likelihood inference

Let $(\hat{\xi}, \hat{\omega}, \hat{\alpha}, \hat{\nu})$ be the maximum likelihood estimator of $(\xi, \omega, \alpha, \nu)$ based on a sample of size n from the $St(\xi, \omega, \alpha, \nu)$ distribution.

The singularity of the expected information matrix for $(\xi, \omega, \alpha, \nu)$ as $\nu \to \infty$ has an undesirable consequence on the efficiency of the estimators when the distribution is normal or close to normal, i.e., when α is in a neighborhood of 0. Table 1 shows the root mean square errors of the estimators of ξ, ω, α, and $\kappa = 1/\nu$, the inverse degrees of freedom, for several samples sizes in the case of the standard normal distribution. The table entries were obtained by a simulation of size 10,000, and they demonstrate that the root mean square errors of $\hat{\xi}$ and $\hat{\alpha}$ fail to decrease at the usual $n^{-1/2}$ rate. The mean square errors of the estimators behave similarly when the distribution is SN with α close to 0. Furthermore, this slow rate of decrease of the mean square errors for the estimators of ξ and α can also be observed in the SN model when α is in the vicinity of 0. In contrast to the behavior of $\hat{\xi}$ and $\hat{\alpha}$, the root mean square errors in Table 1 for $\hat{\omega}$ and $\hat{\kappa}$ do seem to decrease at the regular rate. Thus, it is reasonable to conjecture that the inefficiency of $\hat{\xi}$ and $\hat{\alpha}$ is attributable mainly to the linear dependence between $S_\xi(y)$ and $S_\alpha(y)$ under the normal distribution.

It should be noted, however, although ξ and α cannot be estimated reliably, efficient estimators can be obtained for other functions of the parameter $(\xi, \omega, \alpha, \nu)$ that depend on ξ and α. For example, Table 2 shows the results of a simulation that investigates the efficiency of the maximum likelihood estimators of the probabilities under the St model assigned to the intervals with cutoffs $-\infty, -2, -1, 1, 2, \infty$ when the distribution

Table 1. Root mean square errors of the maximum likelihood estimators of $(\xi, \omega, \alpha, \kappa)$ *and* $(\mu, \sigma^2, \gamma_1, \gamma_2)$ *under the* $N(0, 1)$ *distribution.*

n	ξ	ω	α	κ	μ	σ	γ_1	γ_2
50	0.8340	0.3463	86336.21	0.0667	0.1426	0.3231	5.7941	12.4095
						(74)	(116)	(227)
100	0.7155	0.2650	1.5367	0.0459	0.0994	0.1422	0.7424	2.8703
								(9)
200	0.6200	0.2053	0.9778	0.0328	0.0714	0.1007	0.1915	0.4495
								(1)
500	0.5217	0.1484	0.7553	0.0211	0.0436	0.0635	0.1130	0.1874
1,000	0.4557	0.1165	0.6360	0.0157	0.0311	0.0443	0.0772	0.1253
5,000	0.3453	0.0692	0.4586	0.0073	0.0142	0.0202	0.0334	0.0506
10,000	0.3035	0.0542	0.3977	0.0052	0.0099	0.0139	0.0229	0.0352

The number of cases out of 10,000 simulations omitted for producing $\hat{\nu}$ smaller than the value required for the existence of the corresponding parameter is shown in parentheses.

is standard normal. Despite the inefficiency of the estimators $\hat{\xi}$ and $\hat{\alpha}$, the estimators of the interval probabilities have biases and standard deviations that appear to decrease at satisfactory rates.

Since ν approaches the boundary of the parameter space as $\nu \to \infty$, the resulting singularity of the expected information matrix in the St model when the distribution is normal cannot be avoided by using the reparametrization technique introduced by

Table 2. Biases and standard deviations (in parentheses) of interval probability estimators when the distribution is normal.

Probability (interval)	n 50	100	200	500	1,000	5,000	10,000
0.0228 $(-\infty, -2)$	0.0236 (0.0156)	0.0234 (0.0112)	0.0231 (0.0080)	0.0230 (0.0050)	0.0230 (0.0035)	0.0229 (0.0016)	0.0229 (0.0011)
0.1359 $(-2, -1)$	0.1288 (0.0358)	0.1307 (0.0235)	0.1318 (0.0162)	0.1335 (0.0100)	0.1342 (0.0071)	0.1352 (0.0033)	0.1354 (0.0023)
0.6827 $(-1, 1)$	0.6954 (0.0526)	0.6919 (0.0361)	0.6896 (0.0254)	0.6869 (0.0159)	0.6856 (0.0112)	0.6840 (0.0051)	0.6835 (0.0035)
0.1359 $(1, 2)$	0.1286 (0.0356)	0.1305 (0.0234)	0.1322 (0.0164)	0.1335 (0.0100)	0.1342 (0.0071)	0.1352 (0.0032)	0.1354 (0.0023)
0.0228 $(2, \infty)$	0.0236 (0.0158)	0.0235 (0.0112)	0.0233 (0.0080)	0.0231 (0.0050)	0.0230 (0.0035)	0.0229 (0.0016)	0.0228 (0.0011)

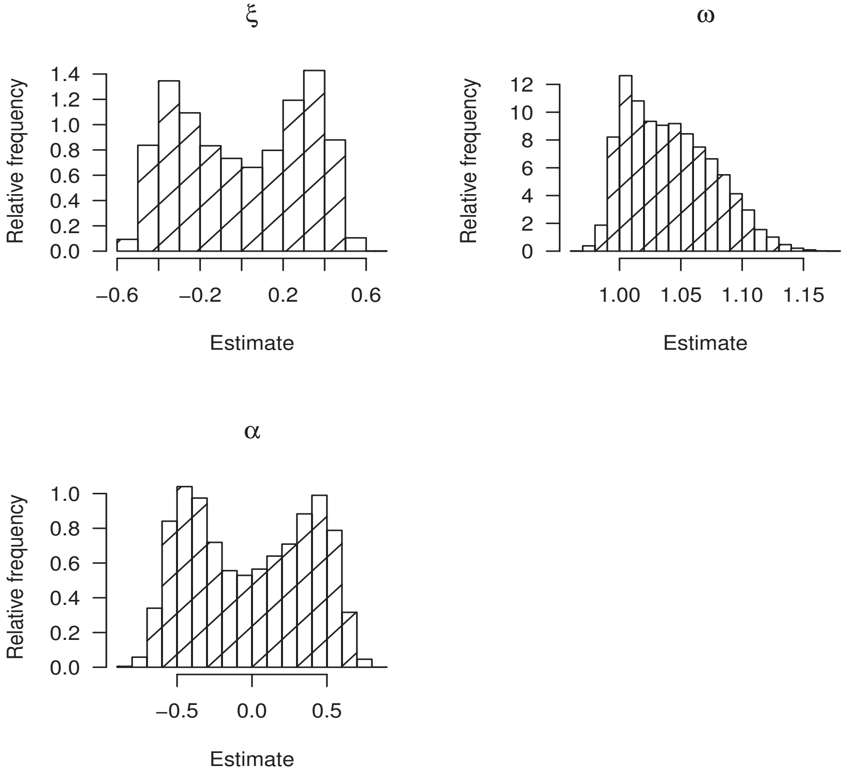

Figure 1. Histograms of the maximum likelihood estimates $\hat{\xi}$, $\hat{\omega}$ and $\hat{\alpha}$ from the St model when the distribution is $N(0,1)$ and the sample size is $n = 10,000$

Rotnitzky *et al*. (2000) as it can in other flexible models, such as in either the SN model (Azzalini, 1985; Chiogna, 2005) or the skew exponential power model (Azzalini, 1986; DiCiccio and Monti, 2004) when the distribution is normal. The dual problems of the singular information matrix and the boundary case affect the asymptotic distributions of the estimators, which fail to be normal as ν diverges. Figure 1 shows histograms, based on $10{,}000$ simulations, of the estimators $\hat{\xi}$, $\hat{\omega}$, and $\hat{\alpha}$ from the St model when the distribution is standard normal and the sample size is $n = 10{,}000$. The distributions of $\hat{\xi}$ and $\hat{\alpha}$ are both bimodal. Similar shapes have been observed for the distributions of estimators of ξ and α in the SN model when α is close to 0 (Arellano -Valle and Azzalini, 2008), i.e., when the distribution is close to normal.

When the St model is assumed and the distribution is a location-scale Student's t, the parameters take values within the parameter space and the expected information matrix is invertible. Consequently, when $\alpha = 0$ and $\nu < \infty$, the maximum likelihood estimators have their usual asymptotic properties.

Sections 4 and 5 provide alternative parameterizations for the St model that improve the properties of the maximum likelihood estimators. The reparameterizations provide consistent estimators having known asymptotic properties when the distribution is SN or normal.

A further inferential problem arising in both the SN and St models is that the profile likelihood of the skewness parameter α can be monotonic, leading to $\hat{\alpha} = \infty$ or $\hat{\alpha} = -\infty$, even though α is finite. Remedies for this problem have been proposed by Sartori (2005), Azzalini and Genton (2008), and Greco (2008). Results given in Appendix A.3 show that, for the St model, both the observed and expected information matrices evaluated at the maximum likelihood estimate, become singular when $\hat{\alpha} = \pm\infty$. In the SN model, only the expected information matrix behaves similarly; the observed information matrix evaluated at the maximum likelihood estimate is not necessarily singular when $\hat{\alpha} = \pm\infty$.

4. Reparameterization by the inverse degrees of freedom

As $\nu \to \infty$, the component $S_\nu(y)$ of the score function vanishes and the expected information matrix for $\big(\xi, \omega, \alpha, \nu\big)$ becomes singular. Reparameterization can remedy this difficulty, however, by using the inverse degrees of freedom, $\kappa = 1/\nu$, in place of ν. Thus, the parameter now becomes $(\xi, \omega, \alpha, \kappa)$, and the SN distribution corresponds to the boundary case $\kappa = 0$.

The component of the score function corresponding to κ is $S_\kappa(y) = -\kappa^{-2} S_\nu(y)$, i.e.,

$$S_\kappa(y) = \frac{1}{2\kappa^2}\Psi\Big(\frac{1}{2\kappa}\Big) - \frac{1}{2\kappa^2}\Psi\Big(\frac{1}{2\kappa}+1\Big) + \frac{(2+\kappa)}{2\kappa(1+\kappa)} + \frac{1}{2\kappa^2}\ln\big(1+\kappa z^2\big) - \frac{(1+\kappa)z^2}{2\kappa\big(1+\kappa z^2\big)} - \frac{\alpha z\big(z^2-1\big)w}{2\big(1+\kappa\big)^{1/2}\big(1+\kappa z^2\big)^{3/2}} - \frac{\gamma}{2\kappa^2 T(\varsigma, 1/\kappa+1)}. \qquad (4.1)$$

As $\kappa \to 0$, $S_\kappa(y)$ tends to

$$S_\kappa^{SN}(y) = \frac{1}{4}\big\{z^4 - 2z^2 - 1 - \alpha z\big(2z^2 + \alpha^2 z^2 - 1\big)w_\phi\big\}. \qquad (4.2)$$

which is clearly non-zero, nor is it a linear combination of $S_\xi^{SN}(y)$, $S_\omega^{SN}(y)$, and $S_\alpha^{SN}(y)$. Consequently, the information matrix for the parameterization $(\xi, \omega, \alpha, \kappa)$ is invertible under the SN distribution.

The components involving κ of the observed information matrix for $(\xi, \omega, \alpha, \kappa)$ in the St model are given in Appendix B, and their expectations under the SN distribution are derived. The corresponding components of the expected information matrix when

$\kappa = 0$ are shown to be

$$I_{\kappa\xi}^{SN} = \frac{b\delta}{4\omega}\left\{9 - 8\delta^2 + 3\delta^4 - \frac{6}{(1+\alpha^2)^2}\right\} - \frac{\alpha^2}{4\omega}\left\{d_1 - \left(2+\alpha^2\right)d_3\right\},$$

$$I_{\kappa\omega}^{SN} = \frac{2}{\omega} - \frac{\alpha^2}{4\omega}\left\{d_2 - \left(2+\alpha^2\right)d_4\right\}, \qquad I_{\kappa\alpha}^{SN} = \frac{1}{4}\alpha\Big\{d_2 - \left(2+\alpha^2\right)d_4\Big\}, \qquad (4.3)$$

$$I_{\kappa\kappa}^{SN} = \frac{7}{2} + \frac{1}{16}\alpha^2\Big\{d_2 - 2\left(2+\alpha^2\right)d_4 + \left(4+4\alpha^2+\alpha^4\right)d_6\Big\}.$$

Although the reparametrization produces a non-singular information matrix, the SN distribution corresponds to the value $\kappa = 0$, which occurs on the boundary of the parameter space. The asymptotic properties of the maximum likelihood estimator $\left(\hat\xi, \hat\omega, \hat\alpha, \hat\kappa\right)$ under the SN distribution, summarized in the following proposition, can be deduced from the results of Self and Liang (1987, Lemma 1 and Case 2).

PROPOSITION 4.1. If $\kappa = 0$, so that the distribution is SN, then $\left(\hat\xi, \hat\omega, \hat\alpha, \hat\kappa\right)$ is consistent with rate of convergence of $n^{-1/2}$. The asymptotic distribution of $n^{1/2}\left(\hat\xi - \xi, \hat\omega - \omega, \hat\alpha - \alpha, \hat\kappa\right)'$ is the same as the distribution of

$$\begin{pmatrix} Z_1 \\ Z_2 \\ Z_3 \\ Z_4 \end{pmatrix} \delta\left(Z_4 > 0\right) + \begin{pmatrix} Z_1 - \left(I_{SN}^{\xi\kappa}/I_{SN}^{\kappa\kappa}\right)Z_4 \\ Z_2 - \left(I_{SN}^{\omega\kappa}/I_{SN}^{\kappa\kappa}\right)Z_4 \\ Z_3 - \left(I_{SN}^{\alpha\kappa}/I_{SN}^{\kappa\kappa}\right)Z_4 \\ 0 \end{pmatrix} \delta\left(Z_4 < 0\right),$$

where $(Z_1, Z_2, Z_3, Z_4)' \sim N\left(0, \left(I^{SN}\right)^{-1}\right)$, the inverse information $\left(I^{SN}\right)^{-1}$ has components denoted by $I_{SN}^{\xi\xi}$ and so forth, and $\delta(\varpi)$ is an indicator function which takes the value 1 when ϖ holds and the value 0 otherwise.

Proposition 4.1 describes the weak convergence properties of the maximum likelihood estimator $\left(\hat\xi, \hat\omega, \hat\alpha, \hat\kappa\right)$. The asymptotic distribution of $n^{1/2}\left(\hat\xi - \xi, \hat\omega - \omega, \hat\alpha - \alpha\right)$ is given by a mixture of correlated normal variables, whereas $n^{1/2}\hat\kappa$ asymptotically takes value 0 with probability $1/2$ and otherwise is distributed as a half normal variable. Therefore, $n^{1/2}\left(\hat\xi - \xi, \hat\omega - \omega, \hat\alpha - \alpha, \hat\kappa\right)$ is not asymptotically unbiased, but rather has asymptotic bias of order $O(1)$ and has variance matrix different than the usual inverse information matrix.

Another boundary case for the St and SN models occurs when $\alpha = \infty$, in which case, the results of Appendix A.3 show that the expected information matrix is singular. By finding a reparameterization that has an invertible information matrix, the results of Self and Liang (1987) could be applied to obtain the asymptotic distributions of the maximum likelihood estimators. The boundary case $\alpha = -\infty$ could be handled similarly.

5. The centered parameterization

The reparameterization with $\kappa = 1/\nu$ does not remedy the singularity of the information matrix that arises from the linear dependence between the score components for ξ and α when the distribution is normal. This problem can be alleviated, however, by adapting the centered parameterization developed by Azzalini (1985), Azzalini and Capitanio (1999), Chiogna (2005), and Arellano-Valle and Azzalini (2008) for the SN model. In the case of the St distribution, the centered parameterization is $\left(\mu, \sigma^2, \gamma_1, \gamma_2\right)$ where μ is the mean, σ^2 is the variance, and γ_1 and γ_2 are the third and fourth standardized cumulants, respectively, which are indices of skewness and kurtosis. The condition $\nu > 4$ is necessary to ensure that cumulants up to the fourth order exist. However, for both the SN distribution and the normal distribution, which is the case of primary interest in the present section, this restriction on ν is satisfied.

Formulae that express $\left(\mu, \sigma^2, \gamma_1, \gamma_2\right)$ in terms of $\left(\xi, \omega, \alpha, \kappa\right)$ for the St model can be obtained from expressions derived by Azzalini and Capitanio (2003), who derived the first four cumulants of the St model using the parameterization $\left(\xi, \omega, \alpha, \nu\right)$. Thus,

$$\mu = \xi + \omega b_\kappa \delta, \quad \sigma^2 = \omega^2 \lambda_{2,\kappa}, \quad \gamma_1 = \frac{\lambda_{3,\kappa}}{\lambda_{2,\kappa}^{3/2}}, \quad \gamma_2 = \frac{\lambda_{4,\kappa}}{\lambda_{2,\kappa}^2} - 3,$$

where

$$b_\kappa = \left(\frac{1}{\kappa\pi}\right)^{1/2} \frac{\Gamma\left\{1/(2\kappa) - 1/2\right\}}{\Gamma\left\{1/(2\kappa)\right\}}, \quad \delta = \frac{\alpha}{\left(1+\alpha^2\right)^{1/2}}, \quad \lambda_{2,\kappa} = \frac{1}{1-2\kappa} - b_\kappa^2 \delta^2,$$

$$\lambda_{3,\kappa} = b_\kappa \delta \left\{ \frac{\left(3-\delta^2\right)}{1-3\kappa} - \frac{3}{1-2\kappa} + 2b_\kappa^2 \delta^2 \right\},$$

$$\lambda_{4,\kappa} = \frac{3}{(1-2\kappa)(1-4\kappa)} - \frac{4b_\kappa^2 \delta^2 \left(3-\delta^2\right)}{1-3\kappa} + \frac{6b_\kappa^2 \delta^2}{1-2\kappa} - 3b_\kappa^4 \delta^4.$$

An expression for the Jacobian $\partial\left(\mu, \sigma^2, \gamma_1, \gamma_2\right)/\partial\left(\xi, \omega, \alpha, \nu\right)$ given by Azzalini (2008, personal communication) yields

$$D = \frac{\partial\left(\mu, \sigma^2, \gamma_1, \gamma_2\right)}{\partial\left(\xi, \omega, \alpha, \kappa\right)} = \begin{pmatrix} 1 & 0 & 0 & 0 \\ b_\kappa \delta & 2\omega\lambda_{2,\kappa} & 0 & 0 \\ \omega b_\kappa \delta' & -2\omega^2 b_\kappa^2 \delta\delta' & \dfrac{\partial\gamma_1}{\partial\delta}\delta' & \dfrac{\partial\gamma_2}{\partial\delta}\delta' \\ -\omega b_\kappa \delta \dfrac{1}{\kappa^2} q\left(\dfrac{1}{\kappa}\right) & \omega^2 \dfrac{\partial\lambda_{2,\kappa}}{\partial\kappa} & \dfrac{\partial\gamma_1}{\partial\kappa} & \dfrac{\partial\gamma_2}{\partial\kappa} \end{pmatrix},$$

where $\delta' = \left(1+\alpha^2\right)^{-3/2}$ and $q(x) = \frac{1}{2}\left\{1/x + \Psi(x/2 - 1/2) - \Psi(x/2)\right\}$. Expressions for the derivatives that appear in D are given in Appendix C. If $S(y)$ and

I are the score function and the expected information matrix of the parameterization $(\xi, \omega, \alpha, \kappa)$, then $(\mu, \sigma^2, \gamma_1, \gamma_2)$ has score function $D^{-1}S(y)$, and the information matrix is $D^{-1}I(D^{-1})'$.

The case $\kappa = 0$, for which the St distribution reduces to the SN, corresponds to the value

$$\gamma_2 = 2(\pi - 3)\left(\frac{2\delta^2}{\pi - 2\delta^2}\right)^2,$$

which is at the boundary of the parameter space for γ_2. Hence, results similar to those presented in Proposition 4.1 can be derived that provide the asymptotic distribution of $n^{1/2}(\hat{\mu} - \mu, \hat{\sigma}^2 - \sigma^2, \hat{\gamma}_1 - \gamma_1, \hat{\gamma}_2 - \gamma_2)$ under the SN distribution.

Calculations summarized in Appendix C show that, when $\alpha = \kappa = 0$, i.e., when the distribution is normal, the components of the score function for $(\mu, \sigma^2, \gamma_1, \gamma_2)$ become

$$S_\mu^N(y) = \frac{z}{\omega}, \quad S_{\sigma^2}^N(y) = \frac{z^2 - 1}{2\omega^2}, \quad S_{\gamma_1}^N(y) = \frac{z^3 - 3z}{6}, \quad S_{\gamma_2}^N(y) = \frac{1}{24}\left(z^4 - 6z^2 + 3\right),$$

while the information matrix for $(\mu, \sigma^2, \gamma_1, \gamma_2)$ is

$$I^N = \operatorname{diag}\left(\frac{1}{\omega^2}, \quad \frac{1}{2\omega^2}, \quad \frac{1}{6}, \quad \frac{1}{24}\right),$$

which is non-singular. Since the normal distribution corresponds to $\gamma_1 = \gamma_2 = 0$, which is on the boundary of the parameter space, the results of Self and Liang (1987) yield the following proposition.

PROPOSITION 5.1. When the distribution is normal, the estimator $(\hat{\mu}, \hat{\sigma}^2, \hat{\gamma}_1, \hat{\gamma}_2)$ is consistent and converges with rate $n^{-1/2}$. Asymptotically, $n^{1/2}(\hat{\mu} - \mu, \hat{\sigma}^2 - \sigma^2, \hat{\gamma}_1 - \gamma_1)'$ has the $N(0, \operatorname{diag}(\omega^2, 2\omega^2, 6))$ distribution, and $\hat{\gamma}_2$ is distributed as $Z\delta(Z > 0)$, where $Z \sim N(0, 24)$, independently of $\hat{\mu}$, $\hat{\sigma}^2$, and $\hat{\gamma}_1$.

Proposition 5.1 shows that, under the normal distribution, the asymptotic distribution of $n^{1/2}(\hat{\mu} - \mu, \hat{\sigma}^2 - \sigma^2, \hat{\gamma}_1 - \gamma_1)$ is the same as that of $n^{1/2}(\tilde{\mu} - \mu, \tilde{\sigma}^2 - \sigma^2, \tilde{\gamma}_1 - \gamma_1)$, where $\tilde{\mu}$ is the sample mean, $\tilde{\sigma}^2$ is the sample variance, and $\tilde{\gamma}_1$ is the sample index of skewness. The asymptotic distribution of $n^{1/2}(\hat{\gamma}_2 - \gamma_2)$ is a mixture of a degenerate distribution at 0 and a half-normal distribution, where the mixing probabilities are both equal to $1/2$. Note that $\hat{\mu}$, $\hat{\sigma}^2$, $\hat{\gamma}_1$, and $\hat{\gamma}_2$ are asymptotically independent.

6. The regression framework

The St model can be applied in the regression framework by assuming a St distribution for the response variable.

Consider the situation where the regression model is expressed in terms of the location parameter ξ; thus, $Y_1, \ldots, Y_n$ is assumed to be a sample generated by a regression model having St errors:

$$Y_i = \beta_0 + \beta_1 x_{i1} + \cdots + \beta_p x_{ip} + \epsilon_i, \qquad \epsilon_i \sim St(0, \omega, \alpha, \nu), \tag{6.1}$$

where $x_{i1}, \ldots, x_{ip}$ are covariate values and $\beta_0, \ldots, \beta_p$ are the regression coefficients. Equivalently, $Y_i \sim St(\xi_i, \omega, \alpha, \nu)$ for $i = 1, \ldots, n$, where $\xi_i = x_i'\beta$, $x_i = \big(x_{i1}, \ldots, x_{ip}\big)'$, and $\beta = \big(\beta_0, \ldots, \beta_p\big)'$.

The score function $S(y_i) = \{S_{\beta_0}(y_i), \ldots, S_{\beta_p}(y_i), S_\omega(y_i), S_\alpha(y_i), S_\nu(y_i)\}'$ for the single observation y_i has components

$$S_{\beta_0}(y_i) = \frac{z_i\tau_i^2}{\omega} - \frac{\nu\alpha\tau_i w_i}{\omega\big(\nu + z_i^2\big)}, \quad S_{\beta_j}(y_i) = \frac{z_i\tau_i^2}{\omega}x_{ij} - \frac{\nu\alpha\tau_i w_i}{\omega\big(\nu + z_i^2\big)}x_{ij}, \quad (j = 1, \ldots, p),$$

$$S_\omega(y_i) = -\frac{1}{\omega} + \frac{z_i^2\tau_i^2}{\omega} - \frac{\alpha\tau_i z_i \nu}{\omega(\nu + z_i^2)}w_i, \qquad S_\alpha(y_i) = z_i\tau_i w_i,$$

$$S_\nu(y_i) = \frac{1}{2}\Bigg\{\Psi\left(\frac{\nu}{2} + 1\right) - \Psi\left(\frac{\nu}{2}\right) - \frac{2\nu + 1}{\nu(\nu + 1)} - \ln\left(1 + \frac{z_i^2}{\nu}\right) + \frac{z_i^2\tau_i^2}{\nu} + \frac{\alpha z_i\big(z_i^2 - 1\big)}{\big(\nu + z_i^2\big)^2\tau_i}w_i + \frac{\gamma_i}{T(\alpha z_i\tau_i; \nu + 1)}\Bigg\},$$

where

$$z_i = \frac{1}{\omega}(y_i - \beta_0 - \beta_1 x_{i1} - \cdots - \beta_p x_{ip}), \quad \tau_i = \left(\frac{\nu + 1}{\nu + z_i^2}\right)^{1/2}, \quad w_i = \frac{t(\alpha z_i\tau_i; \nu + 1)}{T(\alpha z_i\tau_i; \nu + 1)},$$

and

$$\gamma_i = \int_{-\infty}^{\alpha z_i\tau_i}\left\{\frac{(\nu + 2)u^2}{(\nu + 1)(\nu + 1 + u^2)} - \ln\left(1 + \frac{u^2}{\nu + 1}\right)\right\}t(u; \nu + 1)du.$$

The score components $S_\omega(y_i)$, $S_\alpha(y_i)$, $S_\nu(y_i)$ have the same expressions as those given in Section 2 with z replaced by z_i. The components of the observed information matrix are given in Appendix A.4.

Appendix A. Information matrix

Appendix A.1. Observed information matrix of the St model

Let $w = t(\varsigma; \nu + 1)/T(\varsigma; \nu + 1)$; then,

$$w_z = \frac{\partial w}{\partial z} = -\frac{\nu(\nu + 2)\alpha^2 z w}{(\nu + z^2 + \alpha^2 z^2)\big(\nu + z^2\big)} - \frac{\nu\alpha\tau w^2}{\nu + z^2},$$

$$w_\alpha = \frac{\partial w}{\partial \alpha} = -\frac{(\nu+2)\alpha z^2 w}{(\nu + z^2 + \alpha^2 z^2)} - z\tau w^2,$$

$$w_\nu = \frac{\partial w}{\partial \nu} = \frac{w}{2}\left\{\frac{(\nu+2)\alpha^2 z^2}{(\nu+z^2+\alpha^2 z^2)\left(\nu+z^2\right)} - \ln\left(1+\frac{\alpha^2 z^2}{\nu+z^2}\right) - \frac{\gamma}{T(\varsigma;\nu+1)}\right\}$$
$$+ \frac{\alpha z\left(1-z^2\right)w^2}{2\tau\left(\nu+z^2\right)^2}.$$

Furthermore, with $f(y) = f(y;\xi,\omega,\alpha,\nu)$ and

$$S_{zz}(y) = \frac{\partial^2 \ln f(y)}{\partial z \partial z}, \qquad S_{z\alpha}(y) = \frac{\partial^2 \ln f(y)}{\partial z \partial \alpha}, \qquad S_{\alpha\alpha}(y) = \frac{\partial^2 \ln f(y)}{\partial \alpha \alpha},$$

and so forth, we have

$$S_{zz}(y) = \frac{2\tau^2 z^2}{\nu+z^2} - \tau^2 - \frac{3\alpha\tau\nu z w}{(\nu+z^2)^2} + \frac{\alpha\tau\nu w_z}{\nu+z^2}, \qquad S_{z\alpha}(y) = \frac{\nu\tau\left(w+\alpha w_\alpha\right)}{\nu+z^2},$$

$$S_{z\nu}(y) = \frac{z(1-z^2)}{(\nu+z^2)^2} + \frac{\alpha\left\{\nu(3z^2-1)+2z^2\right\}w}{2\tau(\nu+z^2)^3} + \frac{\alpha\tau\nu w_\nu}{\nu+z^2},$$

$$S_{\alpha\alpha}(y) = z\tau w_\alpha, \qquad S_{\alpha\nu}(y) = \frac{z\left(z^2-1\right)w}{2\tau(\nu+z^2)^2} + z\tau w_\nu,$$

and

$$S_{\nu\nu}(y) = \frac{1}{4}\left\{\Psi_1\left(\frac{\nu}{2}+1\right) - \Psi_1\left(\frac{\nu}{2}\right)\right\} + \frac{2\nu^2+2\nu+1}{2\nu^2(\nu+1)^2} + \frac{z^2}{2\nu\left(\nu+z^2\right)}$$
$$- \frac{z^2\left(\nu^2+2\nu+z^2\right)}{2\nu^2\left(\nu+z^2\right)^2} - \frac{\alpha z\left(z^2-1\right)\left(z^2+4\nu+3\right)w}{4\tau(\nu+1)\left(\nu+z^2\right)^3} + \frac{\alpha z\left(1-\tau^2\right)w_\nu}{2\tau\left(\nu+z^2\right)}$$
$$- \frac{\gamma^2}{4T(\varsigma;\nu+1)^2} - \frac{\alpha z\left(z^2-1\right)\gamma w}{4T(\varsigma;\nu+1)\tau\left(\nu+z^2\right)^2} + \frac{2\delta+\beta}{4T(\varsigma;\nu+1)}$$
$$+ \frac{\alpha z\left(z^2-1\right)w}{4\tau\left(\nu+z^2\right)^2}\left\{\frac{(\nu+2)\alpha^2 z^2}{(\nu+1)(\nu+z^2+\alpha^2 z^2)} - \ln\left(1+\frac{\alpha^2 z^2}{\nu+z^2}\right)\right\},$$

where

$$\beta = \int_{-\infty}^{\varsigma}\left\{\frac{(\nu+2)u^2}{(\nu+1)(\nu+1+u^2)} - \ln\left(1+\frac{u^2}{\nu+1}\right)\right\}^2 t(u;\nu+1)du,$$

and

$$\delta = \int_{-\infty}^{\varsigma}\frac{(\nu u^2-2\nu-2)u^2}{(\nu+1)^2(\nu+1+u^2)^2}t(u;\nu+1)du.$$

The second derivatives with respect to ξ and ω are given by

$$S_{\xi\xi}(y) = \frac{1}{\omega^2}S_{zz}(y), \qquad S_{\xi\omega}(y) = \frac{z}{\omega^2}S_{zz}(y) + \frac{1}{\omega^2}S_z(y), \qquad S_{\xi\alpha}(y) = -\frac{1}{\omega}S_{z\alpha}(y),$$

$$S_{\xi\nu}(y) = -\frac{1}{\omega}S_{z\nu}(y), \qquad S_{\omega\omega}(y) = \frac{1}{\omega^2} + \frac{z^2}{\omega^2}S_{zz}(y) + \frac{2z}{\omega^2}S_z(y), \qquad (A.1)$$

$$S_{\omega\alpha}(y) = -\frac{z}{\omega}S_{z\alpha}(y), \qquad S_{\omega\nu}(y) = -\frac{z}{\omega}S_{z\nu}(y).$$

where $S_z(y) = \partial \ln f(y)/\partial z = -\tau^2 z + \alpha\tau\nu w/(\nu+z^2)$.

Appendix A.2. Score function and information matrix under the location-scale Student's t-model

The score components for ξ, ω, and α under the location-scale Student's t-distribution, are obtained by setting $\alpha = 0$ in (2.1). The score component for the degrees of freedom, (2.2), is obtained as

$$S_\nu(y) = \frac{1}{2}\left\{\Psi\left(\frac{\nu}{2}+1\right) - \Psi\left(\frac{\nu}{2}\right) - \frac{2\nu+1}{\nu(\nu+1)} - \ln\left(1+\frac{z^2}{\nu}\right) + \frac{z^2\tau^2}{\nu}\right\} + \gamma^0,$$

where

$$\begin{aligned}\gamma^0 &= \int_{-\infty}^{0}\left\{\frac{(\nu+2)u^2}{(\nu+1)(\nu+1+u^2)} - \ln\left(1+\frac{u^2}{\nu+1}\right)\right\}t(u;\nu+1)du \\ &= \frac{1}{2}\left\{\Psi\left(\frac{\nu}{2}+\frac{1}{2}\right) - \Psi\left(\frac{\nu}{2}+1\right) + \frac{1}{\nu+1}\right\},\end{aligned}$$

and by simplification we obtain formula (2.2).

Under the location-scale Student's t-distribution, the components of the information matrix corresponding to ξ, ω, and ν are the same as those for the location-scale Student's t-model. Thus, to derive (2.3), we only need to derive the information components involving α. In particular,

$$S^t_{z\alpha}(y) = \frac{\nu\tau w^t}{\nu+z^2}, \qquad S^t_{\alpha\alpha}(y) = -\left(z\tau w^t\right)^2, \qquad S^t_{\alpha\nu}(y) = \frac{z\left(z^2-1\right)w^t}{2\tau(\nu+z^2)^2} - z\tau w^t\gamma_0,$$

where $w^t = 2t(0;\nu+1)$. It follows from (A.1) that the required components of the information matrix are $I^t_{\omega\alpha} = I^t_{\alpha\nu} = 0$,

$$I^t_{\xi\alpha} = \frac{2\nu(\nu+1)^{1/2}t(0;\nu+1)}{\omega}E\left\{\frac{1}{\left(\nu+Z^2\right)^{3/2}}\right\}, \quad I^t_{\alpha\alpha} = 4(\nu+1)t(0;\nu+1)^2E\left(\frac{Z^2}{\nu+Z^2}\right),$$

where Z has Student's t-distribution with ν degrees of freedom.

Appendix A.3. Singularity of information matrices as $\alpha \to \pm\infty$ in the St and SN models

Standard calculations show that

$$\lim_{\alpha\to\infty} \alpha w = -\frac{\nu+1}{z\tau}$$

for $z < 0$ and that the limit is 0 for $z > 0$. Consequently, w is of order $O(\alpha^{-1})$ or smaller as $\alpha \to \pm\infty$ for $z \neq 0$. Moreover,

$$\lim_{\alpha\to\pm\infty} \frac{\gamma}{T(\varsigma;\nu+1)}w = 0.$$

These results imply

$$\lim_{\alpha\to\pm\infty} w_\alpha = \lim_{\alpha\to\pm\infty} w_\nu = \lim_{\alpha\to\pm\infty} (w + \alpha w_\alpha) = 0.$$

Thus, $S_{z\alpha}(y)$, $S_{\alpha\alpha}(y)$, and $S_{\alpha\nu}(y)$ all tend to 0 as $\alpha \to \pm\infty$, for $y \neq \xi$, and by (A.1), $S_{\xi\alpha}(y)$ and $S_{\omega\alpha}(y)$ also converge to 0. It follows that, with probability one, the observed information matrix is singular in the St model as $\alpha \to \pm\infty$, so the expected information matrix is also singular in the limit. Note that the observed information matrix evaluated at the maximum likelihood estimate is singular whenever $\hat{\alpha} = \infty$.

In the case of the SN model,

$$S_{\xi\alpha}^{SN}(y) = -\frac{1}{\omega} w_\phi + \frac{\alpha^2}{\omega} z^2 w_\phi + \frac{\alpha}{\omega} z w_\phi^2, \qquad S_{\omega\alpha}^{SN}(y) = -\frac{1}{\omega} z w_\phi + \frac{\alpha^2}{\omega} z^3 w_\phi + \frac{\alpha}{\omega} z^2 w_\phi^2,$$

$$S_{\alpha\alpha}^{SN}(y) = -\alpha z^3 w_\phi - z^2 w_\phi^2,$$

and it can be shown that, for $z > 0$,

$$\lim_{\alpha\to\infty} w_\phi = \lim_{\alpha\to\infty} \alpha w_\phi = \lim_{\alpha\to\infty} \alpha^2 w_\phi = \lim_{\alpha\to\infty} \alpha w_\phi^2 = 0.$$

Consequently, $S_{\xi\alpha}^{SN}(y)$, $S_{\omega\alpha}^{SN}(y)$, and $S_{\alpha\alpha}^{SN}(y)$ vanish when $z > 0$ as $\alpha \to \infty$. Since $z > 0$ with probability tending to 1 as $\alpha \to \infty$, the expected information matrix is singular. A similar argument shows that the expected information matrix is also singular as $\alpha \to -\infty$. Note that the observed information matrix evaluated at the maximum likelihood estimate is not necessarily singular for finite α, even when $\hat{\alpha} = \infty$.

Appendix A.4. Observed information matrix in the regression framework

In the regression model (6.1), the second derivatives with respect to z, α, and ν are

$$S_{zz}(y_i) = \frac{2\tau_i^2 z_i^2}{\nu + z_i^2} - \tau_i^2 - \frac{3\alpha\tau_i\nu z_i w_i}{(\nu + z_i^2)^2} + \frac{\alpha\tau_i\nu w_{zi}}{\nu + z_i^2},$$

$$S_{z\alpha}(y_i) = \frac{\nu\tau_i\big(w_i + \alpha w_{\alpha i}\big)}{\nu + z_i^2}, \quad S_{z\nu}(y_i) = \frac{z_i\big(1 - z_i^2\big)}{\big(\nu + z_i^2\big)^2} + \frac{\alpha\big\{\nu(3z_i^2 - 1) + 2z_i^2\big\} w_i}{2\tau_i(\nu + z_i^2)^3} + \frac{\alpha\tau_i\nu w_{\nu i}}{\nu + z_i^2},$$

$$S_{\alpha\alpha}(y_i) = z_i\tau w_{\alpha i}, \quad S_{\alpha\nu}(y_i) = \frac{z_i\big(z_i^2 - 1\big)w_i}{2\tau_i(\nu + z_i^2)^2} + z_i\tau w_{\nu i},$$

$$\begin{aligned}
S_{\nu\nu}(y_i) = \frac{1}{4}\left\{\Psi_1\left(\frac{\nu}{2} + 1\right) - \Psi_1\left(\frac{\nu}{2}\right)\right\} + \frac{2\nu^2 + 2\nu + 1}{2\nu^2(\nu + 1)^2} + \frac{z_i^2}{2\nu\big(\nu + z_i^2\big)} \\
- \frac{z_i^2\big(\nu^2 + 2\nu + z_i^2\big)}{2\nu^2\big(\nu + z_i^2\big)^2} - \frac{\alpha z_i\big(z_i^2 - 1\big)\big(z_i^2 + 4\nu + 3\big)w_i}{4\tau_i(\nu + 1)\big(\nu + z_i^2\big)^3} + \frac{\alpha z_i\big(1 - \tau_i^2\big)w_{\nu i}}{2\tau_i\big(\nu + z_i^2\big)} \\
- \frac{\gamma_i^2}{4T(\alpha z_i\tau_i, \nu + 1)^2} - \frac{\alpha z_i\big(z_i^2 - 1\big)\gamma_i w_i}{4T(\alpha z_i\tau_i, \nu + 1)\tau_i\big(\nu + z_i^2\big)^2} + \frac{2\delta_i + \beta_i}{4T(\alpha z_i\tau_i, \nu + 1)} \\
+ \frac{\alpha z_i\big(z_i^2 - 1\big)w_i}{4\tau_i\big(\nu + z_i^2\big)^2}\left\{\frac{(\nu + 2)\alpha^2 z_i^2}{(\nu + 1)(\nu + z_i^2 + \alpha^2 z_i^2)} - \ln\left(1 + \frac{\alpha^2 z_i^2}{\nu + z_i^2}\right)\right\},
\end{aligned}$$

where

$$w_{zi} = \frac{\partial w_i}{\partial z_i} = -\frac{\nu(\nu + 2)\alpha^2 z_i w_i}{(\nu + z_i^2 + \alpha^2 z_i^2)\big(\nu + z_i^2\big)} - \frac{\nu\alpha\tau_i w_i^2}{\nu + z_i^2},$$

$$w_{\alpha i} = \frac{\partial w_i}{\partial \alpha} = -\frac{(\nu+2)\alpha z_i^2 w_i}{(\nu + z_i^2 + \alpha^2 z_i^2)} - z_i \tau_i w_i^2,$$

$$w_{\nu i} = \frac{\partial w_i}{\partial \nu} = \frac{w_i}{2}\left\{\frac{(\nu+2)\alpha^2 z_i^2}{(\nu + z_i^2 + \alpha^2 z_i^2)\left(\nu + z_i^2\right)} - \ln\left(1 + \frac{\alpha^2 z_i^2}{\nu + z_i^2}\right) - \frac{\gamma_i}{T(\alpha z_i \tau_i, \nu+1)}\right\} + \frac{\alpha z_i\left(1 - z_i^2\right) w_i^2}{2\tau\left(\nu + z_i^2\right)^2}.$$

$$\beta_i = \int_{-\infty}^{\alpha z_i \tau_i} \left\{\frac{(\nu+2)u^2}{(\nu+1)(\nu+1+u^2)} - \ln\left(1 + \frac{u^2}{\nu+1}\right)\right\}^2 t(u;\nu+1)du,$$

and

$$\delta_i = \int_{-\infty}^{\alpha z_i \tau_i} \frac{(\nu u^2 - 2\nu - 2)u^2}{(\nu+1)^2(\nu+1+u^2)^2} t(u;\nu+1)du.$$

The second derivatives with respect to the $\beta_0, \ldots, \beta_p$, and ω are given by

$$S_{\beta_0\beta_0}(y_i) = \frac{1}{\omega^2}S_{zz}(y_i), \qquad S_{\beta_0\beta_j}(y_i) = S_{\beta_0\beta_0}(y_i)x_{ij}, \qquad S_{\beta_j\beta_r}(y_i) = S_{\beta_0\beta_j}(y_i)x_{ir},$$

$$S_{\beta_0\omega}(y_i) = \frac{z_i}{\omega^2}S_{zz}(y_i) + \frac{1}{\omega^2}S_z(y_i), \qquad S_{\beta_0\alpha}(y_i) = -\frac{1}{\omega}S_{z\alpha}(y_i),$$

$$S_{\beta_0\nu}(y_i) = -\frac{1}{\omega}S_{z\nu}(y_i), \qquad S_{\beta_j\omega}(y_i) = S_{\beta_0\omega}(y_i)x_{ij}, \qquad S_{\beta_j\alpha}(y_i) = S_{\beta_0\alpha}(y_i)x_{ij},$$

$$S_{\beta_j\nu}(y_i) = S_{\beta_0\nu}(y_i)x_{ij}, \qquad S_{\omega\omega}(y_i) = \frac{1}{\omega^2} + \frac{z^2}{\omega^2}S_{zz}(y_i) + \frac{2z_i}{\omega^2}S_z(y_i),$$

$$S_{\omega\alpha}(y_i) = -\frac{z_i}{\omega}S_{z\alpha}(y_i), \qquad S_{\omega\nu}(y_i) = -\frac{z_i}{\omega}S_{z\nu}(y_i),$$

where $S_z(y_i) = -z_i\tau_i^2 + \alpha\tau_i\nu w_i/(\nu + z_i^2)$.

Appendix B. Reparameterization by reciprocal degrees of freedom

The score component for κ is $S_\kappa(y) = S_\nu(y)(\partial\nu/\partial\kappa)$ and its expression is given in (4.1). Taylor expansion of (4.1) about $\kappa = 0$ yields

$$S_\kappa(y) = \frac{1}{4}\left(z^4 - 2z^2 - 2\right) - \frac{1}{2}\alpha z\left(z^2 - 1\right)\frac{t(\varsigma; 1/\kappa + 1)}{T(\varsigma; 1/\kappa + 1)} - \frac{\gamma}{2\kappa^2 T(\varsigma, 1/\kappa + 1)} + O(\kappa),$$

and formula (4.2) follows, since

$$\lim_{\kappa\to 0}\frac{\gamma}{\kappa^2} = \frac{1}{2}\left\{\alpha z(1 + \alpha^2 z^2)\phi(\alpha z) - \Phi(\alpha z)\right\}. \qquad (B.1)$$

In terms of the notation introduced in Appendix A,

$$S_{z\kappa}(y) = \frac{z(z^2-1)}{(1+z^2\kappa)^2} - \frac{\alpha\left(3z^2 - 1 + 2z^2\kappa\right)w}{2(1+\kappa)^{1/2}\left(1+z^2\kappa\right)^{5/2}} + \frac{\alpha(1+\kappa)^{1/2}w_\kappa}{\left(1+z^2\kappa\right)^{3/2}},$$

$$S_{\alpha\kappa}(y) = \frac{z\left(1-z^2\right)w}{2(1+\kappa)^{1/2}\left(1+z^2\kappa\right)^{3/2}} + \frac{z(1+\kappa)^{1/2}w_\kappa}{\left(1+z^2\kappa\right)^{1/2}},$$

and

$$
\begin{aligned}
S_{\kappa\kappa}(y) &= S_{\nu\nu}(y)(\partial\nu/\partial\kappa)^2 + S_\nu(y)(\partial^2\nu/\partial\kappa^2) \\
&= \frac{1}{\kappa^4}\Bigg[\frac{1}{4}\Psi_1\Big(\frac{1}{2\kappa}+1\Big) - \frac{1}{4}\Psi_1\Big(\frac{1}{2\kappa}\Big) + \frac{\big(2+2\kappa+\kappa^2\big)\kappa^2}{2(1+\kappa)^2} + \frac{z^2\kappa^2}{2\big(1+z^2\kappa\big)} \\
&\qquad - \frac{z^2\big(1+2\kappa+z^2\kappa^2\big)\kappa^2}{2\big(1+z^2\kappa\big)^2} - \frac{\alpha z\big(z^2-1\big)\big(4+z^2\kappa+3\kappa\big)\kappa^3 w}{4(1+\kappa)^{3/2}\big(1+z^2\kappa\big)^{5/2}} \\
&\qquad - \frac{\alpha z\kappa^4\big(z^2-1\big)w_\kappa}{2(1+\kappa)^{1/2}\big(1+z^2\kappa\big)^{3/2}} - \frac{\gamma^2}{4T(\varsigma,1/\kappa+1)^2} \\
&\qquad - \frac{\alpha z\kappa^2\big(z^2-1\big)\gamma w}{4T(\varsigma,\nu+1)(1+\kappa)^{1/2}\big(1+z^2\kappa\big)^{3/2}} + \frac{2\delta+\beta}{4T(\varsigma,1/\kappa+1)} \\
&\qquad + \frac{\alpha z\kappa^2\big(z^2-1\big)w}{4(1+\kappa)^{1/2}\big(1+z^2\kappa\big)^{3/2}}\Bigg\{\frac{\alpha^2 z^2(1+2\kappa)\kappa}{(1+\kappa)\big(1+z^2\kappa+\alpha^2 z^2\kappa\big)} \\
&\qquad - \ln\Big(1+\frac{\alpha^2 z^2\kappa}{1+z^2\kappa}\Big)\Bigg\}\Bigg] \\
&\quad + \frac{1}{\kappa^3}\Bigg\{\Psi\Big(\frac{1}{2\kappa}+1\Big) - \Psi\Big(\frac{1}{2\kappa}\Big) - \frac{\kappa(2+\kappa)}{(1+\kappa)} - \ln\big(1+\kappa z^2\big) + \frac{\kappa(1+\kappa)z^2}{\big(1+\kappa z^2\big)} \\
&\qquad + \frac{\alpha z\big(z^2-1\big)\kappa^2 w}{(1+\kappa)^{1/2}\big(1+\kappa z^2\big)^{3/2}} + \frac{\gamma}{T(\varsigma,1/\kappa+1)}\Bigg\},
\end{aligned}
$$

where

$$
\begin{aligned}
w_\kappa = \frac{\partial w}{\partial\kappa} = -\frac{w}{2\kappa^2}&\Bigg\{\frac{\alpha^2 z^2(1+2\kappa)\kappa}{(1+z^2\kappa)\big(1+z^2\kappa+\alpha^2 z^2\kappa\big)} - ln\Big(1+\frac{\alpha^2 z^2\kappa}{1+z^2\kappa}\Big) - \frac{\gamma}{T(\varsigma,\nu+1)}\Bigg\} \\
&+ \frac{\alpha z\big(z^2-1\big)w^2}{2(\kappa+1)^{1/2}\big(1+z^2\kappa\big)^{3/2}}.
\end{aligned}
$$

The observed information matrix for the parameter $\big(\xi,\omega,\alpha,\kappa\big)$ can then be obtained from expressions (A.1).

Taylor expansion of $S_{z\kappa}(y)$, $S_{\alpha\kappa}(y)$, and $S_{\kappa\kappa}(y)$ about $\kappa=0$ yields

$$
\begin{aligned}
S_{z\kappa}(y) =& z(z^2-1) - \frac{\alpha\big(3z^2-1\big)w}{2} - \frac{\alpha^3 z^2\big(4-2z^2-\alpha^2 z^2\big)w}{4} \\
&+ \frac{\alpha\gamma w}{2\kappa^2 T(\varsigma,1/\kappa+1)} + \frac{\alpha^2 z\big(z^2-1\big)w^2}{2} + O(\kappa),
\end{aligned}
$$

$$
\begin{aligned}
S_{\alpha\kappa}(y) = \frac{z\big(1-z^2\big)w}{2} - \frac{\alpha^2 z^3\big(4-2z^2-\alpha^2 z^2\big)w}{4} + \frac{z\gamma w}{2\kappa^2 T(\varsigma,1/\kappa+1)} \\
+ \frac{\alpha z^2\big(z^2-1\big)w^2}{2} + O(\kappa),
\end{aligned}
$$

$$
\begin{aligned}
S_{\kappa\kappa}(y) = & \frac{1}{6}\big(3-2z^6+3z^4\big)+\frac{\alpha z}{4}\big(3z^4-2z^2-1\big)w+\frac{\alpha^3 z^3\big(z^2-1\big)\big(3-z^2-\alpha^2 z^2\big)w}{4} \\
& -\frac{\alpha z\big(z^2-1\big)\gamma w}{2\kappa^2 T(\varsigma,\nu+1)}-\frac{\alpha^2 z^2\big(z^2-1\big)^2 w^2}{4}-\frac{\gamma^2}{4\kappa^4 T(\varsigma,1/\kappa+1)^2} \\
& +\frac{\beta}{4\kappa^4 T(\varsigma,1/\kappa+1)}+\frac{\delta+2\kappa\gamma}{2\kappa^4 T(\varsigma,1/\kappa+1)}+O(\kappa),
\end{aligned}
$$

since

$$
w_\kappa = -\frac{\alpha^2 z^2\big(4-2z^2-\alpha^2 z^2\big)w}{4}+\frac{\gamma w}{2\kappa^2 T(\varsigma,1/\kappa+1)}+\frac{\alpha z\big(z^2-1\big)w^2}{2}+O(\kappa).
$$

By taking the limit as $\kappa \to 0$ in the expansions for $S_{z\kappa}(y)$ and $S_{\alpha\kappa}(y)$, it follows from (B.1) that

$$
\begin{aligned}
S^{SN}_{z\kappa}(y) = & z(z^2-1)-\frac{\alpha\big(3z^2-1\big)w_\phi}{2}-\frac{\alpha^3 z^2\big(4-2z^2-\alpha^2 z^2\big)w_\phi}{4} \\
& +\frac{\alpha w_\phi}{4}\Big\{\alpha z(1+\alpha^2 z^2)w_\phi-1\Big\}+\frac{\alpha^2 z\big(z^2-1\big)w_\phi^2}{2},
\end{aligned}
$$

and

$$
\begin{aligned}
S^{SN}_{\alpha\kappa}(y) = & \frac{z\big(1-z^2\big)w_\phi}{2}-\frac{\alpha^2 z^3\big(4-2z^2-\alpha^2 z^2\big)w_\phi}{4}+\frac{z w_\phi}{4}\Big\{\alpha z(1+\alpha^2 z^2)w_\phi-1\Big\} \\
& +\frac{\alpha z^2\big(z^2-1\big)w_\phi^2}{2}.
\end{aligned}
$$

Similarly, the expansion for $S_{\kappa\kappa}(y)$ shows that

$$
\begin{aligned}
S^{SN}_{\kappa\kappa}(y) = & \frac{1}{6}\big(3z^4-2z^6\big)+\frac{1}{16}\alpha z\big(1-4z^2+12z^4\big)w_\phi \\
& -\frac{1}{48}\alpha^3 z^3\Big\{7-48z^2+12z^4-\alpha^2 z^2\big(19-12z^2\big)+3\alpha^4 z^4\Big\}w_\phi \\
& -\frac{1}{16}\alpha^2 z^2\Big\{1-4z^2+4z^4-2\alpha^2 z^2\big(1-2z^2\big)+\alpha^4 z^4\Big\}w_\phi^2,
\end{aligned}
$$

since

$$
\lim_{\kappa\to 0}\frac{\beta}{\kappa^4} = \frac{1}{4}\Big\{57\Phi(\alpha z)-\big(\alpha^7 z^7+3\alpha^5 z^5+19\alpha^3 z^3+57\alpha z\big)\phi(\alpha z)\Big\},
$$

and

$$
\lim_{\kappa\to 0}\frac{\delta+2\kappa\gamma}{\kappa^4} = -8\Phi(\alpha z)+\left(\frac{2}{3}\alpha^5 z^5+\frac{10}{3}\alpha^3 z^3+8\alpha z\right)\phi(\alpha z).
$$

Expression (4.3) follows by taking expected values of $S^{SN}_{z\kappa}(y)$, $S^{SN}_{\alpha\kappa}(y)$, and $S^{SN}_{\kappa\kappa}(y)$ and applying (A.1).

Appendix C. Centered Parametrization

The derivatives appearing in the matrix $D = \partial\big(\mu,\sigma^2,\gamma_1,\gamma_2\big)/\partial\big(\xi,\omega,\alpha,\kappa\big)$ are

$$
\frac{\partial\lambda_{2,\kappa}}{\partial\kappa} = 2\left\{\frac{1}{(1-2\kappa)^2}+\frac{b_\kappa^2\delta^2}{\kappa^2}q\left(\frac{1}{\kappa}\right)\right\},
$$

$$\frac{\partial\gamma_1}{\partial\delta} = \frac{\gamma_1}{\delta} - \frac{2b_\kappa\delta^2\{1/(1-3\kappa)-2b^2\}}{\lambda_{2,\kappa}^{3/2}} + \frac{3\gamma_1 b_\kappa^2\delta}{\lambda_{2,\kappa}},$$

$$\frac{\partial\gamma_2}{\partial\delta} = \frac{4b_\kappa^2\delta}{\lambda_{2,\kappa}}\left[\gamma_2 + 3 - \frac{1}{\lambda_{2,\kappa}}\left\{\frac{2(3-2\delta^2)}{1-3\kappa} - \frac{3}{1-2\kappa} + 3b_\kappa^2\delta^2\right\}\right],$$

$$\frac{\partial\gamma_1}{\partial\kappa} = -\frac{\gamma_1}{\kappa^2}q\left(\frac{1}{\kappa}\right) - \frac{b_\kappa\delta}{\lambda_{2,\kappa}^{3/2}}\left\{-\frac{3(3-\delta^2)}{(1-3\kappa)^2} + \frac{6}{(1-2\kappa)^2} + \frac{4b_\kappa^2\delta^2}{\kappa^2}q\left(\frac{1}{\kappa}\right)\right\} - \frac{3\gamma_1}{2\lambda_{2,\kappa}}\frac{\partial\lambda_{2,\kappa}}{\partial\kappa},$$

and

$$\begin{aligned}\frac{\partial\gamma_2}{\partial\kappa} = \frac{1}{\lambda_{2,\kappa}^2}\Bigg[&\frac{6(3-8\kappa)}{(1-2\kappa)^2(1-4\kappa)^2} + \frac{4b_\kappa^2\delta^2(3-\delta^2)}{(1-3\kappa)^2}\left[\left\{2q\left(\frac{1}{\kappa}\right)\frac{1}{\kappa}+1\right\}\left(\frac{1}{\kappa}-3\right) - \frac{1}{\kappa}\right]\\ &- \frac{6b_\kappa^2\delta^2}{(1-2\kappa)^2}\left[\left\{2q\left(\frac{1}{\kappa}\right)\frac{1}{\kappa}+1\right\}\left(\frac{1}{\kappa}-2\right) - \frac{1}{\kappa}\right] + \frac{12b_\kappa^4\delta^4}{\kappa^2}q\left(\frac{1}{\kappa}\right)\Bigg]\\ &- \frac{2\lambda_{4,\kappa}}{\lambda_{2,\kappa}^3}\frac{\partial\lambda_{2,\kappa}}{\partial\kappa}.\end{aligned}$$

To determine the score function and information matrix under SN and normal distributions, first take $\kappa = 0$, and now let $b = (2/\pi)^{1/2}$ and $\lambda_2 = 1 - b^2\delta^2$. Then, in obvious notation, the Jacobian D becomes

$$D^{SN} = \begin{pmatrix} 1 & 0 & 0 & 0 \\ b\delta & 2\omega\lambda_2 & 0 & 0 \\ \omega b\delta' & -\frac{4}{\pi}\omega^2\delta\delta' & \frac{3b\delta^2(4-\pi)\delta'}{\pi\lambda_2^{5/2}} & \frac{32\delta^3(1-3/\pi)\delta'}{\pi\lambda_2^{3,\kappa}} \\ \frac{3}{4}\omega b\delta & \omega^2\left(2-\frac{3}{\pi}\delta^2\right) & \frac{3a_1 b\delta}{4\pi\lambda_2^{5/2}} & -\frac{2a_2}{\pi^2\lambda_2^3} \end{pmatrix},$$

where $a_1 = 4\pi - 12\delta^2 - \delta^2\pi + 4\delta^4$ and $a_2 = 18\delta^2\pi - 36\delta^4 - 2\delta^4\pi + 12\delta^6 - 3\pi^2$.

The inverse $D_{SN} = \left(D^{SN}\right)^{-1} = \partial(\xi,\omega,\alpha,\kappa)/\partial(\mu,\sigma^2,\gamma_1,\gamma_2)$, necessary to calculate the score function of the centered parameterization under the SN distribution, is

$$D_{SN} = \begin{pmatrix} 1 & 0 & 0 & 0 \\ -\frac{b\delta}{2\omega\lambda_2} & \frac{1}{2\omega\lambda_2} & 0 & 0 \\ -\frac{\omega\lambda_2^{3/2}(2\delta^4-4\delta^2+\pi)}{\delta^2 c} & \frac{\omega\lambda_2^{3/2}ba_3}{3\delta c} & -\frac{\lambda_2^{3/2}ba_2}{6\delta^2\delta' c} & -\frac{8\lambda_2^{3/2}b\delta(\pi-3)}{3c} \\ -\frac{\omega\pi\lambda_2^2 b(\delta^2-2)}{4\delta c} & \frac{\omega\lambda_2^2(\delta^2-6+\pi)}{2c} & -\frac{\lambda_2^2 a_1}{8\delta\delta' c} & -\frac{\lambda_2^2(\pi-4)}{2c} \end{pmatrix}.$$

The score function of the centered parameterization under the SN distribution has components

$$S_\mu(y)^{SN} = \frac{z}{\omega} - \frac{\alpha w}{\omega}, \qquad S_{\sigma^2}^{SN}(y) = \frac{1}{2\omega^2\lambda_2}\{(z-\alpha w)(z-b\delta)-1\},$$

$$S_{\gamma_1}^{SN}(y) = \frac{\lambda_2^{3/2}}{c}\Big[-(z-\alpha w)\frac{2\delta^4-4\delta^2+\pi}{\delta^2} + (z^2-1-\alpha z w)\frac{ba_3}{3\delta} - zw\frac{ba_2}{6\delta^2\delta'} - \frac{2}{3}\big\{z^4-2z^2-1-\alpha z(2z^2+\alpha^2 z^2-1)w\big\}\frac{b\delta(\pi-3)}{3}\Big],$$

$$S_{\gamma_2}^{SN}(y) = \frac{\lambda_2^{2}}{c}\Big[-\frac{1}{4}b\pi(z-\alpha w)\frac{\delta^2-2}{\delta} + \frac{1}{2}(z^2-1-\alpha z w)(\delta^2-6+\pi) - \frac{1}{8}zw\frac{a_1}{\delta\delta'} - \frac{1}{8}\big\{z^4-2z^2-1-\alpha z(2z^2+\alpha^2 z^2-1)w\big\}(\pi-4)\Big].$$

By expanding these components with respect to α and eventually taking the limit as $\alpha \to 0$ yields the expressions for $S_\mu(y)^N$, $S_{\sigma^2}(y)^N$, $S_{\gamma_1}(y)^N$, and $S_{\gamma_2}(y)^N$ given in Section 5. The expected information matrix I^N is obtained by calculating the covariance matrix of these components under the normal model.

Acknowledgements: This research was partially supported by MIUR, Italy, with grant PRIN 2006132978. The numerical work was undertaken on the R package `sn` written by Azzalini and available at http://cran.r-project.org/src/contrib/PACKAGES.html.

References

Arellano-Valle, R. B., Azzalini, A. (2008), The centred parametrization for the multivariate skew-normal distribution, *Journal of Multivariate Analysis*, 99, 1362–1382.

Azzalini, A. (1985), A class of distributions which includes the normal ones, *Scandinavian Journal of Statistics*, 12, 171–178.

Azzalini, A. (1986), Further results on a class of distributions which includes the normal ones, *Statistica*, 46, 199–208.

Azzalini, A., Capitanio, A. (1999), Statistical applications of the multivariate skew normal distribution, *Journal of the Royal Statistical Society, Series B*, 61, 579–602.

Azzalini, A., Capitanio, A. (2003), Distributions generated by perturbation of symmetry with emphasis on a multivariate skew t distribution, Journal of the Royal Statistical Society, Series B, 65, 367–389.

Azzalini, A., Genton, M. G. (2008), Robust likelihood methods based on the skew-t and related distributions, *International Statistical Review*, 76, 106–29.

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

Chiogna, M. (2005), A note on the asymptotic distribution of the maximum likelihood estimator for the scalar skew-normal distribution, *Statistical Methods and Applications*, 14, 331–341.

DiCiccio, T. J., Monti, A. C. (2004), Inferential aspects in the skew exponential power distribution, *Journal of the American Statistical Association*, 99, 439–450.

Greco, L. (2011), Minimum Hellinger based inference for scalar skew-normal and skew-t distribution, *Test*, 20, 120-137

Rotnitzky, A., Cox, D. R., Bottai, M., Robins, J. (2000), Likelihood-based inference with singular information matrix, *Bernoulli*, 6, 243–284.

Sartori, N. (2006), Bias prevention of maximum likelihood estimates for scalar skew normal and skew t distributions, *Journal of Statistical Planning and Inference*, 136, 4259–4275.

Self, S. G., Liang K. Y. (1987), Asymptotic properties of maximum likelihood estimators and likelihood ratio test under nonstandard conditions. *Journal of the American Statistical Association*, 82, 605–610.