Inference in the Skew-t and Related Distributions

Anna Clara Monti

Pe.Me.Is Department, Statistical division

University of Sannio

Einaudi Institute for Economics and Finance

Roma 8 maggio 2008

Outline

- Brief overview of flexible models
- The skew-t model
- Problems connected to inference
- Reparameterizations
- Applications in robustness context

The basic lemma

- A continuous density function symmetric around 0: $g(y)$
- A skewing function $\pi : \mathbb{R}^d \to [0,1]$

$$f(y) = 2g(y-\xi)\pi(y-\xi)$$

is a proper density function.

Azzalini (1985) and further literature

Often

$$\pi(y) = G(w(y))$$

- $G(y)$ is a scalar distribution function such that $G(-y) = 1 - G(y)$
- $w(y): \mathbb{R}^d \to \mathbb{R}$ is such that $w(-y) = -w(y)$

The Skew Normal distribution

$$Y \sim SN(\xi, \Omega, \alpha)$$

$$f_{SN}(y;\xi,\Omega,\alpha) = 2\varphi_d(y-\xi;\Omega)\Phi\left(\alpha^T\omega^{-1}(y-\xi)\right)$$

Azzalini (1985), Azzalini & Dalla Valle (1996)

$$\omega = diag\left(\omega_{11}^{1/2},\ldots,\omega_{dd}^{1/2}\right)$$

<u>*When*</u>

$$Y \sim SN(\xi, \Omega, \alpha)$$

$$E(Y) = \xi + \omega\mu_z$$

$$Var(Z) = \Omega - \omega\mu_z\mu_z^T\omega$$

$$\mu_Z = b\delta$$

$$\delta = \frac{1}{\left(1+\alpha'\bar{\Omega}\alpha\right)^{1/2}}\bar{\Omega}\alpha$$

$$b = \left(\frac{2}{\pi}\right)^{1/2}$$

$$\bar{\Omega} = \omega^{-1}\Omega\omega^{-1}$$

<u>*In the scalar case*</u>

$$\gamma_1 = \frac{1}{2}(4-\pi)\left\{\frac{(b\delta)^2}{1-(b\delta)^2}\right\}^{3/2}$$

$$\gamma_2 = 2(\pi-3)\left\{\frac{(b\delta)^2}{1-(b\delta)^2}\right\}^2$$

$$\delta = \frac{\alpha}{\left(1+\alpha^2\right)^{1/2}}$$

$-.995 < \gamma_1 < .995$

Stocastic representations

Conditioning method

$$\begin{pmatrix} U_0 \\ U \end{pmatrix} \sim N_{d+1}\left(0, \Omega^*\right) \qquad \Omega^* = \begin{bmatrix} 1 & \delta \\ \delta & \bar{\Omega} \end{bmatrix}$$

$$U \mid U_0 > 0 \sim SN\left(0, \bar{\Omega}, \alpha\right) \qquad \alpha = \frac{1}{\left(1 - \delta' \bar{\Omega}^{-1} \delta\right)^{1/2}} \bar{\Omega}^{-1} \delta$$

Transformation method

$$\begin{pmatrix} U_0' \\ U' \end{pmatrix} \sim N_{d+1}\left(0, \begin{bmatrix} 1 & 0 \\ 0 & \Psi \end{bmatrix}\right) \qquad Z_j = \delta_j \left|U_0'\right| + \left(1 - \delta_j^2\right)^{1/2} U_j'$$

$$-1 < \delta_j < 1$$

$$\left(Z_1, \ldots, Z_d\right) \sim SN\left(0, \bar{\Omega}, \alpha\right)$$

Maximum (d=2)

$$\begin{pmatrix} U_0 \\ U_1 \end{pmatrix} \sim N_2\left(0, \begin{bmatrix} 1 & \rho \\ \rho & 1 \end{bmatrix}\right) \quad \Rightarrow \quad \max\left(U_0, U_1\right) \sim SN\left(0, 1, \alpha\right) \qquad \alpha = \left(\frac{1-\rho}{1+\rho}\right)^{1/2}$$

Main properties

Marginal distribution

$$Y = (Y_1, Y_2) \sim SN(0, \Omega, \alpha) \quad \Rightarrow \quad Y_1 \sim SN(0, \Omega_{11}, \tilde{\alpha}_1)$$

$$\tilde{\alpha}_1 = \frac{\alpha_1 + \Omega_{11}^{-1}\Omega_{12}\alpha_2}{\left(1 + \alpha_2^T \Omega_{22.1} \alpha_2\right)^{1/2}} \qquad \Omega_{22.1} = \Omega_{22} - \Omega_{21}\Omega_{11}^{-1}\Omega_{12}$$

Linear transformation

$$Y \sim SN(0, \Omega, \alpha)$$

A non-singular $\qquad AY \sim SN\left(0, A\Omega A^T, A^{-1}\alpha\right)$

$A^T \Omega A$ Correlation matrix

Quadratic forms

$$Y \sim SN(\xi, \Omega, \alpha)$$

B symmetric positive semidefinite $\Rightarrow \quad (Y - \xi)^T B (Y - \xi) \sim \chi_p^2$

$B\Omega B = B \quad rank(B) = p$

Even functions

$$Y \sim f(y) = 2g(y)G(w(y)) \qquad Z \sim g(z)$$

$$t(Y) \overset{d}{=} t(Z)$$

Skew Normal – Inferential issues

- Inflection point of the profile likelihood when $\alpha=0$ $\Rightarrow$ The information matrix is singular
- The shape of the likelihood function can be far from quadratic

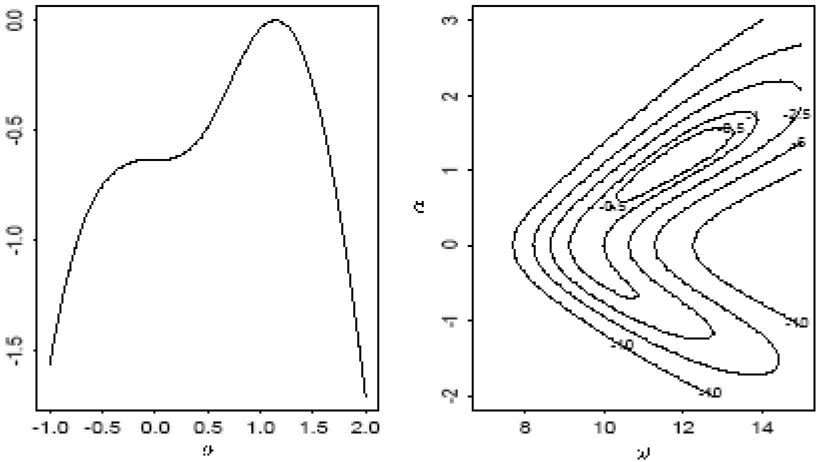


Figure 2: *Twice relative profile loglikelihood of α (left) and contour levels of the similar function of (ω, α) (right) for the Otis data, when the direct parametrization is used*

Azzalini & Capitanio, 1999

These problems can be remedied by reparameterization $CP=\left(\mu, Var\left(Y\right), \gamma_1\right)$

Azzalini (1985), Azzalini & Capitanio (1999), Arellano-Valle & Azzalini (2007)

- The estimate of the skewness parameter is often infinite

Liseo & Loperfido (2004), Sartori (2005), Greco (2008), Azzalini e Genton (2008)

Skew elliptical distribution

Purpose: to deal with both skewness and heavy tails

$$g(y;\xi,\Omega)=\frac{c_d}{|\Omega|^{1/2}}\tilde{g}\left\{(y-\xi)^T\Omega^{-1}(y-\xi)\right\}$$

Density function constant on ellipsoids

$$f(y)=2g(y;\xi,\Omega)G(w(y))$$

Stocastic representations

- Conditioning
- Transformation
- Maximum (d=2)

Azzalini & Capitanio (2003), Branco & Dey (2001)

Skew-t

$$Z \sim SN(0, \Omega, \alpha)$$
$$V \sim \chi_\nu^2 / \nu$$
$$Z \text{ and } V \text{ independent}$$

$$\Rightarrow \quad Y = \xi + V^{-1/2} Z \sim St(\xi, \Omega, \alpha, \nu)$$

$$f(y; \xi, \Omega, \alpha, \nu) = 2 t_d(y; \xi, \Omega, \nu) T_1\left(\alpha^T \omega^{-1}(y - \xi) \frac{\nu + 1}{Q_y + \nu}; \nu + d \right)$$

$$Q_Y = (Y - \xi)^T \Omega^{-1} (Y - \xi)$$

$$t_d(y; \xi, \omega, \nu) = \frac{\Gamma\left(\frac{\nu + d}{2}\right)\left(1 + \frac{Q_y}{\nu}\right)^{-\nu/2 - d/2}}{|\Omega|^{1/2} (\pi\nu)^{d/2} \Gamma\left(\frac{\nu}{2}\right)}$$

$$T(y; \nu) = \int_{-\infty}^{y} t(u; 0, 1, \nu) du$$

Main properties of the Skew-t

Linear transformation

$$Y \sim St\left(\xi, \Omega, \alpha, \nu\right) \quad \Rightarrow \quad AY + b \sim St\left(\xi', A\Omega A^T, \alpha', \nu\right)$$

Marginal distributions

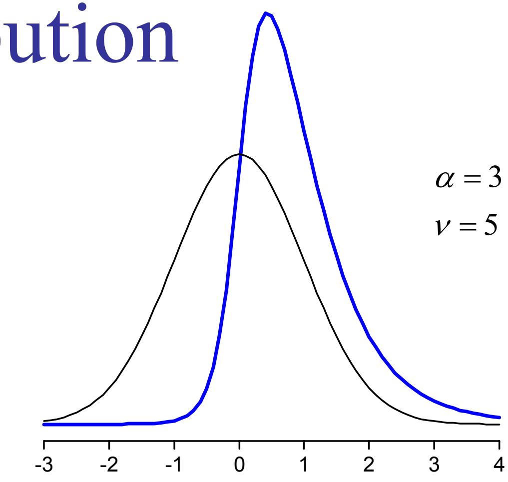
$$Y \sim St\left(\xi, \Omega, \alpha, \nu\right) \quad \Rightarrow \quad Y_j \sim St\left(\xi_j, \omega_{jj}, \alpha'_j, \nu\right)$$

Quadratic forms

$$Y \sim St\left(\xi, \Omega, \alpha, \nu\right) \quad \Rightarrow \quad \frac{\left(Y - \xi\right)^T \Omega^{-1} \left(Y - \xi\right)}{d} \sim F_{d,\nu}$$

Azzalini & Capitanio, 2003

Univariate Skew-*t* distribution



$$Y \sim St(\xi, \omega, \alpha, \nu)$$

$$f(y;\xi,\omega,\alpha,\nu) = 2t(y;\xi,\omega,\nu)T(\zeta;\nu+1)$$

$$\zeta = \alpha z \tau \qquad z = \frac{y-\xi}{\omega} \qquad \tau = \left(\frac{\nu+1}{\nu+z^2}\right)^{1/2}$$

$$t(y;\xi,\omega,\nu) = \frac{\Gamma\left(\frac{\nu+1}{2}\right)\left(1+\frac{z^2}{\nu}\right)^{-\nu/2-1/2}}{\omega(\pi\nu)^{1/2}\,\Gamma\left(\frac{\nu}{2}\right)} \qquad T(y;\nu) = \int_{-\infty}^{y} t(u;0,1,\nu)\,du$$

Moments of the Skew-t distribution

$$b_\nu = \left(\frac{\nu}{\pi}\right)^{1/2} \frac{\Gamma(\nu/2 + 1/2)}{\Gamma(\nu/2)}$$

$$\mu = \xi + \omega\delta b_\nu$$

$$\delta = \frac{\alpha}{\left(1+\alpha^2\right)^{1/2}}$$

$$\sigma^2 = \omega^2 \left\{\frac{\nu}{\nu-2} - \delta^2 b_\nu^2\right\}$$

$$\gamma_1 = \delta b_\nu \left\{\frac{\nu\left(3-\delta^2\right)}{\nu-3} - \frac{3\nu}{\nu-2} + 2\delta^2 b_\nu^2\right\}\left\{\frac{\nu}{\nu-2} - \delta^2 b_\nu^2\right\}^{-3/2}$$

$$\gamma_2 = \left\{\frac{3\nu^2}{(\nu-2)(\nu-4)} - \frac{4\delta^2 b_\nu^2 \nu\left(3-\delta^2\right)}{\nu-3} + \frac{6\delta^2 b_\nu^2 \nu}{\nu-2} - 3\delta^4 b_\nu^4\right\}\left\{\frac{\nu}{\nu-2} - \delta^2 b_\nu^2\right\}^{-2} - 3$$

Stew-t – Main features

Model able to deal with
- Skewness
- Heavy tails

Special cases

$\alpha = 0 \quad \Rightarrow \quad Y \sim t(\xi, \omega, \nu)$

$\nu \to \infty \quad \Rightarrow \quad Y \sim SN(\xi, \omega, \alpha)$

$\alpha = 0 \quad and \quad \nu \to \infty \quad \Rightarrow \quad Y \sim N(\xi, \omega^2)$

Topics of the talk

- Inferential issues under special cases
- Applications in robustness contexts

Score functions and information matrix

$$\vartheta = (\xi, \omega, \alpha, \nu)$$

$$S_\xi(y) = \frac{\partial \ln f(y;\vartheta)}{\partial \xi} = \frac{z\tau^2}{\omega} - \frac{\alpha\tau\nu}{\omega(\nu+z^2)} w \qquad S_\omega(y) = \frac{\partial \ln f(y;\vartheta)}{\partial \omega} = -\frac{1}{\omega} + \frac{z^2\tau^2}{\omega} - \frac{\alpha\tau z\nu}{\omega(\nu+z^2)} w$$

$$S_\alpha(y) = \frac{\partial \ln f(y;\vartheta)}{\partial \alpha} = z\tau w$$

$$S_\nu(y) = \frac{\partial \ln f(y;\vartheta)}{\partial \nu} = \frac{1}{2}\left\{\Psi\left(\frac{\nu}{2}+1\right) - \Psi\left(\frac{\nu}{2}\right) - \frac{2\nu+1}{\nu(\nu+1)} - \ln\left(1+\frac{z^2}{\nu}\right) + \frac{z^2\tau^2}{\nu} + \frac{\alpha z(1-\tau^2)}{\tau(\nu+z^2)} w + \frac{\gamma}{T(\zeta,\nu+1)}\right\}$$

$$z = \frac{y-\xi}{\omega} \quad \zeta = \alpha z\tau \quad \tau = \left(\frac{\nu+1}{\nu+z^2}\right)^{1/2} \quad w = \frac{t(\varsigma,\nu+1)}{T(\varsigma,\nu+1)} \quad \Psi(u) = \frac{\partial \ln\Gamma(u)}{\partial u} \quad \gamma = \int_{-\infty}^{\zeta}\left\{\frac{(\nu+2)u^2}{(\nu+1)(\nu+1+u^2)} - \ln\left(1+\frac{u^2}{\nu+1}\right)\right\} t_{\nu+1}(u)\,du$$

Maximum Likelihood Estimators

$$Y_1, Y_2, \ldots, Y_n \qquad \hat{\vartheta} = (\hat{\xi}, \hat{\omega}, \hat{\alpha}, \hat{\nu}): \qquad \sum_{i=1}^{n} S_{\vartheta_j}(y_i) = 0 \qquad j = 1, \ldots, 4$$

Information matrix $$I = E\left\{S_\vartheta(Y) S_\vartheta(Y)^T\right\} = -E\left\{\frac{\partial^2 \ln f(y;\vartheta)}{\partial\vartheta\,\partial\vartheta^T}\right\}$$

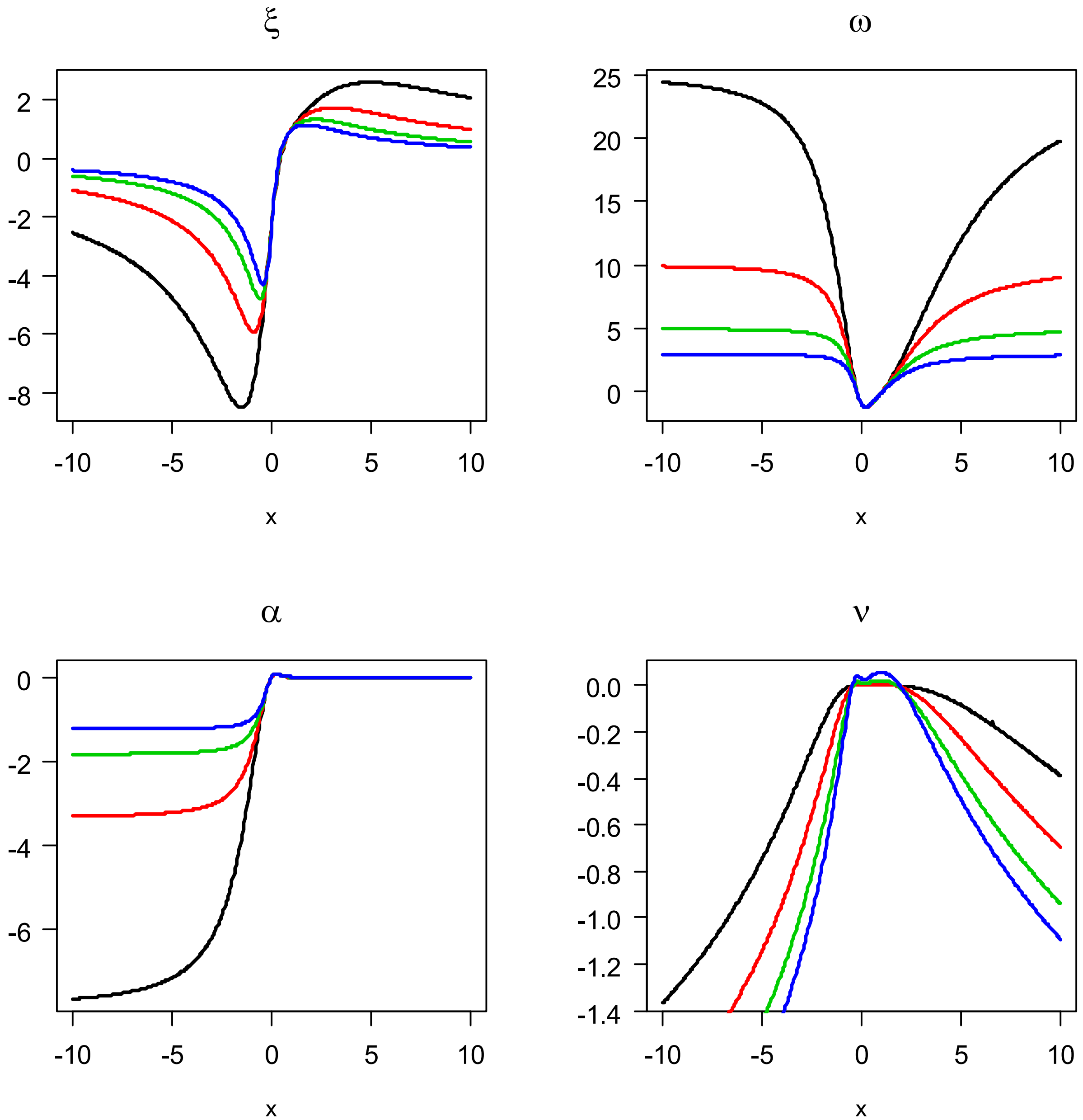
Score functions

$\xi = 0$

$\omega = 1$

$\alpha = 3$

ν=3 (blue)
ν =5 (green)
ν =10 (red)
ν =25 (black)



Issues related to likelihood inference

- Nice likelihood function behaviour { No stationary point; Quadratic shape }

(Azzalini and Capitanio 2003,Azzalini and Genton, 2008)

- The information matrix can be singular
- Normality or skew normality ⇨ parameter on the boundary
- The estimate of α can be infinite

Sartori (2005), Azzalini e Genton (2008), Greco (2008)

Score functions and information matrix under the *t* distribution

$$S^t_\xi(y)=\frac{z\tau^2}{\omega} \qquad S^t_\omega(y)=-\frac{1}{\omega}+\frac{z^2\tau^2}{\omega} \qquad S^t_\alpha(y)=2z\tau t_{\nu+1}(0)$$

$$S^t_\nu(y)=\frac{1}{2}\left\{\Psi\left(\frac{\nu}{2}+1\right)-\Psi\left(\frac{\nu}{2}\right)-\ln\left(1+\frac{z^2}{\nu}\right)+-\frac{z^2-1}{\nu+z^2}\right\}$$

$I^t_{\xi\omega}=I^t_{\xi\nu}=I^t_{\omega\alpha}=I^t_{\alpha\nu}=0$ ⇒ $(\hat{\xi},\hat{\alpha})$ $(\hat{\omega},\hat{\nu})$ *are asymptotically uncorrelated*

- ❑ No singularity of the information matrix when α=0
- ❑ The likelihood ratio statistic for the null hypothesis α=0 has a regular asymptotic behaviour

Score functions when ν diverges

Skew Normal

$$S_\xi^{SN}(y) = \frac{z}{\omega} - \frac{\alpha}{\omega}\frac{\phi(\alpha z)}{\Phi(\alpha z)}$$

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

$$S_\alpha^{SN}(y) = z\frac{\phi(\alpha z)}{\Phi(\alpha z)}$$

$$S_\nu^{SN}(y) = 0$$

Normal

$$S_\xi^{N}(y) = \frac{z}{\omega}$$

$$S_\omega^{N}(y) = -\frac{1}{\omega} + \frac{z^2}{\omega}$$

$$S_\alpha^{N}(y) = z\left(\frac{2}{\pi}\right)^{1/2}$$

$$S_\nu^{N}(y) = 0$$

$$S_\nu(y) = O(\nu^{-2})$$

rank(I^N)=2

The information matrix is singular!

Inferential problems when ν diverges

- Singularity of the information matrix
- Parameter ν on the boundary

⇓

- Efficiency of the estimators
- Asymptotic distribution of the estimators
- Likelihood ratio test

Root mean square errors

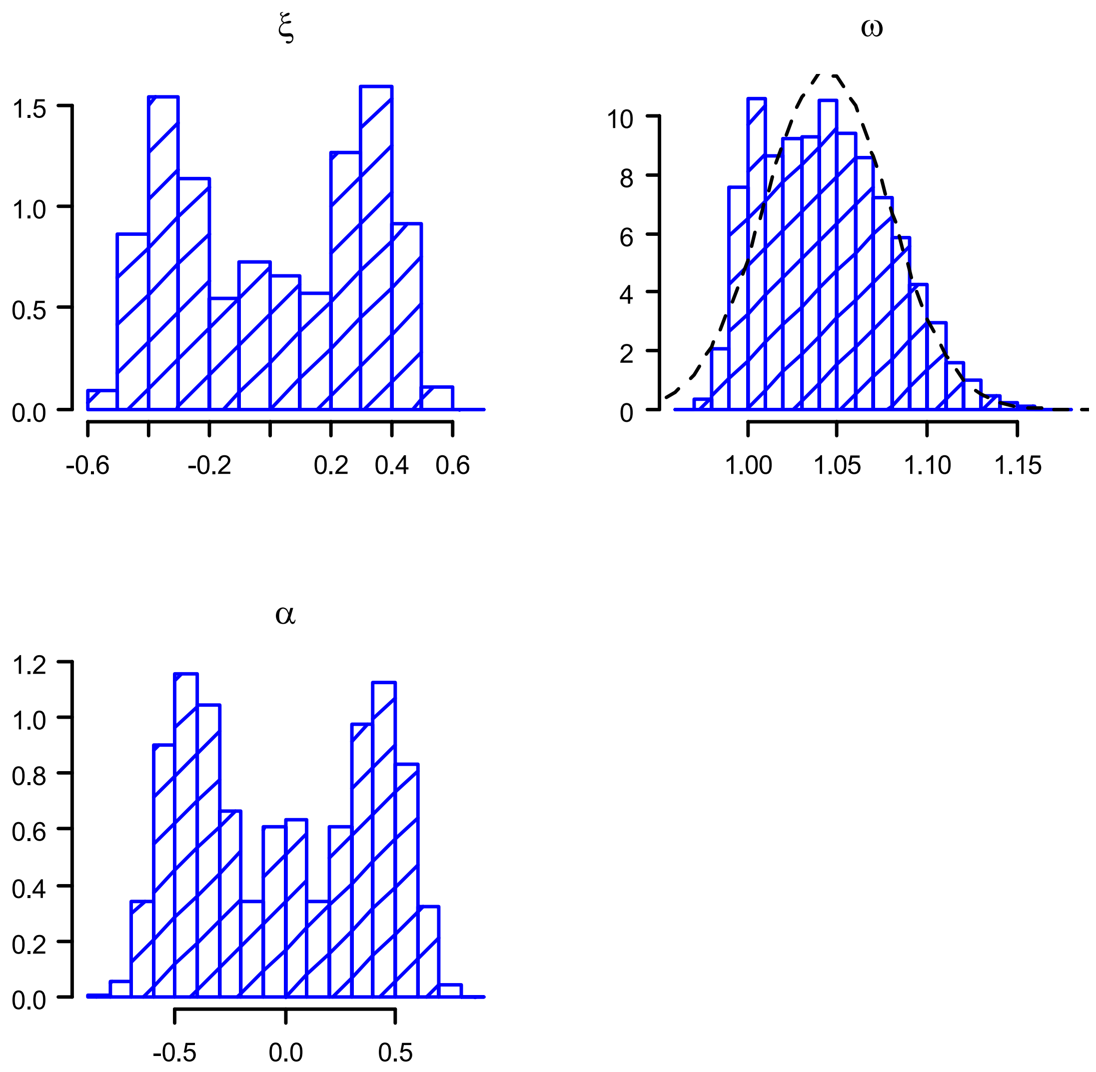
Root mean square errors of the estimators of the direct parameters ξ, ω, α, e κ=1/ν when the underlying distribution is $N\left(\xi,\omega^2\right)$

n	ξ	ω	α	κ
50	0.8362	0.3467	350.7858	0.0667
100	0.7159	0.2651	1.5205	0.0459
200	0.6210	0.2054	0.9790	0.0328
500	0.5240	0.1489	0.7582	0.0211
1000	0.4584	0.1170	0.6394	0.0157
5000	0.3520	0.0705	0.4672	0.0073
10000	0.3115	0.0558	0.4083	0.0052

Simulations=10,000

Simulated distribution of the estimators under the Normal model

n = 10,000
Simulations = 10,000



Distribution of the likelihood ratio statistic to test normality or skew normality

$H_0 : Y \sim N\left(\xi, \omega^2\right)$

$H_0 : Y \sim SN\left(\xi, \omega, \alpha\right)$

The likelihood statistic does not converge in distribution to a chi-square !!!

Reparameterization of the degrees of freedom

The singularity at the *SN* Model can be removed by a reparameterization

$$\kappa = \frac{1}{\nu} \quad \Longrightarrow \quad \vartheta_\kappa = (\xi, \omega, \alpha, \kappa)$$

$$S_\kappa(y) = \frac{1}{2\kappa^2}\Psi\left(\frac{1}{2\kappa}+1\right) - \frac{1}{2\kappa^2}\Psi\left(\frac{1}{2\kappa}\right) - \frac{\kappa(\kappa+2)}{(\kappa+1)} - \ln\left(1+\kappa z^2\right) - \frac{(\kappa+1)z^2}{2\kappa\left(\kappa z^2+1\right)}$$
$$- \frac{\alpha z\left(z^2-1\right)}{2(1+\kappa)^{1/2}\left(1+\kappa z^2\right)^{3/2}}\frac{t_{1/\nu+1}(\zeta)}{T_{1/\nu+1}(\zeta)} - \frac{\gamma}{2\kappa^2 T_{1/\nu+1}(\zeta)}$$

$$S_\kappa^{SN}(y) = \frac{1}{4}\left\{z^4 - 2z^2 - 1 - \alpha z\left(2z^2 + \alpha^2 z^2 - 1\right)\frac{\phi(\alpha z)}{\Phi(\alpha z)}\right\}$$

Asymptotic properties of the skew-t MLE under the SN Model

$\kappa = 0$

$\hat{\vartheta}_\kappa = \left(\hat{\xi}, \hat{\omega}, \hat{\alpha}, \hat{\kappa}\right)$ is consistent for $\vartheta_\kappa = \left(\xi, \omega, \alpha, \kappa\right)$

$$n^{1/2}\begin{bmatrix} \hat{\xi}-\xi \\ \hat{\omega}-\omega \\ \hat{\alpha}-\alpha \\ \hat{\kappa} \end{bmatrix} \overset{d}{=} \begin{bmatrix} Z_1 \\ Z_2 \\ Z_3 \\ Z_4 \end{bmatrix} I\left(Z_4>0\right) + \begin{bmatrix} Z_1-\left(I_{\xi\kappa}/I_{\kappa\kappa}\right)Z_4 \\ Z_2-\left(I_{\omega\kappa}/I_{\kappa\kappa}\right)Z_4 \\ Z_3-\left(I_{\alpha\kappa}/I_{\kappa\kappa}\right)Z_4 \\ 0 \end{bmatrix} I\left(Z_4<0\right)$$

$$\left(Z_1, Z_2, Z_3, Z_4\right) \sim N\left(0, \left(I^{SN}\right)^{-1}\right)$$

Asymptotic distribution of the likelihood statistic under the *SN* distribution

$$H_0 : Y \sim SN(\xi, \omega, \alpha)$$

$$-2\ln\lambda_{SN} = 2\left(\hat{\ell}_{St} - \hat{\ell}_{SN}\right)$$

$$-2\ln\lambda \xrightarrow{d} Z^2 I(Z > 0) \qquad Z \sim N(0,1)$$

Simulated level of the likelihood ratio test for skew-normality with nominal level 0.05

n	$\alpha = 0.5$	$\alpha = 1.5$	$\alpha = 3$	$\alpha = 5$	$\alpha = 10$
100	0.0179	0.0265	0.0765	0.1676	0.1569
200	0.0228	0.0290	0.0306	0.0473	0.0652
500	0.0273	0.0329	0.0378	0.0369	0.0331
1000	0.0320	0.0338	0.0372	0.0416	0.0393

Replications 10,000

Information matrix at the normal model with $\kappa=1/\nu$

$$S_\xi^N(y) = \frac{z}{\omega}$$

$$S_\omega^N(y) = -\frac{1}{\omega} + \frac{z^2}{\omega}$$

$$S_\alpha^N(y) = z\left(\frac{2}{\pi}\right)^{1/2}$$

$$S_\kappa^N(y) = \frac{1}{4}\left(z^4 - 2z^2 - 1\right)$$

$$I_\kappa^N = \begin{bmatrix} \frac{1}{\omega^2} & 0 & \frac{1}{\omega}\left(\frac{2}{\pi}\right)^{1/2} & 0 \\ 0 & \frac{2}{\omega^2} & 0 & \frac{2}{\omega} \\ \frac{1}{\omega}\left(\frac{2}{\pi}\right)^{1/2} & 0 & \frac{2}{\pi} & 0 \\ 0 & \frac{2}{\omega} & 0 & \frac{7}{2} \end{bmatrix}$$

- The information matrix is still singular due to the link between S_ξ and S_α
- The parameter κ is on the boundary

Pseudo centred parameterization

$$\vartheta = (\xi, \omega, \alpha, \nu) \Rightarrow \vartheta_\kappa = (\xi, \omega, \alpha, \kappa) \Longrightarrow CP = (\mu, \sigma^2, \gamma_1, \gamma_2) \quad \nu > 4$$

$$S_{CP} = \left(D^T\right)^{-1} S_{DP^\kappa} \qquad D = \left[\frac{\partial CP_i}{\partial \vartheta_{\kappa,j}}\right]$$

Azzalini 2007

Normal model

$$S_\mu(y) = \frac{z}{\sigma}$$

$$S_{\sigma^2}(y) = \frac{1}{2}\frac{z^2-1}{\sigma^2}$$

$$S_{\gamma_1}(y) = \frac{1}{6}\left(z^3 - z\right)$$

$$S_{\gamma_2}(y) = \frac{1}{24}\left(z^4 - 6z^2 + 3\right)$$

$$I_{CP}^N = E\left\{S_{CP}^N(Y)\, S_{CP}^N(Y)'\right\}$$

$$\Downarrow$$

$$\begin{bmatrix} 1/\omega^2 & 0 & 0 & 0 \\ 0 & 1/2\omega^2 & 0 & 0 \\ 0 & 0 & 1/6 & 0 \\ 0 & 0 & 0 & 1/24 \end{bmatrix}$$

Root mean square errors with centered parameterization

Root mean square errors of the estimators of the direct parameters ξ, ω, α, e κ=1/ν and the centred parameters μ, σ, γ_1 e $\gamma_2$when the underlying distribution is $N\left(\xi,\omega^2\right)$

n	ξ	ω	α	κ	μ	σ	γ_1	γ_2
50	0.8362	0.3467	350.7858	0.0667	0.1463	0.1171 (1)	5.7735 (43)	12.3627 (151)
100	0.7159	0.2651	1.5205	0.0459	0.0994	0.0712	0.7424	2.8703 (9)
200	0.6210	0.2054	0.9790	0.0328	0.0714	0.0505	0.1915	0.4495 (1)
500	0.5240	0.1489	0.7582	0.0211	0.0436	0.0318	0.1131	0.1874
1000	0.4584	0.1170	0.6394	0.0157	0.0311	0.0222	0.0773	0.1253
5000	0.3520	0.0705	0.4672	0.0073	0.0142	0.0101	0.0338	0.0507
10000	0.3115	0.0558	0.4083	0.0052	0.0099	0.0070	0.0234	0.0353

(In parenthesis there is the number of omitted cases, in 10,000 simulations, due to estimates $\hat{\nu}$ lower than the value required for the existence of the corresponding moments)

Skew t MLEs of the centred parameters at the Normal model

$\kappa = 0$ ⟶ γ_2 is on the boundary

Consistency

$\left(\hat{\mu}, \hat{\sigma}^2, \hat{\gamma}_1, \hat{\gamma}_2\right)$ is consistent for $\left(\mu, \sigma^2, \gamma_1, \gamma_2\right)$

Asymptotic distribution of the estimators

$$n^{1/2}\begin{bmatrix} \hat{\mu} - \mu \\ \hat{\sigma}^2 - \sigma^2 \\ \hat{\gamma}_1 \\ \hat{\gamma}_2 \end{bmatrix} \overset{d}{=} \begin{bmatrix} Z_1 \\ Z_2 \\ Z_3 \\ Z_4 I\left(Z_4 > 0\right) \end{bmatrix}$$

$$\left(Z_1, Z_2, Z_3, Z_4\right) \sim N\left(0, I^{-1}\right) \quad I^{-1} = diag\left(\omega^2, 2\omega^2, 6, 24\right)$$

Likelihood ratio normality test

$$H_0 : Y \sim N\left(\xi, \omega^2\right)$$

$$-2\ln\lambda_N = 2\left(\hat{\ell}_{St} - \hat{\ell}_N\right)$$

$$-2\ln\lambda \xrightarrow{d} Z_1^2 I\left(Z_1 > 0\right) + Z_2^2 \qquad Z_i \sim N\left(0,1\right) \quad for \quad i = 1, 2$$

Simulated level of the likelihood ratio test for normality with Self and Liang percentiles

$$-2\ln\lambda \xrightarrow{d} Z_1^2 I\left(Z_1 > 0\right) + Z_2^2 \text{ with } Z_i \sim N\left(0,1\right) \quad per \quad i = 1, 2$$

Nominal	n					
level	100	200	500	1000	5000	10,000
0.10	0.0801	0.0764	0.0796	0.0826	0.0881	0.0830
0.05	0.0406	0.0371	0.0371	0.0405	0.0448	0.0436
0.01	0.0095	0.0071	0.0075	0.0084	0.0083	0.0090

Replications 10,000

Skew *t* and robusteness

- The model is skew-*t*.

  Are the MLE robust?

- The skew-t is assumed to deal with deviations from a simpler model.

Influence function of the skew-t MLE

$$IF(x,F) = I(F)^{-1} S_{\vartheta}(y,F)$$

$$\gamma_s = \sup_y \left\{ S_{\vartheta}(y,F)^T I(F) S_{\vartheta}(y,F) \right\}^{1/2}$$

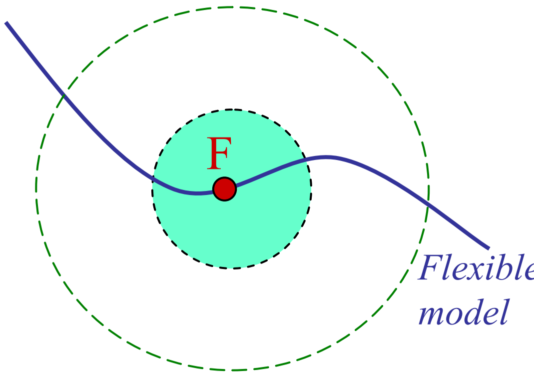
- The degrees of freedom are estimated

  The influence functions are unbounded

- The degrees of fredom are fixed
  - The influence functions are bounded
  - The gross error sensitivity is limited if α is finite

Robustness versus flexible models

- Robust methods work in a neighbourhood
- Flexible models are enlarged models



ADAPTIVE ROBUST ESTIMATION

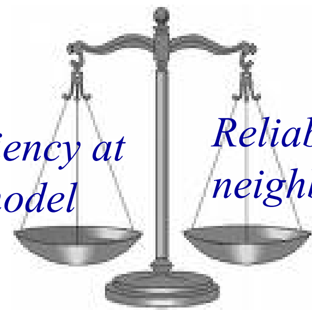
Only some parameters are of interest

$$Y = X\beta + \varepsilon$$

The skew-t score functions are *modified* estimating equations designed to deal with deviations from normality

- parameters have a clear meaning
- inference can be based on likelihood function
- normality can be tested through hypotheses on the parameters
- prediction intervals

Robustness properties



Efficiency at the model

Reliability in a neighbourhood of the model

What is the loss of efficiency if the simpler model is the correct?

How do the skew-t MLEs compare with robust estimators?

Efficiency at the normal model

$$X \sim N\left(\xi, \omega^2\right)$$

Efficiency in the estimation of the location parameter under the normal model

n	Huber	$St-DP$	$St-CP$
50	0.9522	0.0289	0.9425
100	0.9489	0.0193	0.9983
200	0.9591	0.0132	0.9997
500	0.9526	0.0069	1.0000
1000	0.9499	0.0046	1.0001
5000	0.9475	0.0016	1.0000
10000	0.9547	0.0010	1.0000

Efficiency in the estimation of the scale parameter (standard deviation) under the normal model

n	Huber	$St-DP$	$St-CP$
50	0.1642	0.0835	0.7317
100	0.1604	0.0717	0.9928
200	0.1648	0.0604	0.9985
500	0.1677	0.0455	0.9999
1000	0.1619	0.0359	1.0000
5000	0.1680	0.0206	1.0001
10000	0.1649	0.0156	1.0000

Huber Proposal 2 - 95% efficiency at the normal model (k=1.35)

Efficiency of the skew-t MLEs when the undelying model is Skew Normal

Simulations results

n = 1000

Montecarlo replications = 10,000

DIRECT PARAMETER

n	$\alpha = 0.5$	$\alpha = 1.5$	$\alpha = 3$	$\alpha = 5$	$\alpha = 10$
ξ	1.1310	0.7569	0.8608	0.8996	0.9437
ω	1.0931	0.6141	0.6215	0.5913	0.6395
α	1.1417	0.8469	0.9630	1.0019	1.0260

CENTRED PARAMETERS

n	$\alpha = 0.5$	$\alpha = 1.5$	$\alpha = 3$	$\alpha = 5$	$\alpha = 10$
μ	1.0000	1.0004	0.9988	0.9868	0.9846
σ^2	0.9999	0.9975	0.9824	0.9722	0.9412
γ_1	0.6774	0.8846	0.5387	0.1953	0.0337
γ_2	0.0648	0.1452	0.0555	0.0140	0.0022

Efficiency of the skew t MLE when the underlying distribution is t

- No loss of efficiency in the estimation of ω and ν

Efficiency of the skew t MLE of the location parameter ξ when the underlying distribution is t

ν	2	3	5	7	10	15	20	30	60
DP	0.1326	0.0778	0.0363	0.0210	0.0114	0.0055	0.0033	0.0015	0.0004
CP	0.2980	0.5743	0.8110	0.8958	0.9462	0.9751	0.9858	0.9936	0.9984

Robustness in regression models

$$Y = X\beta + \varepsilon$$

Parameter of interest:
Regression coefficients

- Skew t MLE (with and without fixed degrees of freedom)

- Least squares

- t MLE

- M-estimators (Huber's ψ, k=1.345, scale=MAD)

$$\psi(u) = x \min\left\{1, \frac{k}{|x|}\right\}$$

- MM – estimators (BDP = 0.5, Eff=95%)

  Tukey's biweigth ψ
  Step 1: S-estimates
  Step 2: M-estimates with fixed scale

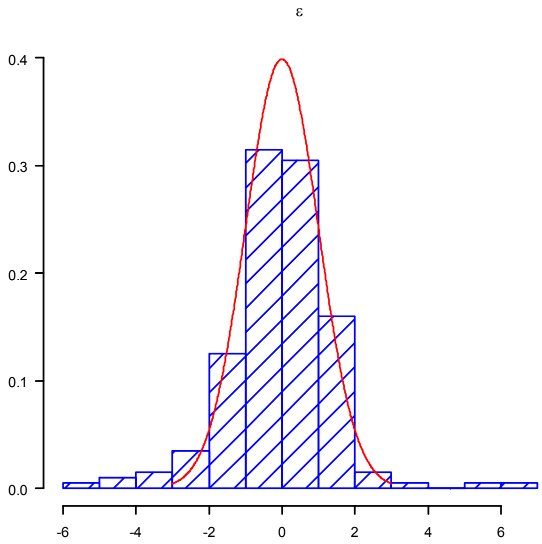
$$\psi(u) = u\left\{1 - \left(\frac{u}{k}\right)^2\right\}^2 I(|u| \le k)$$

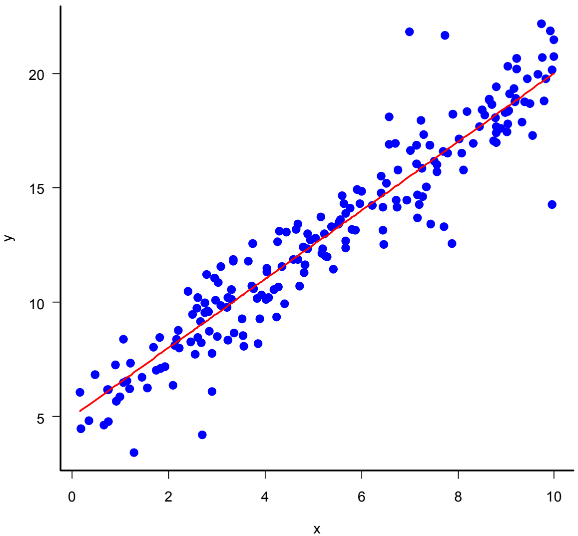
Monte Carlo replications=1000, n=200

Estimate of the intercept adjusted so that the error has zero mean

Model 1

$$Y = 5 + 1.5x + \varepsilon \qquad \varepsilon \sim 0.9N(0,1) + 0.1N(0,9)$$





Root mean square errors of the estimators of the regression coefficients – Model 1

	St	*t*	**LS**	M	MM	*St* (ν=3)	*St* (ν=5)	*St* (ν=7)	*St* (ν=10)	*St* (ν=15)
β_1	0.1806	0.1786	**0.2038**	0.1771	0.1760	0.1837	0.1791	0.1793	0.1810	0.1841
β_2	0.0294	0.0293	**0.0337**	0.0290	0.0289	0.0297	0.0291	0.0291	0.0293	0.0298

Not substantially different from robust estimators

Models from 2 to 4

Model 2

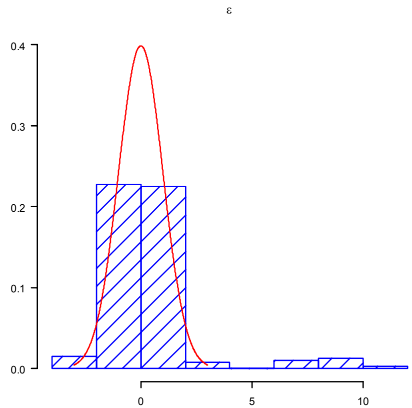
$$Y = 5 + 1.5x + \varepsilon$$

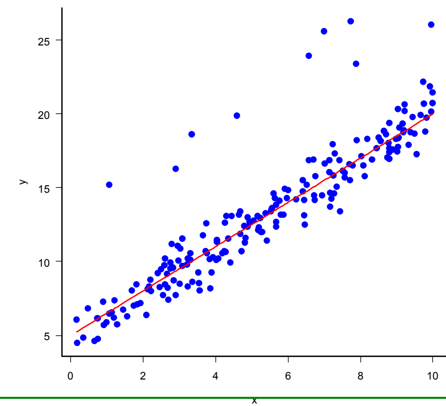
$$\varepsilon \sim 0.9N(0,1) + 0.1N(0,25)$$

Model 3

$$Y = 5 + 1.5x + \varepsilon$$

$$\varepsilon \sim 0.95N(0,1) + 0.05N(8,1)$$

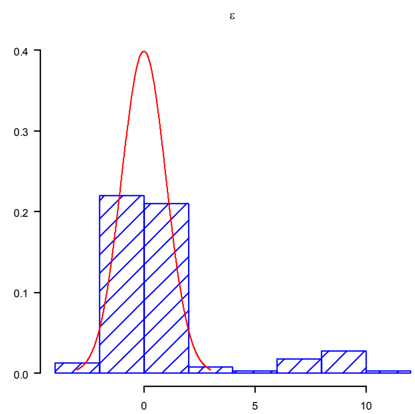


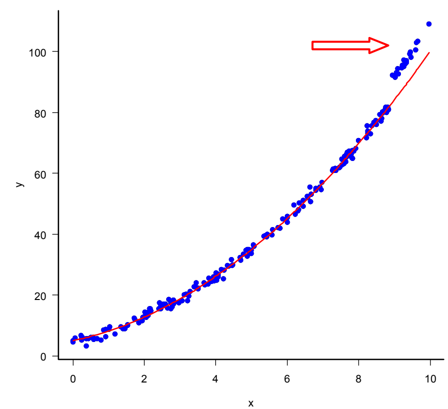


Model 4

$$Y = 5 + 2.5x + 0.7x^2 + \varepsilon$$

$$\varepsilon_{last\ 20} \sim N(8,1)$$





Root mean square errors of the estimators of the regression coefficients

Model 2

	St	*t*	**LS**	M	MM	*St* (ν=3)	*St* (ν=5)	*St* (ν=7)	*St* (ν=10)	*St* (ν=15)
β_1	0.2002	0.1841	**0.2771**	0.1827	0.1765	0.1921	0.1889	0.1924	0.1998	0.2117
β_2	0.0303	0.0302	**0.0460**	0.0299	0.0291	0.0297	0.0295	0.0299	0.0308	0.0323

Model 3

	St	*t*	**LS**	M	MM	*St* (ν=3)	*St* (ν=5)	*St* (ν=7)	*St* (ν=10)	*St* (ν=15)
β_1	0.3913	0.1826	**0.5079**	0.1992	0.1645	0.3388	0.3337	0.3614	0.4066	0.4652
β_2	0.0297	0.0300	**0.0517**	0.0291	0.0271	0.0292	0.0292	0.0299	0.0309	0.0321

Model 4

	St	*t*	**LS**	M	MM	*St* (ν=3)	*St* (ν=5)	*St* (ν=7)	*St* (ν=10)	*St* (ν=15)
β_1	1.0282	0.9714	**2.0716**	1.4624	0.2859	0.9088	1.0373	1.1616	1.2858	1.3961
β_2	0.2505	0.5931	**1.4048**	0.9877	0.1299	0.2689	0.3554	0.4186	0.4823	0.5422
β_3	0.0291	0.0703	**0.1727**	0.1220	0.0129	0.0312	0.0419	0.0496	0.0574	0.0647

The skew-t MLE estimators are

- definitely better than L.S.
- not as efficient as MM estimators

Should be the degrees of freedom be held fixed?