

MULTIVARIATE SKEW- T DISTRIBUTIONS IN ECONOMETRICS AND
ENVIRONMETRICS

A Dissertation

by

YULIA V. MARCHENKO

Submitted to the Office of Graduate Studies of
Texas A&M University
in partial fulfillment of the requirements for the degree of

DOCTOR OF PHILOSOPHY

December 2010

Major Subject: Statistics

MULTIVARIATE SKEW- T DISTRIBUTIONS IN ECONOMETRICS AND
ENVIRONMETRICS

A Dissertation

by

YULIA V. MARCHENKO

Submitted to the Office of Graduate Studies of
Texas A&M University
in partial fulfillment of the requirements for the degree of

DOCTOR OF PHILOSOPHY

Approved by:

Chair of Committee,	Marc G. Genton
Committee Members,	Raymond J. Carroll
	Mohsen Pourahmadi
	Kenneth P. Bowman
Head of Department,	Simon J. Sheather

December 2010

Major Subject: Statistics

ABSTRACT

Multivariate Skew- t Distributions in Econometrics and
Environmetrics. (December 2010)

Yulia V. Marchenko, Diploma, Belarussian State University, Minsk, Belarus;

M.S., Texas A&M University

Chair of Advisory Committee: Dr. Marc G. Genton

This dissertation is composed of three articles describing novel approaches for analysis and modeling using multivariate skew-normal and skew- t distributions in econometrics and environmetrics.

In the first article we introduce the Heckman selection- t model. Sample selection arises often as a result of the partial observability of the outcome of interest in a study. In the presence of sample selection, the observed data do not represent a random sample from the population, even after controlling for explanatory variables. Heckman introduced a sample-selection model to analyze such data and proposed a full maximum likelihood estimation method under the assumption of normality. The method was criticized in the literature because of its sensitivity to the normality assumption. In practice, data, such as income or expenditure data, often violate the normality assumption because of heavier tails. We first establish a new link between sample-selection models and recently studied families of extended skew-elliptical distributions. This then allows us to introduce a selection- t model, which models the error distribution using a Student's t distribution. We study its properties and investigate the finite-sample performance of the maximum likelihood estimators for this model. We compare the performance of the selection- t model to the Heckman selection model and apply it to analyze ambulatory expenditures.

In the second article we introduce a family of multivariate log-skew-elliptical distributions, extending the list of multivariate distributions with positive support. We investigate their probabilistic properties such as stochastic representations, marginal and conditional distributions, and existence of moments, as well as inferential properties. We demonstrate, for example, that as for the log- t distribution, the positive moments of the log-skew- t distribution do not exist. Our emphasis is on two special cases, the log-skew-normal and log-skew- t distributions, which we use to analyze U.S. precipitation data.

Many commonly used statistical methods assume that data are normally distributed. This assumption is often violated in practice which prompted the development of more flexible distributions. In the third article we describe two such multivariate distributions, the skew-normal and the skew- t , and present commands for fitting univariate and multivariate skew-normal and skew- t regressions in the statistical software package Stata.

To my parents, my husband, and all my extended family

In the memory of my mother

ACKNOWLEDGMENTS

I would like to thank all members of my committee: Dr. Raymond J. Carroll, Dr. Mohsen Pourahmadi, and Dr. Kenneth P. Bowman, for their helpful comments and suggestions. I would especially like to thank my committee chair, Dr. Marc G. Genton, for his guidance and patience during the whole dissertation process. His expertise and advice were invaluable.

I would also like to thank Dr. Reinaldo B. Arellano-Valle, Dr. Kenneth P. Bowman, and the referee for their helpful comments for the article “Multivariate Log-Skew-Elliptical Distributions with Applications to Precipitation Data”.

I would like to thank my family and friends for their support during the completion of my dissertation. Special thanks to my husband for his enormous patience, support, and encouragement, to my stepmother for being there for me during difficult and good times, to my grandparents for the fondest memories I have from visiting them, and to my parents-in-law for their love, patience, and understanding. I am in debt to my mother, who is greatly missed, for her unbounded love, her courage, and her strength, and to my father, who has always been an inspiration to me and taught me integrity, honesty, and persistence. My parents’ example led me to try to excel in all that I do.

Finally I would like to thank StataCorp for its flexibility as I was working toward my Ph.D. and my Stata family for putting up with me ignoring them socially to work on my dissertation.

TABLE OF CONTENTS

CHAPTER		Page
I	INTRODUCTION	1
	A. Overview of Skewed Distributions	1
	B. Motivation	6
II	A HECKMAN SELECTION- T MODEL	9
	A. Heckman Selection Models	11
	1. Classical Heckman sample-selection model	12
	2. Link to extended skew-elliptical distributions	15
	3. Selection- t model	18
	B. Monte Carlo Simulations	22
	1. Finite-sample properties of the MLEs	22
	2. Test of selection bias when the error distribution is bivariate t	26
	C. Numerical Application to Ambulatory Expenditures	29
	D. Conclusion	30
III	MULTIVARIATE LOG-SKEW-ELLIPTICAL DISTRIBUTIONS WITH APPLICATIONS TO PRECIPITATION DATA	32
	A. Multivariate Log-Skew-Elliptical Distributions	35
	1. Definitions	35
	2. Stochastic representations	39
	3. Marginal and conditional distributions	42
	4. Moments	44
	5. Inference with log-skew-elliptical distributions	45
	B. Precipitation Data Analysis	47
	1. U.S. national scale	48
	2. U.S. regional scale	57
	C. Discussion	64
IV	A SUITE OF COMMANDS FOR FITTING THE SKEW- NORMAL AND SKEW- T MODELS	67
	A. Motivating Example	69
	B. The Skew-Normal and Skew- t Regression Models	74

CHAPTER	Page
1. Definition and some properties	74
2. Regression models	76
3. Centered parametrization	78
C. A Suite of Commands for Fitting Skewed Regressions . . .	80
1. Syntax	80
2. Description	81
3. Options	82
D. Analyses of Australian Institute of Sport Data	84
1. Univariate analysis	84
2. Multivariate analysis	92
E. Analysis of Automobile Prices	96
F. Log-Skew-Normal and Log-Skew- t Distributions	99
G. Conclusion	102
V SUMMARY	103
REFERENCES	105
APPENDIX A	111
APPENDIX B	121
VITA	125

LIST OF TABLES

TABLE		Page
1	Simulation results in the presence of the exclusion restriction.	24
2	Simulation results in the absence of the exclusion restriction.	25
3	Empirical significance levels (as %) of the tests of selection bias for the nominal significance level $\alpha = 0.01$	27
4	Powers (as %) of the tests of selection bias for $\nu = 100$ for the nominal significance level $\alpha = 0.01$	28
5	Powers (as %) of the SLt test of $\rho = 0$ for $\nu = 3$ and $\nu = 5$ for the nominal significance level $\alpha = 0.01$	28
6	Monthly (January-June) precipitation minimum, mean, median, and maximum values for nine U.S. climatic regions.	49
7	Monthly (July-December) precipitation minimum, mean, median, and maximum values for nine U.S. climatic regions.	50
8	Description of the Australian Institute of Sport data.	69
9	Regression of lean body mass on weight and height.	70
10	Skew-normal regression of lean body mass on weight and height.	71
11	Skew- t regression of lean body mass on weight and height.	72
12	Skew-normal regression analysis of the AIS data.	85
13	Skew-normal regression of lean body mass on weight and height in the DP metric.	86
14	Skew- t regression analysis of the AIS data.	88
15	Likelihood-ratio test of the skew- t regression versus the skew- normal regression.	89

TABLE	Page
16	AIC and BIC for the skew-normal and skew- t models. 90
17	Combined regression results. 91
18	Bivariate skew-normal fit to lean body mass and body mass index. . . 93
19	Joint test of bivariate normality for lean body mass and body mass index. 93
20	Bivariate skew-normal fit to lean body mass and body mass index in the DP metric. 94
21	Bivariate skew- t fit to lean body mass and body mass index. 95
22	Gender-specific bivariate skew- t fit to lean body mass and body mass index. 96
23	Summary statistics of automobile prices. 97
24	Skew- t model for automobile prices. 98
25	Estimated moments of the skew- t fit to automobile prices. 98
26	Log-skew-normal fit to U.S. national precipitation data in January. . . 100

LIST OF FIGURES

FIGURE	Page
1	Plot of $\lambda_\nu(\cdot)$ for different values of ν with $\lambda_\infty(\cdot) = \lambda(\cdot)$, corresponding to the normal case. 20
2	Plot of $\lambda'_\nu(\cdot)$ for different values of ν with $\lambda'_\infty(\cdot) = \lambda'(\cdot)$, corresponding to the normal case. 22
3	Left panel: skew-normal (solid curves) and log-skew-normal (dashed curves) densities with $\alpha = 0$ (thick curves) and $\alpha = 0.5, 2, 20$. Right panel: skew- t (solid curves) and log-skew- t (dashed curves) densities with $\alpha = 0$ (thick curves) and $\alpha = 0.5, 2, 20$ for fixed $\nu = 3$. 37
4	Log-skew-normal densities $\text{LSN}(\xi, \omega^2, 2)$ for varying ξ (left panel) and varying ω (right panel). 38
5	Contour plots of the standard bivariate log-skew-normal density plotted at levels 0.03, 0.05, 0.1, 0.15, 0.2, 0.3, 0.4, 0.8 for varying values of α 40
6	U.S. climatic regions from NCDC. 48
7	PP-plots for the univariate normal model (top left panel) and the univariate skew-normal model (top right panel) fitted to the log-precipitation for January. 52
8	Estimated mean of the log-skew-normal distribution fitted to precipitation for each month. 53
9	Estimated skewness indexes (log scale) of the log-skew-normal distribution fitted to precipitation for each month. 54
10	Estimated degrees of freedom of the log-skew- t distribution fitted to the national precipitation data for each month. 56
11	Bivariate scatter plots overlaid with contour plots for the multivariate skew- t model fitted to log-precipitation for November. 58

FIGURE	Page
12	Estimated degrees of freedom of the multivariate log-skew- t distribution fitted to the regional precipitation data for each month. 60
13	Estimated marginal skewness indexes of the multivariate skew- t distributions fitted to log-precipitation for each month. 62
14	The estimated 95th marginal percentiles of the multivariate log-skew- t distributions fitted to the precipitation for each month. 63
15	Normal residuals density estimate. 70
16	Skew-normal residuals density estimate. 72
17	Skew- t residuals density estimate. 73
18	The syntax for fitting univariate skew-normal regression. 80
19	The syntax for fitting multivariate skew-normal regression. 80
20	The syntax for fitting univariate skew- t regression. 80
21	The syntax for fitting multivariate skew- t regression. 81
22	The syntax for obtaining predictions. 81
23	The syntax for producing goodness-of-fit plots. 81
24	Probability plot for the skew-normal model. 87
25	Probability plots for the skew-normal model (left panel) and the skew- t model (right panel). 89
26	Scatter plot of <code>lbn</code> and <code>bmi</code> 92
27	Probability plot for bivariate skew-normal and Normal model. 95
28	Normal probability plot of automobile prices. 97
29	Skew- t probability plot of automobile prices. 99
30	Skewness indexes over months with 95% confidence intervals. 101

CHAPTER I

INTRODUCTION

A. Overview of Skewed Distributions

In the past decade, there has been much growth in application and in research in the area of skewed distributions, a class of flexible parametric distributions accommodating departures from symmetry. This class originates from a perturbation of symmetric density functions in a multiplicative fashion which introduces skewness. Besides the flexibility in modeling skewed data, often arising in practice, the increasing popularity of skewed distributions can be attributed to a number of reasons. First, skewed distributions are simple extensions of their symmetric counterparts and include the latter as special cases. Second, skewed distributions retain a number of formal properties of standard symmetric distributions. Third, skewed distributions have tractable multivariate versions. Fourth, many of the skewed distributions arise as a result of some selection mechanism.

Although the idea of skewed distributions existed for a long time in the literature, it was formally introduced in Azzalini (1985). The (standard) skew-normal distribution, $\text{SN}(\alpha)$, is the simplest representative of the class of skewed distributions and has the density

$$f_{\text{SN}}(z; \alpha) = 2 \phi(z) \Phi(\alpha z), \quad z \in \mathbb{R}, \quad (1.1)$$

where the shape parameter $\alpha \in \mathbb{R}$ regulates the skewness of the distribution, and $\phi(\cdot)$ and $\Phi(\cdot)$ denote the density and the cumulative distribution function, respectively, of the standard normal distribution. When $\alpha = 0$ the density (1.1) becomes the density

This dissertation follows the style of *American Statistical Association*.

of the standard normal distribution, when $\alpha > 0$ the density (1.1) is skewed to the right, and when $\alpha < 0$ the density (1.1) is skewed to the left.

An alternative representation of the skew-normal distribution arises from a selection mechanism (Arellano-Valle et al. 2006). Suppose that a random vector (U, V) follows a standard bivariate normal distribution with correlation δ :

$$\begin{pmatrix} U \\ V \end{pmatrix} \sim N_2 \left\{ \begin{pmatrix} 0 \\ 0 \end{pmatrix}, \begin{pmatrix} 1 & \delta \\ \delta & 1 \end{pmatrix} \right\},$$

then $Z \stackrel{d}{=} (V|U > 0)$ has the density

$$f_{\text{SN}}(z; \delta) = 2 \phi(z) \Phi \left(\frac{\delta}{\sqrt{1 - \delta^2}} z \right), \quad z \in \mathbb{R}, \quad (1.2)$$

where $\delta \in (-1, 1)$. This representation is one of the stochastic representations of a skew-normal random variate and is referred to as the *conditioning method*. The parameterization (1.2) also corresponds to the so-called “ δ -parameterization” of the skew-normal distribution whereas (1.1) corresponds to the “ α -parameterization”. There is one-to-one correspondence between the two parameterizations with $\alpha = \delta/\sqrt{1 - \delta^2}$ and $\delta = \alpha/\sqrt{1 + \alpha^2}$.

The location-scale version of the skew-normal distribution $\text{SN}(\xi, \omega^2, \alpha)$ is defined in a standard fashion: if $Z \sim \text{SN}(\alpha)$, then $Y = \xi + \omega Z \sim \text{SN}(\xi, \omega^2, \alpha)$ and its density is

$$f_{\text{SN}}(y; \xi, \omega, \alpha) = \frac{2}{\omega} \phi(z) \Phi(\alpha z), \quad y \in \mathbb{R}, \quad (1.3)$$

where $z = (y - \xi)/\omega$. Unlike the normal distribution, ξ and ω^2 do not correspond to the mean μ and variance σ^2 of the skew-normal random variate unless $\alpha = 0$. The mean and variance are functions of ξ , ω , and α .

There is yet another parameterization associated with the skew-normal distri-

bution that is appealing from estimation and interpretation standpoints. The most common method of estimation of the parameters in (1.3) is maximum likelihood. It has been shown by Azzalini (1985), among others, that the profile log-likelihood for α (or δ) has a stationary point at $\alpha = 0$ (or $\delta = 0$). This unfortunate property leads to a singular Fisher information matrix at $\alpha = 0$ and, thus, creates difficulty for testing the important hypothesis of normality $H_o: \alpha = 0$ or, equivalently, $H_o: \delta = 0$ within the likelihood framework. Also, the sampling distributions of estimated parameters are far from being symmetric and sometimes are even bimodal for moderate sample size. To alleviate this, Azzalini (1985) proposed the *centered parameterization*. Instead of maximizing the log-likelihood based on *direct parameters* (ξ, ω, α) , the log-likelihood is reformulated and is maximized using *centered parameters* (μ, σ, γ) , where γ is the skewness index. The estimates of the direct parameters can then be obtained from the estimates of centered parameters using the one-to-one correspondence between the two parameterizations.

As we mentioned earlier, the skew-normal distribution is the simplest member of a class of skewed distributions. Generally, any symmetric (univariate or multivariate) density $f(\cdot)$, a base density, can be used instead of $\phi(\cdot)$ and any univariate differentiable distribution function $G\{w(z)\}$, a skewing function, with a density symmetric about 0 can be used instead of $\Phi(\alpha z)$ in (1.1), where $w(z)$ is a real-valued function such that $w(-z) = -w(z)$ for all z . More generally, if one employs the selection approach, then for continuous random vectors $\mathbf{V} \in \mathbb{R}^d$ and $\mathbf{U} \in \mathbb{R}^p$, the distribution with density $f_{\mathbf{Z}}(\mathbf{z}) = f_{\mathbf{V}}(\mathbf{z})P(\mathbf{U} \in C|\mathbf{V} = \mathbf{z})/P(\mathbf{U} \in C)$ underlying the selection mechanism $\mathbf{Z} \stackrel{d}{=} (\mathbf{V}|\mathbf{U} \in C)$, where C is a measurable subset of \mathbb{R}^p such that $0 < P(\mathbf{U} \in C) < 1$, belongs to a class of skewed distributions.

In fact, the multivariate version of (1.1), $\text{SN}_d(\mathbf{0}, \bar{\boldsymbol{\Omega}}, \boldsymbol{\alpha})$, arises when $\phi(\cdot)$ is replaced with a multivariate standard normal density $\phi_d(\cdot)$ with the correlation matrix $\bar{\boldsymbol{\Omega}}$ and

the function αz is replaced by a linear function $\boldsymbol{\alpha}^\top \mathbf{z}$, where the shape is now regulated by a vector $\boldsymbol{\alpha} = (\alpha_1, \dots, \alpha_d)^\top \in \mathbb{R}^d$:

$$f_{\text{SN}_d}(\mathbf{z}; \bar{\boldsymbol{\Omega}}, \boldsymbol{\alpha}) = 2 \phi_d(\mathbf{z}; \bar{\boldsymbol{\Omega}}) \Phi(\boldsymbol{\alpha}^\top \mathbf{z}), \quad \mathbf{z} \in \mathbb{R}^d.$$

The location-scale version can be defined in a similar manner as for the univariate case by considering a random vector $\mathbf{Y} = \boldsymbol{\xi} + \boldsymbol{\omega} \mathbf{Z}$, where $\mathbf{Z} \sim \text{SN}_d(\mathbf{0}, \bar{\boldsymbol{\Omega}}, \boldsymbol{\alpha})$, $\boldsymbol{\xi} \in \mathbb{R}^d$ is the location vector, and $\boldsymbol{\Omega} = \{(\omega_{ij})\}_{i,j=1}^d \in \mathbb{R}^{d \times d}$ is the scale matrix, such that $\bar{\boldsymbol{\Omega}} = \boldsymbol{\omega}^{-1} \boldsymbol{\Omega} \boldsymbol{\omega}^{-1}$, where $\boldsymbol{\omega} = \text{diag}\{\sqrt{\omega_{11}}, \dots, \sqrt{\omega_{dd}}\}$. The multivariate skew-normal density is:

$$f_{\text{SN}_d}(\mathbf{y}; \boldsymbol{\xi}, \boldsymbol{\Omega}, \boldsymbol{\alpha}) = 2 \phi_d(\mathbf{y}; \boldsymbol{\xi}, \boldsymbol{\Omega}) \Phi(\boldsymbol{\alpha}^\top \mathbf{z}), \quad \mathbf{y} \in \mathbb{R}^d, \quad (1.4)$$

where $\phi_d(\mathbf{y}; \boldsymbol{\xi}, \boldsymbol{\Omega})$ is the density of a d -dimensional normal distribution with mean $\boldsymbol{\xi}$ and covariance $\boldsymbol{\Omega}$, and $\mathbf{z} = \boldsymbol{\omega}^{-1}(\mathbf{y} - \boldsymbol{\xi})$.

Another popular representative of the family of skewed distributions is the skew- t distribution, introduced by Azzalini and Capitanio (2003), for which the symmetric base distribution is a heavy-tailed, Student's t distribution:

$$f_{\text{ST}}(z; \alpha, \nu) = 2 t(z; \nu) T \left\{ \left(\frac{\nu + 1}{\nu + z^2} \right)^{1/2} \alpha z; \nu + 1 \right\}, \quad z \in \mathbb{R}, \quad (1.5)$$

where $t(\cdot; \nu)$ and $T(\cdot; \nu)$ denote the density and the cumulative distribution function, respectively, of the standard Student's t distribution with ν degrees of freedom. The skew-normal distribution is a special case of the skew- t distribution (1.5) when $\nu = \infty$. When $\alpha = 0$, the density (1.5) reduces to the Student's t density and to the normal density when in addition $\nu = \infty$. The popularity of the skew- t distribution is attributed to its flexibility in capturing two common departures from normality arising in practice — asymmetry and heavy tails. Similar to the skew-normal distribution, the conditioning method can be used to formulate the skew- t distribution in

the δ -parameterization. The centered parameterization of the skew- t distribution is still under development. The location-scale and multivariate versions of the skew- t distribution can be defined analogously to the skew-normal distribution. For example, the density of a multivariate skew- t distribution with location $\boldsymbol{\xi} \in \mathbb{R}^d$, scale matrix $\boldsymbol{\Omega} \in \mathbb{R}^{d \times d}$, the vector of shape parameters $\boldsymbol{\alpha} \in \mathbb{R}^d$, and the degrees-of-freedom parameter $\nu > 0$, $\text{ST}_d(\boldsymbol{\xi}, \boldsymbol{\Omega}, \boldsymbol{\alpha}, \nu)$, is

$$f_{\text{ST}_d}(\mathbf{y}; \boldsymbol{\xi}, \boldsymbol{\Omega}, \boldsymbol{\alpha}, \nu) = 2 t_d(\mathbf{y}; \boldsymbol{\xi}, \boldsymbol{\Omega}, \nu) T \left\{ \left(\frac{\nu + d}{\nu + Q_y} \right)^{1/2} \boldsymbol{\alpha}^\top \mathbf{z}; \nu + d \right\}, \mathbf{y} \in \mathbb{R}^d, \quad (1.6)$$

where $\mathbf{z} = \boldsymbol{\omega}^{-1}(\mathbf{y} - \boldsymbol{\xi})$, $Q_y = (\mathbf{y} - \boldsymbol{\xi})^\top \boldsymbol{\Omega}^{-1}(\mathbf{y} - \boldsymbol{\xi})$, and $t_d(\cdot)$ is the density of a d -dimensional Student's t distribution with location $\boldsymbol{\xi}$, scale matrix $\boldsymbol{\Omega}$ and degrees of freedom ν .

Besides the extra flexibility in modeling the tail behavior, the skew- t distribution has another advantage over a more simple skew-normal model. The singularity of the Fisher information matrix at $\alpha = 0$ (or $\delta = 0$) is not observed for the skew- t model unless $\nu = \infty$. This as well as its robustness properties make it an appealing alternative to the skew-normal or normal distribution.

The distributions discussed above have a number of appealing properties such as the distribution of the quadratic forms does not depend on the shape parameter α , the closure under affine transformations and marginalization. One useful property they are lacking is the closure under conditioning. This issue is circumvented by considering the extended versions of the skewed distributions described above introducing an extra, shift parameter $\tau \in \mathbb{R}$. Specifically, the multivariate extended skew-normal density (Capitanio et al. 2003) is

$$f_{\text{ESN}_d}(\mathbf{y}; \boldsymbol{\xi}, \boldsymbol{\Omega}, \boldsymbol{\alpha}, \tau) = \phi_d(\mathbf{y}; \boldsymbol{\xi}, \boldsymbol{\Omega}) \frac{\Phi(\boldsymbol{\alpha}^\top \mathbf{z} + \tau)}{\Phi(\tau / \sqrt{1 + \boldsymbol{\alpha}^\top \boldsymbol{\Omega} \boldsymbol{\alpha}})}, \mathbf{y} \in \mathbb{R}^d. \quad (1.7)$$

The multivariate extended skew- t density (Arellano-Valle and Genton 2010a) is

$$f_{\text{EST}_d}(\mathbf{y}; \boldsymbol{\xi}, \boldsymbol{\Omega}, \boldsymbol{\alpha}, \tau, \nu) = t_d(\mathbf{y}; \boldsymbol{\xi}, \boldsymbol{\Omega}, \nu) \frac{T \left\{ \left(\frac{\nu+d}{\nu+Q_y} \right)^{1/2} (\boldsymbol{\alpha}^\top \mathbf{z} + \tau); \nu + d \right\}}{T(\tau/\sqrt{1 + \boldsymbol{\alpha}^\top \boldsymbol{\Omega} \boldsymbol{\alpha}}; \nu)}, \mathbf{y} \in \mathbb{R}^d. \quad (1.8)$$

Densities (1.4) and (1.6) are special cases of (1.7) and (1.8), respectively, when $\tau = 0$. The extended versions of the skew-normal and skew- t distributions are closed under conditioning and also can model lighter tails than the normal distribution. In fact, the conditional distribution of the skew-normal and skew- t random vectors are of the form (1.7) and (1.8), respectively. The extended skewed distributions also arise naturally in the sample-selection setting, as we discuss in more detail in Chapter II.

The normal and Student's t distributions are special cases of a more general family of elliptically-contoured distributions, distributions with contours of constant density representing ellipsoids. We will also consider in Chapters II and III a more general family of skew-elliptical distributions which arises from (1.1), when the base symmetric (multivariate) density and the skewing function are the density and the cumulative distribution function, respectively, of an elliptically-contoured distribution.

B. Motivation

Below I briefly describe the proposed novel approaches for analysis and modeling using multivariate skew-normal and skew- t distributions in econometrics and environmetrics as well as motivation for each topic.

Heckman selection- t model. This work was prompted by the link between selection mechanisms such as (1.2) and the family of skewed distributions. More specifically, there is a link between the classical Heckman sample-selection model, one of the most popular models in econometrics, and the extended skew-normal distribution

describing the continuous part of the selection model. The maximum likelihood estimation method, originally proposed by Heckman under the assumption of normality (Heckman 1974), despite its efficiency, was criticized for sensitivity to the violation of normality and also for collinearity problems arising in certain cases. Taking into account the robustness properties of the skew- t distribution, it is natural to consider the extended skew- t distribution for modeling the continuous part of the selection model. Using the link between selections and skewed distributions, we demonstrate that the extended skew- t model arises when the Student's t distribution is assumed for the error distribution. Using this fact, we introduce a more robust sample-selection model, the selection- t model, and study its properties.

Multivariate log-skew-elliptical distributions with applications to precipitation data. The log-normal distribution is often used to analyze data with nonnegative support such as precipitation data, for example. When heavy tails are likely, as is the case for various financial data, the log- t distribution is considered for the analysis. Both of these distributions, however, assume a symmetric distribution for the data in the log scale which is not always plausible in practice. The assumption of symmetry can be relaxed by considering log-skew-normal and log-skew- t distributions. Azzalini et al. (2003) successfully applied the univariate log-skew-normal and log-skew- t distributions for the analysis of family income data. Their encouraging findings prompted us to introduce and investigate formal properties of a more general family of multivariate log-skew-elliptical distributions and to extend the areas of application to environmetrics by using these distributions to analyze U.S. precipitation data (Marchenko and Genton 2010*b*).

A suite of commands for fitting the skew-normal and skew- t models. A large number of research articles appeared in the literature demonstrating the use of skewed distributions in various applications, but the use of these distributions by

practitioners is still limited. Two common reasons for low popularity of some methodologies in practice are the lack of awareness by practitioners of new methods and the absence of user-friendly software implementing the methods to perform the analysis. Currently, a number of methods for fitting the skew-normal and skew- t distributions are available in the package `sn`, developed by Azzalini, with a free software package, R (Azzalini 2006). We develop a suite of commands to perform analysis using skewed distributions in a commercial software package, Stata (StataCorp 2009), as a free user-written add-on and distribute it via publication in the peer-reviewed Stata Journal. This work was motivated by the desire to raise awareness of these distributions among the large Stata community as well as to satisfy the requests from a number of current Stata users. This suite of commands also offers some features which are not yet available in other software.

CHAPTER II

A HECKMAN SELECTION- T MODEL

Sample selection arises frequently in applications in many fields including economics, biostatistics, finance, sociology, and political science, to name a few. Sample selection is a special case of a more general concept known in the econometrics literature as limited dependent variables — variables observed over a limited range of their support.

Let $Y^* \in \mathbb{R}$ be our outcome of interest. Suppose that we observe $Y = Y^*$ only when some unobserved random variable $U^* \in \mathbb{R}$ belongs to a subset $C \subset \mathbb{R}$ of its support, such that $0 < P(U^* \in C) < 1$. That is, Y is subject to hidden truncation (or simply truncation when $U^* = Y^*$). Model parameters underlying Y^* are then estimated from the observed Y using the conditional density $f(Y|U^* \in C)$. In practice, truncation arises when the collected sample represents only a subset of a full population, for example, a sample of individuals with incomes below or above some threshold. Sometimes the collected sample does represent a full population but because of some hidden truncation $U^* \in \mathbb{R}$ the outcome of interest Y^* is not observed for all of the participants. In this case, Y^* is subject to incidental truncation or sample selection (e.g., Greene 2008). The problem of sample-selection or, more specifically, sample-selection bias, arises when Y^* and U^* are correlated and, thus, must be modeled jointly. That is, inference based on only observed Y would not be valid. This problem is also known as data missing not at random (MNAR, Rubin 1976). For example, in a study of incomes, people with high (or low) income may be less likely to report it than people with average income. In the presence of sample selection we observe an indicator $U = I(U^* \in C)$ and other explanatory variables for all of a sample and the outcome $Y = Y^*$ for part of the sample. Thus, the

sample-selection model is comprised of the continuous component $f(Y|U = 1)$ and the discrete component $P(U)$.

The classical sample-selection model was introduced by Heckman (1974) in the mid 1970s when he proposed a parametric approach to the estimation under the assumption of bivariate normality between Y^* and U^* . The main criticism of the proposed method was the sensitivity of the parameter estimates to the assumption of normality, often violated in practice, which led to his developing a more robust estimation procedure in the late 1970s, known as Heckman's two-step estimator (Heckman 1979). Both estimation methods were found to be sensitive to high correlation between variables of the outcome and selection equations, as is often encountered in practice (see, for example, Puhani 2000 and references therein). Various other robust-to-normality methods have been proposed over the years for the analysis of a sample-selection model, including a number of semiparametric (e.g., Ahn and Powell 1993, Newey 1999) and nonparametric (e.g., Das et al. 2003) methods, relaxing the distributional assumption in general. See, for example, Vella (1998), Li and Racine (2007, p. 315) for a general overview of methods for selection models.

In this chapter, we develop and study the properties of yet another estimation approach. Our approach maintains the original parametric framework of the model but considers a bivariate Student's t distribution as the underlying joint distribution of (Y^*, U^*) and estimates the parameters via maximum likelihood. Among various departures from normality occurring in practice, one of the most common is when the distribution of the data has heavier tails than the normal distribution, such as the distribution of (log) incomes in the population. This makes it natural to consider a Student's t distribution as an underlying joint distribution for the selection model. The robustness properties of the Student's t distribution (Lange et al. 1989, Azzalini and Genton 2008, DiCiccio and Monti 2009) also make it an attractive parametric

alternative to the normal distribution. For example, this distribution has been used recently to relax the assumption of normality in various statistical models such as censored regression (Muñoz-Gajardo et al. 2010), treatment models (Chib and Hamilton 2000), and switching regression (Scruggs 2007), to name a few. The selection bias test based on the selection-normal model can be affected by heavy tails, as we demonstrate in our simulation, and our method addresses this issue. Our motivation for considering the Student's t distribution is also prompted by the existence of a link between the continuous part of the selection model and an extended skew- t distribution, studied extensively in the recent literature (Arellano-Valle and Genton 2010a).

The chapter is organized as follows. Section A describes the classical sample-selection model and introduces the selection- t model. The finite-sample performance of the selection- t model is evaluated numerically and compared to that of the classical selection-normal model in Section B. A numerical application of the selection- t model is presented in Section C. The chapter concludes with a discussion in Section D. The relevant analytical results are given in the Appendix A.

A. Heckman Selection Models

In this section we first describe the classical sample-selection model and its two commonly used estimation methods, maximum likelihood and two step. Next, we comment on the link between sample-selection models and a family of skew-elliptical distributions. Finally, we formulate the sample selection- t model and study its properties.

1. Classical Heckman sample-selection model

Suppose that the regression model of primary interest is

$$y_i^* = \mathbf{x}_i^\top \boldsymbol{\beta} + \epsilon_i, \quad i = 1, \dots, N. \quad (2.1)$$

However, due to a certain selection mechanism,

$$u_i^* = \mathbf{w}_i^\top \boldsymbol{\gamma} + \eta_i, \quad i = 1, \dots, N, \quad (2.2)$$

we observe only N_1 out of N observations y_i^* for which $u_i^* > 0$:

$$\begin{aligned} u_i &= I(u_i^* > 0), \\ y_i &= y_i^* u_i. \end{aligned} \quad (2.3)$$

Latent variables $y_i^* \in \mathbb{R}$ and $u_i^* \in \mathbb{R}$ are associated with primary and selection regressions, respectively; y_i is the observed counterpart of y_i^* and u_i is an indicator of whether the primary dependent variable is observed. The vectors $\boldsymbol{\beta} \in \mathbb{R}^p$ and $\boldsymbol{\gamma} \in \mathbb{R}^q$ are unknown parameters; the vectors $\mathbf{x}_i \in \mathbb{R}^p$ and $\mathbf{w}_i \in \mathbb{R}^q$ are observed characteristics; and ϵ_i and η_i are error terms from a bivariate normal distribution:

$$\begin{pmatrix} \epsilon_i \\ \eta_i \end{pmatrix} \sim N_2 \left\{ \begin{pmatrix} 0 \\ 0 \end{pmatrix}, \begin{pmatrix} \sigma^2 & \rho\sigma \\ \rho\sigma & 1 \end{pmatrix} \right\}. \quad (2.4)$$

The selection-normal model (2.1)-(2.3) is known as “type 2 tobit model” in the econometrics literature (Amemiya 1985, p. 385) and is sometimes also referred to as the “Heckman model”.

Because we only observe the sign of u_i^* in (2.3), its variance is nonidentifiable and, w.l.g., is set to 1 in (2.4). The parameter $\rho \in (-1, 1)$ governs the “selection bias” which arises when $\rho \neq 0$ and the standard ordinary least squares (OLS) regression is used to estimate $\boldsymbol{\beta}$ in (2.1). The zero threshold in (2.3) is arbitrary; any other

constant threshold $c \neq 0$ would be absorbed by the intercept in (2.2).

As we mentioned in the introduction, the density of the sample-selection model is comprised of a continuous component corresponding to the conditional density $f(y|U = 1)$ and a discrete component $P(U)$. The discrete component is described by the probit model $P(U = u) = \{\Phi(\mathbf{w}^\top \boldsymbol{\gamma})\}^u \{\Phi(-\mathbf{w}^\top \boldsymbol{\gamma})\}^{1-u}$, where $\Phi(\cdot)$ is the standard normal cumulative distribution function. The conditional density is

$$f(y|U = 1) = \frac{1}{\sigma} \phi\left(\frac{y - \mathbf{x}^\top \boldsymbol{\beta}}{\sigma}\right) \frac{\Phi\left\{\frac{\rho}{\sqrt{1-\rho^2}}\left(\frac{y - \mathbf{x}^\top \boldsymbol{\beta}}{\sigma}\right) + \frac{\mathbf{w}^\top \boldsymbol{\gamma}}{\sqrt{1-\rho^2}}\right\}}{\Phi(\mathbf{w}^\top \boldsymbol{\gamma})}, \quad (2.5)$$

where $\phi(\cdot)$ denotes the standard normal density.

Then, the (partial) log-likelihood function for this model based on a single pair of observations (y, u) can be written as

$$\begin{aligned} l(\boldsymbol{\beta}, \boldsymbol{\gamma}, \rho, \sigma; y, u) &= u \ln f(y|U = 1) + u \ln \Phi(\mathbf{w}^\top \boldsymbol{\gamma}) + (1 - u) \ln \Phi(-\mathbf{w}^\top \boldsymbol{\gamma}) \\ &= u \left[\ln \Phi\left\{\frac{\mathbf{w}^\top \boldsymbol{\gamma} + \rho(y - \mathbf{x}^\top \boldsymbol{\beta})/\sigma}{\sqrt{1-\rho^2}}\right\} - \frac{1}{2} \left(\frac{y - \mathbf{x}^\top \boldsymbol{\beta}}{\sigma}\right)^2 - \ln(\sqrt{2\pi}\sigma) \right] \\ &\quad + (1 - u) \ln \Phi(-\mathbf{w}^\top \boldsymbol{\gamma}). \end{aligned} \quad (2.6)$$

The term partial likelihood reflects the fact that only observed values y contribute to the likelihood. Heckman (1974) showed that the maximum likelihood estimators (MLEs) obtained from the maximization of (2.6) have the same properties as the conditional MLEs, although (2.6) does not correspond to a true conditional density. When the distributional assumption of bivariate normality is correct, MLEs are fully efficient.

A method more robust to the normality assumption was proposed by Heckman (1979) and is known as the two-step estimation. The motivation for the two-step estimator is based on the fact that the conditional expectation of the observed data

is

$$E(Y|U^* > 0, \mathbf{x}, \mathbf{w}) = \mathbf{x}^\top \boldsymbol{\beta} + \rho\sigma\lambda(\mathbf{w}^\top \boldsymbol{\gamma}), \quad (2.7)$$

where $\lambda(\cdot) = \phi(\cdot)/\Phi(\cdot)$ denotes the inverse Mills ratio. The inconsistency of the classical OLS estimator from regression of y on \mathbf{x} is explained by the existence of an extra term $\rho\sigma\lambda(\mathbf{w}^\top \boldsymbol{\gamma})$ when $\rho \neq 0$ involved in the regression function (2.7), which is omitted from the regression. The two-step method is designed to correct the OLS regression for the omitted term and consists of two stages. At the first stage, the probit model $P(U|\mathbf{w})$ is fit to the data and the ML estimates of $\boldsymbol{\gamma}$ are obtained. At the second stage, $\boldsymbol{\beta}$ and $\beta_\lambda = \rho\sigma$ are estimated by least squares regression of y on \mathbf{x} and $\hat{\lambda}$, where $\hat{\lambda} = \phi(\mathbf{w}^\top \hat{\boldsymbol{\gamma}})/\Phi(\mathbf{w}^\top \hat{\boldsymbol{\gamma}})$. The consistent estimates of ρ and σ can be obtained from $\hat{\beta}_\lambda$, least squares residual variance and average predicted probabilities from the probit model (e.g., Greene 2008, p. 886). The advantage of the two-step method over ML is that no distributional assumption is required for the error terms for the consistency of the estimator. However, this method suffers from possible collinearity problems when \mathbf{w} includes some of the covariates from the primary regression because of the linearity of the inverse Mills ratio $\lambda(\cdot)$ in the wide range of its support. In fact, both methods tend to perform poorly in the presence of high correlation between error terms and high collinearity among regressors in the primary and selection equations.

The sensitivity of the two methods to the collinearity among regressors in the primary and selection equations has been shown to be even more of an issue in practical applications than the misspecified error distribution (Puhani 2000). When $\mathbf{w} = \mathbf{x}$, which is not an unusual assumption in many practical settings, the identifiability of the parameters relies heavily on the functional form of the distribution. In particular, for the selection-normal model the identification of the regression parameters

is achieved through the nonlinearity of the inverse Mills ratio. The problem arises because of the linearity of the inverse Mills ratio $\lambda(\cdot)$ in the wide range of its support (see Subsection 3 of Section A of this chapter). To alleviate this problem the econometrics literature suggests to impose an exclusion restriction according to which at least one extra variable which is a good predictor of u_i^* is included in the selection equation and does not appear in the primary regression. In practice, however, it can be difficult to find such variables because strong predictors of the selection equation are usually also strong predictors of the primary equation and thus should be included in the primary regression as well.

2. Link to extended skew-elliptical distributions

The continuous component of the sample-selection density, the conditional density (2.5), corresponds to the extended skew-normal distribution, studied in more detail recently by Arellano-Valle and Genton (2010a),

$$f_{\text{ESN}}(y; \mu, \sigma^2, \alpha, \tau) = \frac{1}{\sigma} \phi\left(\frac{y - \mu}{\sigma}\right) \frac{\Phi\left(\alpha \frac{y - \mu}{\sigma} + \tau\right)}{\Phi(\tau / \sqrt{1 + \alpha^2})}, \quad y \in \mathbb{R}, \quad (2.8)$$

with parameterization $\mu = \mathbf{x}^\top \boldsymbol{\beta} \in \mathbb{R}$, $\alpha = \rho / \sqrt{1 - \rho^2} \in \mathbb{R}$ and $\tau = \mathbf{w}^\top \boldsymbol{\gamma} / \sqrt{1 - \rho^2} \in \mathbb{R}$. The parameterization of the mean μ is simply the conventional parameterization used in the regression setting. The parameterization of the shape parameter α corresponds to the so-called “ δ -parameterization” ($\delta = \rho$ in this case) arising from the stochastic representation of a skew-normal random variable (Azzalini 2005), also discussed below. The key distinction here is the parameterization of the shift parameter τ . In the sample-selection setting, similar to the mean μ , τ is parameterized as a linear function of the predictors \mathbf{w} from the selection equation. Thus, the model includes a $q \times 1$ vector of unknown coefficients $\boldsymbol{\gamma}$ rather than a single shift parameter τ .

Azzalini (1985) and Copas and Li (1997), among others, noted that the distribution of $Y^*|U^* > 0$, arising from hidden truncation when Y^* and U^* are jointly normal and the marginal distribution of U^* is standard normal, belongs to the family of skew-normal distributions (Azzalini 1985, 2005, Genton 2004), which is a special case of (2.8) with $\tau = 0$. In fact, the selection mechanism $Y^*|U^* > 0$ corresponds to one of the stochastic representations of a skew-normal random variate, when the location of U^* is zero, and an extended skew-normal random variate, when the location of U^* is different from zero. Based on this link, it would be natural to use the skew-normal distribution to model data arising from hidden truncation under the assumption of the underlying bivariate normal distribution. For example, we can fit the skew-normal model, available in statistical packages R (Azzalini 2006) and Stata (Marchenko and Genton 2010a), to the observed data to account for asymmetry in the distribution often induced by hidden truncation. Even in the sample-selection setting, if we fit the extended skew-normal model (2.8) with regression parameterization of τ and μ to the observed data only we will obtain consistent (albeit inefficient) results, unlike the OLS regression.

We can also use this link to help study the properties of the sample-selection model. From (2.6), the log-likelihood for the selection-normal model consists of two components. The first component is the log-likelihood of the extended skew-normal model for the observed data and the second component is the probit log-likelihood describing the selection process. It is known that both the profile log-likelihood for α (or ρ in the “ δ -parameterization”) of the skew-normal and extended skew-normal models has a stationary point at $\alpha = 0$ (or $\rho = 0$) which leads to the singularity of the Fisher information and observed information matrices at that point. The singularity is caused by the chosen parameterization as opposed to the unidentifiability of the model parameters in general. The sample-selection model uses a similar parameteriza-

tion. Also, in the sample-selection framework, the hypothesis $H_0: \rho = 0$ is important for testing the existence of the sample-selection bias. Thus, the stationarity of the likelihood at $\rho = 0$ would create difficulty in testing this hypothesis within the likelihood framework. As it turns out, the selection-normal model does not exhibit this property. The stationarity issue for the extended skew-normal model arises because the scores of the parameters are linearly dependent at $\alpha = 0$ (or $\rho = 0$) (Arellano-Valle and Genton 2010a). The scores for β , σ , and ρ for the selection-normal model are linearly related. However, the score for γ is not zero, unlike the score for τ for the extended skew-normal model, and is not linearly dependent with any of the other scores. Hence, the observed information is not singular at $\rho = 0$. See Appendix A3 for details.

More generally, Arellano-Valle et al. (2006) unified all of the distributions arising from selection mechanisms $\mathbf{Y}^*|\mathbf{U}^* \in C$, where $\mathbf{Y}^* \in \mathbb{R}^{d_1}$, $\mathbf{U}^* \in \mathbb{R}^{d_2}$, and C is a measurable subset in \mathbb{R}^{d_2} such that $0 < P(\mathbf{U}^* \in C) < 1$, in a broad class of selection distributions. For example, if (Y^*, U^*) follow a bivariate elliptically contoured distribution (e.g., Fang et al. 1990), then $Y^*|U^* > 0$ has an extended skew-elliptical distribution (Arellano-Valle and Azzalini 2006, Arellano-Valle and Genton 2010b). Specifically, let $\text{EC}_2(\boldsymbol{\xi}, \boldsymbol{\Omega}, g^{(2)})$ denote a family of bivariate elliptically contoured distributions (with existing density) with a generator function $g^{(2)}(\cdot)$ defining a spherical bivariate density, a location column vector $\boldsymbol{\xi} \in \mathbb{R}^2$, and a positive definite scale matrix $\boldsymbol{\Omega} \in \mathbb{R}^{2 \times 2}$. If $\mathbf{X} \sim \text{EC}_2(\boldsymbol{\xi}, \boldsymbol{\Omega}, g^{(2)})$, then its density is $f_2(\mathbf{x}; \boldsymbol{\xi}, \boldsymbol{\Omega}, g^{(2)}) = |\boldsymbol{\Omega}|^{-1/2} g^{(2)}(Q_x)$, where $Q_x = (\mathbf{x} - \boldsymbol{\xi})^\top \boldsymbol{\Omega}^{-1} (\mathbf{x} - \boldsymbol{\xi})$ and $\mathbf{x} \in \mathbb{R}^2$ (Fang et al. 1990, p. 46). If

$$\begin{pmatrix} Y^* \\ U^* \end{pmatrix} \sim \text{EC}_2 \left\{ \boldsymbol{\xi} = \begin{pmatrix} \mu \\ \mu_u \end{pmatrix}, \boldsymbol{\Omega} = \begin{pmatrix} \sigma^2 & \rho\sigma \\ \rho\sigma & 1 \end{pmatrix}, g^{(2)} \right\}, \quad (2.9)$$

then $Y^*|U^* > 0 \sim \text{ESE}(\mu, \sigma^2, \alpha, \tau, g_z)$, an extended skew-elliptical distribution with

the location parameter μ , the scale parameter σ^2 , the shape parameter $\alpha = \rho/\sqrt{1-\rho^2}$, and the shift parameter $\tau = \mu_u/\sqrt{1-\rho^2}$. The corresponding extended skew-elliptical density is

$$f_{\text{ESE}}(y; \mu, \sigma^2, \alpha, \tau, g_z) = \frac{1}{\sigma} f(z; g^{(1)}) \frac{F(\alpha z + \tau; g_z)}{F(\tau/\sqrt{1+\alpha^2}; g^{(1)}), \quad y \in \mathbb{R}, \quad (2.10)$$

where $z = (y - \mu)/\sigma$, $g_z(v) = g^{(2)}(v + z^2)/g^{(1)}(v)$, $f(\cdot; g^{(1)})$ and $F(\cdot; g^{(1)})$ are the density and cumulative distribution function, respectively, of a univariate standard elliptical distribution with a generator function $g^{(1)}(\cdot)$, and $F(\cdot; g_z)$ is the cumulative distribution function of a univariate standard elliptical distribution with a generator function $g_z(\cdot)$. The extended skew-normal density (2.8) is a special case of (2.10) with the normal generator function $g^{(2)}(v) = \frac{1}{2\pi} e^{-v/2}$. When $\mu_u = 0$, (2.10) reduces to the family of skew-elliptical distributions studied by Branco and Dey (2001).

We can use this more general link to build more flexible parametric sample-selection models relaxing the classical assumption of underlying normality as we demonstrate in the next subsection using the Student's t distribution. If we consider an underlying bivariate elliptically-contoured distribution (2.9), the continuous component of the resulting sample-selection model will correspond to the extended skew-elliptical distribution (2.10). Following (2.6), the likelihood function will include the likelihood for the extended skew-elliptical model and the likelihood for the corresponding binary elliptical model. In the spirit of the selection-normal model, we can investigate the properties of such flexible sample-selection models using some properties established for extended skew-elliptical distributions.

3. Selection- t model

Using the link between sample-selection models and extended skew-elliptical distributions, discussed in the previous subsection, we relax the assumption of bivariate

normality and consider the case when the underlying error distribution is a bivariate Student's t distribution. That is, a selection- t model is defined by (2.1)-(2.3) with bivariate Student's t error distribution:

$$\begin{pmatrix} \epsilon_i \\ \eta_i \end{pmatrix} \sim t_2 \left\{ \begin{pmatrix} 0 \\ 0 \end{pmatrix}, \begin{pmatrix} \sigma^2 & \rho\sigma \\ \rho\sigma & 1 \end{pmatrix}, \nu \right\}, \quad (2.11)$$

where $t_2(\mathbf{y}; \boldsymbol{\mu}, \boldsymbol{\Omega}, \nu) = \frac{1}{2\pi} |\boldsymbol{\Omega}|^{-1/2} \left\{ 1 + \frac{(\mathbf{y}-\boldsymbol{\mu})^\top \boldsymbol{\Omega}^{-1} (\mathbf{y}-\boldsymbol{\mu})}{\nu} \right\}^{-(\nu+2)/2}$ is the density of a bivariate Student's t distribution. In this case $\rho = 0$ does not imply independence of the primary and selection equations as for the selection-normal model, unless $\nu = \infty$.

When errors are nonnormally distributed, Lee (1982, 1983) proposed a general estimation method transforming the error components to bivariate normality and applied it to the case when error distribution is bivariate t . We focus on the full maximum likelihood estimation of the selection- t model (2.1)-(2.3), (2.11), which is fully efficient under the bivariate- t assumption.

The bivariate Student's t distribution from (2.11) corresponds to the bivariate elliptical distribution (2.9) with the generator function $g^{(2)}(v) = \frac{1}{2\pi} \nu^{(\nu+2)/2} (1+v)^{-(\nu+2)/2}$. Then, from (2.10), the distribution of $Y^*|U^* > 0$ is extended skew- t with density

$$f_{\text{EST}}(y; \mu, \sigma^2, \alpha, \tau, \nu) = \frac{1}{\sigma} t(z; \nu) \frac{T \left\{ (\alpha z + \tau) \left(\frac{\nu+1}{\nu+z^2} \right)^{1/2}; \nu+1 \right\}}{T(\tau/\sqrt{1+\alpha^2}; \nu)}, \quad y \in \mathbb{R}, \quad (2.12)$$

where $z = (y - \mu)/\sigma$, and $t(\cdot; \nu)$ and $T(\cdot; \nu)$ are the density and the cumulative distribution function of a univariate Student's t distribution with ν degrees of freedom. Thus, the density $f(y|U = 1)$, corresponding to the continuous component of the selection- t model, is described by (2.12) with parameterization $\mu = \mathbf{x}^\top \boldsymbol{\beta}$, $\alpha = \rho/\sqrt{1-\rho^2}$, and $\tau = \mathbf{w}^\top \boldsymbol{\gamma}/\sqrt{1-\rho^2}$. The extended skew- t distribution was

studied in detail in Arellano-Valle and Genton (2010a).

For the selection- t model, the conditional expectation of the observed data is

$$E(Y|U^* > 0, \mathbf{x}, \mathbf{w}) = \mathbf{x}^\top \boldsymbol{\beta} + \rho\sigma\lambda_\nu(\mathbf{w}^\top \boldsymbol{\gamma}), \quad \nu > 1, \quad (2.13)$$

where $\lambda_\nu(v) = \frac{\nu+v^2}{\nu-1} \frac{t(v;\nu)}{T(v;\nu)}$ (Lee 1983, Arellano-Valle and Genton 2010a). We can see that, similar to the selection-normal model, the conventional OLS regression of y on \mathbf{x} will produce inconsistent results when $\rho \neq 0$. We can visualize the impact of using the selection-normal model to model the regression function (2.13) by plotting functions $\lambda(\cdot)$ from (2.7) and $\lambda_\nu(\cdot)$ in Figure 1.

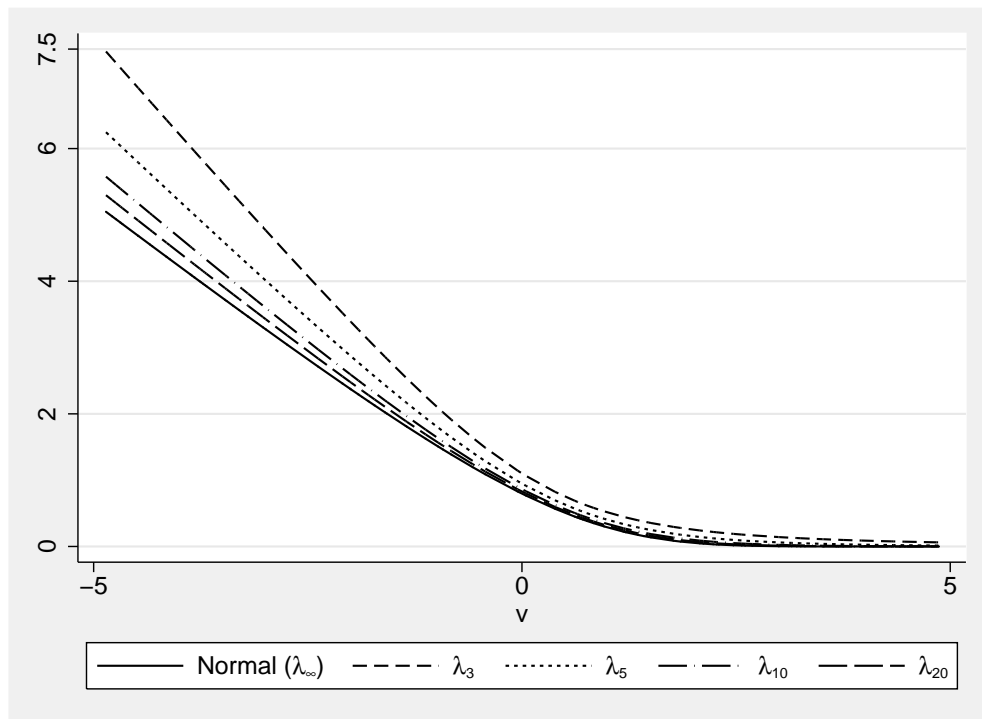


Figure 1. Plot of $\lambda_\nu(\cdot)$ for different values of ν with $\lambda_\infty(\cdot) = \lambda(\cdot)$, corresponding to the normal case.

From Figure 1 we can see that for negative values of the selection linear predictor $\mathbf{w}^\top \boldsymbol{\gamma}$, the conditional expectation will be underestimated under the selection-normal model for moderate values of the degrees of freedom ν . The difference diminishes as the degrees of freedom increase.

Marginal effects of the predictors on y in the observed sample are often of interest in practice. Suppose that $w_k = x_k$, then the conditional marginal effect of x_k on y under the selection- t model is

$$\frac{\partial E(Y|U^* > 0, \mathbf{x}, \mathbf{w})}{\partial x_k} = \beta_k + \rho\sigma\gamma_k\lambda'_\nu(\mathbf{w}^\top \boldsymbol{\gamma}), \quad \nu > 1,$$

where $\lambda'_\nu(v) = \frac{\partial \lambda_\nu(v)}{\partial v} = -\lambda_\nu(v) \left\{ v \frac{\nu+1}{\nu+v^2} + \lambda_\nu(v) \right\}$. Figure 2 depicts functions $\lambda'(\cdot)$ and $\lambda'_\nu(\cdot)$ for several degrees of freedom. From the graph, the conditional marginal effect of x_k on y will be overestimated by the selection-normal model for negative values of $\mathbf{w}^\top \boldsymbol{\gamma}$ and moderate degrees of freedom ν .

Similar to the selection-normal model, the log-likelihood function of the selection- t model can be decomposed into the log-likelihood of the extended skew- t distribution and the log-likelihood for the binary t model. From (2.12), the log-likelihood for the selection- t model based on a single pair of observations (y, u) is

$$\begin{aligned} l(\boldsymbol{\beta}, \boldsymbol{\gamma}, \rho, \sigma, \nu; y, u) &= u \ln f(y|U=1) + u \ln T(\mathbf{w}^\top \boldsymbol{\gamma}; \nu) + (1-u) \ln T(-\mathbf{w}^\top \boldsymbol{\gamma}; \nu) \\ &= u \ln t(z; \nu) - u \ln \sigma + u \ln T \left\{ \left(\frac{\nu+1}{\nu+z^2} \right)^{1/2} \frac{\rho z + \mathbf{w}^\top \boldsymbol{\gamma}}{\sqrt{1-\rho^2}}; \nu+1 \right\} \\ &\quad + (1-u) \ln T(-\mathbf{w}^\top \boldsymbol{\gamma}; \nu), \end{aligned} \quad (2.14)$$

where $z = (y - \mathbf{x}^\top \boldsymbol{\beta})/\sigma$. There are no closed-form expressions for the MLEs of the parameters in (2.14). Thus, the MLEs are obtained numerically using the Newton-Raphson algorithm. The scores and the Hessian matrix corresponding to (2.14) are given in Appendix A1 and Appendix A2, respectively.

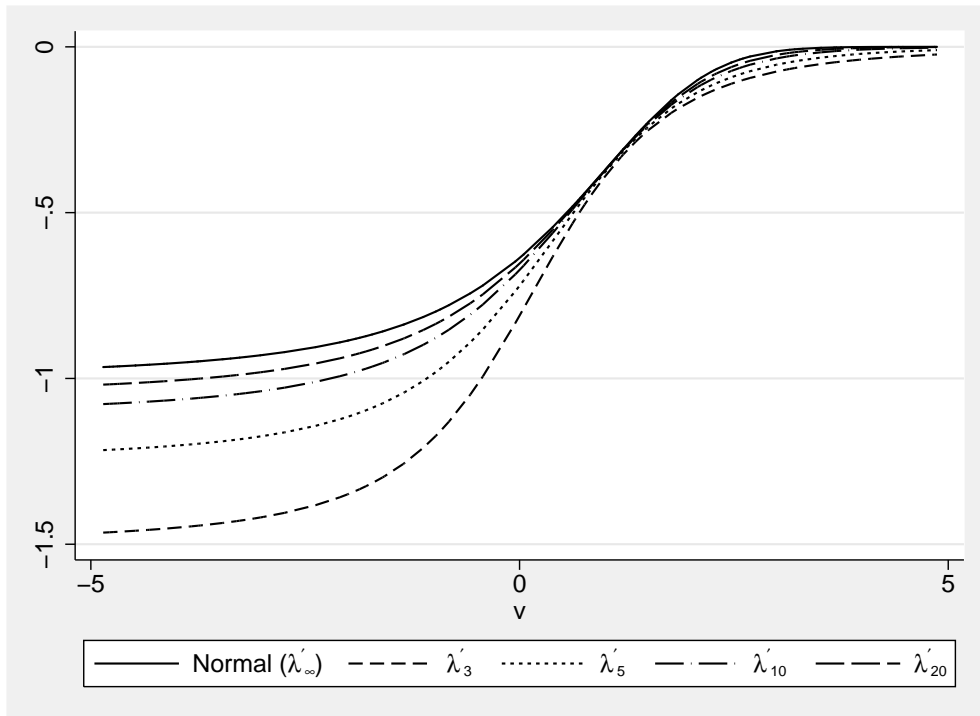


Figure 2. Plot of $\lambda'_\nu(\cdot)$ for different values of ν with $\lambda'_\infty(\cdot) = \lambda'(\cdot)$, corresponding to the normal case.

B. Monte Carlo Simulations

1. Finite-sample properties of the MLEs

To study finite-sample properties of the MLEs for the selection- t model, we consider several simulation scenarios. The primary regression is $y_i^* = 0.5 + 1.5x_i + \epsilon_i$, where $x_i \stackrel{\text{iid}}{\sim} N(0, 1)$ and $i = 1, \dots, N = 1000$. We consider two types of selection regressions: $u_i^* = 1 + x_i + 1.5w_i + \eta_i$, with exclusion restriction $w_i \stackrel{\text{iid}}{\sim} N(0, 1)$, and $u_i^* = 1 + x_i + \eta_i$, without the exclusion restriction. The covariates x_i and w_i are independent and

are also independent from the error terms ϵ_i and η_i . The error terms (ϵ_i, η_i) are generated from a bivariate Student's t distribution with $\nu = 5$ degrees of freedom and with the scale matrix $\mathbf{\Omega} = \begin{pmatrix} \sigma^2 & \rho\sigma \\ \rho\sigma & 1 \end{pmatrix}$, where $\sigma = 1$. We consider several values of correlation $\rho \in \{0, 0.2, 0.5, 0.7\}$. We observe only values y_i for which $u_i^* > 0$, that is $y_i = u_i y_i^*$, where $u_i = I(u_i^* > 0)$. In the considered scenarios, the degree of censoring corresponds to about 30% in the absence of the exclusion restriction and about 40% in the presence of the exclusion restriction.

We compare the performance of the selection- t model (SLt) to the selection-normal model (SLN) and the Heckman's two-step method (TS) when errors come from a bivariate Student's t distribution. Simulation results based on $R = 1000$ replications are presented in the Tables 1 and 2. The results are presented in the estimation metric with the support $(-\infty, \infty)$, where $\text{atanh } \rho$ is the inverse hyperbolic tangent of ρ , $\text{atanh } \rho = \ln\{(1 + \rho)/(1 - \rho)\}/2$.

Our simulations demonstrate good performance of the selection- t ML estimators (SLt) in a finite sample under the correct specification of the model and error distribution. Biases and mean squared errors of the parameter estimates are close to zero and standard error estimates obtained using the inverse of the negative Hessian matrix, presented in the Appendix A2, adequately reflect variability in the parameter estimates.

Compared to other methods, SLt leads to smaller biases, in general, and smaller mean squared errors of the parameter estimates. In the presence of the exclusion restriction (Table 1), the results for the primary regression coefficients are comparable across the three methods, with SLt being slightly more efficient. In the absence of the exclusion restriction (Table 2), the SLN and TS methods demonstrate some bias in the estimates of the primary regression coefficients as the correlation between

Table 1. Simulation results in the presence of the exclusion restriction. Standard error estimates of $\operatorname{atanh} \rho$ and $\ln(\sigma)$ are not available for the TS method so the mean squared errors are not reported.

	Bias			MSE			SE diff.
	SLt	SLN	TS	SLt	SLN	TS	SLt
$\rho = 0$							
β_1	0.0024	0.0014	0.0012	0.0050	0.0063	0.0064	0.0001
β_0	0.0005	0.0025	0.0028	0.0073	0.0098	0.0104	0.0006
γ_2	0.0067	-0.1752	-0.1753	0.0179	0.0404	0.0404	0.0021
γ_1	0.0104	-0.2619	-0.2623	0.0305	0.0834	0.0836	0.0050
γ_0	0.0046	-0.1737	-0.1738	0.0161	0.0385	0.0385	0.0008
$\operatorname{atanh} \rho$	0.0035	0.0021	0.0014	0.0298	0.0293	–	0.0000
$\ln(\sigma)$	-0.0014	0.2366	0.2369	0.0041	0.0591	–	0.0005
$\ln(\nu)$	0.0306	–	–	0.0819	–	–	0.0051
$\rho = 0.2$							
β_1	0.0004	-0.0007	0.0012	0.0052	0.0065	0.0065	0.0016
β_0	-0.0039	-0.0004	-0.0046	0.0073	0.0098	0.0103	0.0002
γ_2	0.0059	-0.1737	-0.1739	0.0178	0.0397	0.0398	0.0023
γ_1	0.0033	-0.2651	-0.2655	0.0282	0.0837	0.0839	0.0022
γ_0	0.0044	-0.1726	-0.1726	0.0159	0.0379	0.0379	0.0000
$\operatorname{atanh} \rho$	0.0083	-0.0059	0.0051	0.0300	0.0315	–	0.0022
$\ln(\sigma)$	0.0023	0.2367	0.2372	0.0040	0.0592	–	0.0005
$\ln(\nu)$	0.0394	–	–	0.0809	–	–	0.0018
$\rho = 0.5$							
β_1	0.0009	0.0041	0.0048	0.0047	0.0062	0.0062	0.0003
β_0	0.0024	0.0023	0.0008	0.0065	0.0091	0.0105	0.0023
γ_2	0.0038	-0.1777	-0.1766	0.0171	0.0411	0.0409	0.0016
γ_1	0.0079	-0.2684	-0.2646	0.0288	0.0869	0.0848	0.0025
γ_0	0.0048	-0.1747	-0.1731	0.0152	0.0384	0.0380	0.0029
$\operatorname{atanh} \rho$	0.0049	-0.0057	0.0020	0.0332	0.0417	–	0.0076
$\ln(\sigma)$	-0.0015	0.2361	0.2366	0.0042	0.0590	–	0.0000
$\ln(\nu)$	0.0285	–	–	0.0857	–	–	0.0122
$\rho = 0.7$							
β_1	-0.0010	0.0075	0.0040	0.0045	0.0058	0.0061	0.0007
β_0	-0.0009	-0.0090	-0.0014	0.0053	0.0071	0.0097	0.0001
γ_2	0.0032	-0.1831	-0.1763	0.0158	0.0423	0.0402	0.0009
γ_1	0.0004	-0.2832	-0.2686	0.0274	0.0941	0.0860	0.0037
γ_0	-0.0005	-0.1829	-0.1767	0.0157	0.0414	0.0395	0.0024
$\operatorname{atanh} \rho$	0.0125	0.0214	0.0074	0.0354	0.0450	–	0.0043
$\ln(\sigma)$	0.0021	0.2378	0.2367	0.0044	0.0601	–	0.0010
$\ln(\nu)$	0.0351	–	–	0.0840	–	–	0.0032

Table 2. Simulation results in the absence of the exclusion restriction. Standard error estimates of $\operatorname{atanh} \rho$ and $\ln(\sigma)$ are not available for the TS method so the mean squared errors are not reported.

	Bias			MSE			SE diff.
	SLt	SLN	TS	SLt	SLN	TS	SLt
$\rho = 0$							
β_1	0.0013	-0.0009	-0.0029	0.0141	0.0253	0.0467	0.0023
β_0	-0.0005	0.0046	0.0082	0.0228	0.0434	0.0906	0.0032
γ_1	0.0021	-0.1573	-0.1565	0.0149	0.0326	0.0323	0.0049
γ_0	-0.0001	-0.1430	-0.1419	0.0113	0.0262	0.0258	0.0011
$\operatorname{atanh} \rho$	0.0009	-0.0089	-0.0109	0.1074	0.1720	–	0.0078
$\ln(\sigma)$	0.0049	0.2408	0.2532	0.0041	0.0615	–	0.0002
$\ln(\nu)$	0.0197	–	–	0.0796	–	–	0.0022
$\rho = 0.2$							
β_1	-0.0041	-0.0056	0.0072	0.0136	0.0264	0.0451	0.0033
β_0	0.0074	0.0134	-0.0055	0.0209	0.0448	0.0856	0.0031
γ_1	0.0088	-0.1521	-0.1498	0.0142	0.0310	0.0301	0.0002
γ_0	0.0051	-0.1392	-0.1371	0.0112	0.0254	0.0247	0.0003
$\operatorname{atanh} \rho$	-0.0108	-0.0213	0.0350	0.1026	0.1805	–	0.0096
$\ln(\sigma)$	0.0045	0.2390	0.2496	0.0042	0.0614	–	0.0004
$\ln(\nu)$	0.0289	–	–	0.0817	–	–	0.0042
$\rho = 0.5$							
β_1	-0.0075	0.0178	0.0131	0.0102	0.0210	0.0376	0.0045
β_0	0.0074	-0.0137	-0.0094	0.0139	0.0325	0.0701	0.0058
γ_1	0.0023	-0.1626	-0.1525	0.0143	0.0345	0.0312	0.0038
γ_0	0.0058	-0.1417	-0.1342	0.0116	0.0262	0.0240	0.0027
$\operatorname{atanh} \rho$	-0.0103	0.0466	0.0900	0.0809	0.1686	–	0.0149
$\ln(\sigma)$	0.0027	0.2400	0.2431	0.0051	0.0631	–	0.0030
$\ln(\nu)$	0.0480	–	–	0.0910	–	–	0.0153
$\rho = 0.7$							
β_1	-0.0026	0.0342	0.0038	0.0065	0.0133	0.0335	0.0025
β_0	0.0044	-0.0243	0.0141	0.0077	0.0174	0.0613	0.0017
γ_1	0.0008	-0.1693	-0.1485	0.0126	0.0363	0.0297	0.0032
γ_0	-0.0012	-0.1543	-0.1378	0.0112	0.0299	0.0250	0.0013
$\operatorname{atanh} \rho$	-0.0078	0.0800	0.0586	0.0577	0.1213	–	0.0043
$\ln(\sigma)$	0.0003	0.2376	0.2279	0.0052	0.0615	–	0.0015
$\ln(\nu)$	0.0520	–	–	0.0924	–	–	0.0118

errors increases, whereas the SLt estimates maintain low biases and mean squared errors. In both cases, the estimates of the selection regression coefficients and the scale parameter $\ln(\sigma)$ are severely biased under SLN and TS.

We also repeated the above simulation scenarios (not shown here) for $\nu = 3$, $\nu = 100$, and $\nu = \infty$, corresponding to the Normally distributed errors. Our findings for $\nu = 3$ were similar to the above with biases of the primary regression coefficients for SLN and TS being even more prominent in the absence of the exclusion restriction. As the degrees of freedom increases, the Student's t distribution approaches the normal distribution and, as expected, the results from the selection- t model were similar to those from the selection-normal model with the latter being slightly more efficient.

2. Test of selection bias when the error distribution is bivariate t

In this subsection we investigate the performance of three tests, commonly used for testing the presence of sample-selection bias in the OLS regression, when the error distribution is bivariate t . The tests under consideration are the selection-normal Wald test of $H_0: \rho = 0$ (or, equivalently, of $H_0: \operatorname{atanh}\rho = 0$), SLN, the likelihood ratio test of independent equations under the selection-normal model (LRT), and the Wald test of $H_0: \beta_\lambda = \rho\sigma = 0$ under the two-step estimation (TS). We also compare the performance of these tests to the selection- t Wald test of $H_0: \rho = 0$ (SLT) obtained under the correct, selection- t , model specification.

The data are simulated as described in the previous subsection. We consider the scenario in the presence of an exclusion restriction with varying sample sizes ($N = 500, 1000$), varying degrees of freedom ($\nu = 3, 5, 100$), and varying values of ρ . The simulation results are based on $R = 5000$ replications. We considered two nominal levels $\alpha = 0.01$ and $\alpha = 0.05$ and obtained similar findings. We present results for the nominal level $\alpha = 0.01$.

Under the null hypothesis, in the case of uncorrelated errors ($\rho = 0$), only the selection- t Wald test maintains correct nominal levels for small degrees of freedom; see Table 3. The other tests are known to be sensitive to the normality assumption and, thus, they often attribute nonnormality of errors to the presence of selection bias ($\rho \neq 0$) which leads to an inflated type I error. As degrees of freedom increase to 100, the significance levels of all tests are similar and close to the nominal level. As such, we compare powers of all tests only for $\nu = 100$ (Table 4) and report powers of the SLt test for other degrees of freedom (Table 5). The powers of all tests are similar for $\nu = 100$ and increase with sample size N and correlation ρ . The SLN test has slightly higher powers than the SLt test which is expected with large degrees of freedom. The TS test has lowest powers among the considered tests. From Table 5, the powers of the SLt test tend to be slightly higher for $\nu = 3$ than for $\nu = 5$.

Table 3. Empirical significance levels (as %) of the tests of selection bias for the nominal significance level $\alpha = 0.01$. Standard errors ranged between .13 (SLt) and .33 (SLN).

ν	$N = 500$				$N = 1000$			
	SLt	SLN	TS	LRT	SLt	SLN	TS	LRT
3	1.1	5.8	4.7	4.5	1	4.8	4.4	3.9
5	1.2	2.5	2.8	2	0.9	2	2.3	1.8
100	1.2	1.3	1.3	1.2	1	1	1.1	1

Table 4. Powers (as %) of the tests of selection bias for $\nu = 100$ for the nominal significance level $\alpha = 0.01$. Standard errors ranged between .21 and .7.

ρ	$N = 500$				$N = 1000$			
	SLt	SLN	TS	LRT	SLt	SLN	TS	LRT
.1	2.4	2.5	2.4	2.4	4.6	4.7	4.8	4.7
.2	8.7	8.9	8.5	8.8	19.3	19.8	19.6	19.5
.3	22.9	23.4	22.5	23.2	51.2	52.0	51.1	51.8
.4	45.3	46.3	43.5	46.0	81.2	81.5	80.0	81.4
.5	72.2	72.9	68.4	72.7	95.2	95.2	94.4	95.2
.6	89.4	89.4	86.0	89.4	97.0	97.0	96.8	97.0
.7	95.1	95.2	93.9	95.1	97.5	97.5	97.5	97.5

Table 5. Powers (as %) of the SLt test of $\rho = 0$ for $\nu = 3$ and $\nu = 5$ for the nominal significance level $\alpha = 0.01$. Standard errors ranged between .04 and .7.

ρ	$N = 500$		$N = 1000$	
	$\nu = 3$	$\nu = 5$	$\nu = 3$	$\nu = 5$
.1	2.5	2.6	4.9	4.2
.2	8.7	7.8	19.6	18.1
.3	22.3	21.8	50.6	48.4
.4	47.0	45.1	83.1	80.8
.5	73.4	73.0	97.5	96.7
.6	91.4	91.4	99.9	99.8
.7	97.3	96.8	99.8	99.9

C. Numerical Application to Ambulatory Expenditures

We consider the data on ambulatory expenditures from the 2001 Medical Expenditure Panel Survey analyzed by Cameron and Trivedi (2010). The data consist of 3,328 observations, of which 526 (15.8%) correspond to zero values of expenditures. The dataset includes several explanatory variables such as age, gender, education status, and others. The distribution of expenditures is highly skewed so the analysis is performed using the log scale. Because the decision to spend is likely to be related to the spending amount, it is natural to consider a sample-selection model for the analysis. Cameron and Trivedi (2010, p. 561), among other models, use the classical Heckman sample-selection model to analyze these data; see Appendix A4 for the results from Stata's `heckman` command (StataCorp 2009). The primary regression includes such factors as age, gender, ethnicity, education status, insurance status, and the number of chronic diseases. The selection equation also includes income imposing the exclusion restriction on the model, although the use of income for this purpose is debatable. All of the considered factors are strong predictors of the decision to spend. All factors other than the insurance status are also strong predictors of the spending amount. The reported Wald test ($p = 0.380$) of no sample selection, $H_0: \rho = 0$ or, more precisely, $H_0: \operatorname{atanh}\rho = 0$, does not provide sufficient evidence to reject this hypothesis, implying that spending amount is unrelated to the decision to spend and can be analyzed separately using standard OLS regression. This conclusion seems implausible.

As noted by Cameron and Trivedi, the assumption of underlying normality is suspect for these data. Thus, we use the proposed selection- t model to analyze these data; see the results from Stata's user-written command `heckt` in Appendix A4. (This command is available from the authors on request.)

We obtain results similar to those from the selection-normal model regarding the coefficients in the primary and selection equations. There is a significant change, however, in the inference about the existence of sample-selection bias. The Wald test provides sufficient evidence ($p = 0.009$) to reject the null hypothesis of no sample-selection bias at a 1% significance level, in agreement with our intuition. The estimate of ρ reported by the selection-normal model is -0.131 whereas the selection- t model reports an estimate of -0.322 , which is also more in agreement with the two-step estimate of ρ , -0.359 . Our simulations showed that the estimate of ρ from the selection-normal model can be biased when data come from a Student's t distribution which is likely what we observe in this example. The estimated degrees of freedom are 13 with the 95% confidence interval of $(8, 20)$, indicating some deviation from normality.

D. Conclusion

We introduced the sample selection- t model which extends the conventional sample-selection model of Heckman (1974) to have a bivariate Student's t error distribution. This model provides a greater flexibility for modeling heavier-tailed data than the selection-normal model by introducing only one extra parameter, the degrees of freedom, controlling the tails of the distribution. We considered maximum likelihood estimation of the parameters. Monte Carlo simulations demonstrated good performance of the MLEs in finite samples. Monte Carlo simulations also showed that the selection- t model performs better than the selection-normal model for heavier-tailed data and is also more robust to collinearity between the primary and selection regressors for moderate degrees of freedom. The robustness to the collinearity is appealing because there is no need to impose exclusion restrictions which are often difficult to

formulate in practice. Our simulations demonstrated high sensitivity (inflated type I error) of the selection bias test based on the selection-normal model to the presence of heavy tails, whereas the selection bias test based on the selection- t maintained nominal coverage. We also provided some insight into the power of this test.

Although the considered parametric selection- t model is not as flexible in modeling various shapes of the distribution compared with semiparametric and nonparametric methods, it is useful to model heavy-tailed data which occur rather often in practice. Its advantages include a gain of efficiency within a class of heavy-tailed Student's t distributions; an ability to identify an intercept, which, as was noted by Heckman, is an important parameter of interest in many economic applications; and also the relative simplicity and speed efficiency of the implementation.

In this paper we used a bivariate Student's t distribution to allow for heavier tails in the error distribution. It is straightforward to extend the presented results to some other parametric distributions. For example, one can use another version of a bivariate t distribution where a separate degrees-of-freedom parameter is considered for each dimension. From a practical standpoint, it would be even more appealing to consider some flexible parametric distributions accommodating the two most common deviations from normality — the heavier tails and the asymmetry of the distribution. A bivariate skew-normal distribution (Azzalini and Dalla Valle 1996, Azzalini and Capitanio 1999, Azzalini 2005) and a bivariate skew- t distribution (Azzalini and Capitanio 2003) are two appealing candidates. Keeping in mind the link described in Subsection 2 of Section A of this chapter, we can justify the use of the skewed distributions to model errors as a natural way of modeling some hidden truncation already present in the population from which the data were sampled. For example, Little and Rubin (2002, p. 324) give an example of such a population.

CHAPTER III

MULTIVARIATE LOG-SKEW-ELLIPTICAL DISTRIBUTIONS WITH
APPLICATIONS TO PRECIPITATION DATA*

In recent years, there has been a growing interest for more flexible parametric families of non-normal distributions with additional parameters allowing to regulate skewness and tails. This is especially important with environmental data which are often skewed and heavy tailed. The simplest representative of such families, as defined by Azzalini (1985), is the so-called skew-normal distribution. It extends the conventional normal model by introducing an additional parameter controlling the asymmetry of the distribution, the shape parameter. Azzalini and Dalla Valle (1996) proposed a multivariate analog of the univariate skew-normal distribution. Branco and Dey (2001) and Azzalini and Capitanio (2003) introduced the univariate and multivariate skew- t distributions, which extend the respective skew-normal distributions by allowing to control the tails of the distribution with the additional degrees-of-freedom parameter. A more detailed description of these and other skewed models may be found in the book edited by Genton (2004) and in the review by Azzalini (2005).

The support of the univariate skew-normal, skew- t , and, more generally, skew-elliptical distributions is the real line. For data that cannot be negative, such as income or precipitation, distributions with positive support, such as gamma, exponential, and log-normal, are often used for modeling purpose. The problem of modeling is exacerbated in the multivariate setting, where tractable distributions besides the multivariate log-normal are lacking. We expand the list of such distributions by

* Reprinted with permission from “Multivariate log-skew-elliptical distributions with applications to precipitation data” by Y. V. Marchenko and M. G. Genton, 2010, *Environmetrics*, 21, 318–340. Copyright [2010] by John Wiley & Sons.

introducing a family of multivariate log-skew-elliptical distributions.

Azzalini et al. (2003) introduced the univariate log-skew-normal and log-skew- t distributions, although without formal investigation of their properties and characteristics, and use them to model family income data. We extend their definition more generally to the class of multivariate log-skew-elliptical distributions. We also examine probabilistic properties of the multivariate log-skew-elliptical distributions, such as stochastic representations, marginal and conditional distributions, and existence of moments.

In climatology, various distributions have been used to model the distribution of precipitation data; among them there are the exponential, gamma (e.g., Wilks 2006, p. 98), and log-normal (e.g., Crow and Shimizu 1988). There is no definitive physical justification to what distribution has to be used to model precipitation. The distribution of precipitation data is often skewed and sometimes has heavy tails. For example, Wilson and Toumi (2005) study univariate heavy precipitation and point out that its distribution exhibits a heavy-tailed behavior. These characteristics motivate us to consider the log-skew-normal and log-skew- t distributions to model precipitation data. One appealing feature of these distributions is that their extension to the multivariate case is straightforward. Moreover, model parameters can be estimated using readily available estimation methods developed for the multivariate skew-normal and the multivariate skew- t distributions (Azzalini and Capitanio 1999, 2003). We define these and, more generally, the multivariate skew-elliptical distributions next.

Let $EC_d(\boldsymbol{\xi}, \boldsymbol{\Omega}, g^{(d)})$ denote a family of d -dimensional elliptically contoured distributions (with existing probability density function) with a generator function $g^{(d)}(u)$, $u \geq 0$, defining a spherical d -dimensional density, a location column vector $\boldsymbol{\xi} \in \mathbb{R}^d$, and a $d \times d$ positive definite dispersion matrix $\boldsymbol{\Omega}$. If $\mathbf{X} \sim EC_d(\boldsymbol{\xi}, \boldsymbol{\Omega}, g^{(d)})$, then its density is $f_d(\mathbf{x}; \boldsymbol{\xi}, \boldsymbol{\Omega}, g^{(d)}) = |\boldsymbol{\Omega}|^{-1/2} g^{(d)}(Q_{\mathbf{x}}^{\boldsymbol{\xi}, \boldsymbol{\Omega}})$, where $Q_{\mathbf{x}}^{\boldsymbol{\xi}, \boldsymbol{\Omega}} = (\mathbf{x} - \boldsymbol{\xi})^\top \boldsymbol{\Omega}^{-1} (\mathbf{x} - \boldsymbol{\xi})$ and

$\mathbf{x} \in \mathbb{R}^d$ (Fang et al. 1990, p. 46).

Genton (2004) summarizes various approaches for defining skew-elliptical distributions. We consider a class of skew-elliptical distributions with a density of the form

$$f_{\text{SE}_d}(\mathbf{x}; \Theta) = 2 f_d(\mathbf{x}; \boldsymbol{\xi}, \boldsymbol{\Omega}, g^{(d)}) F\{\boldsymbol{\alpha}^\top \boldsymbol{\omega}^{-1}(\mathbf{x} - \boldsymbol{\xi}); g_{Q_{\mathbf{x}}^{\boldsymbol{\xi}, \boldsymbol{\Omega}}}\}, \quad \mathbf{x} \in \mathbb{R}^d, \quad (3.1)$$

where $\Theta = (\boldsymbol{\xi}, \boldsymbol{\Omega}, \boldsymbol{\alpha}, g^{(d)})$, $\boldsymbol{\alpha} \in \mathbb{R}^d$ is a shape parameter, $\boldsymbol{\omega} = \text{diag}(\boldsymbol{\Omega})^{1/2}$ is a $d \times d$ scale matrix, $f_d(\mathbf{x}; \boldsymbol{\xi}, \boldsymbol{\Omega}, g^{(d)})$ is the density of $\text{EC}_d(\boldsymbol{\xi}, \boldsymbol{\Omega}, g^{(d)})$ defined above, and $F(u; g_{Q_{\mathbf{x}}^{\boldsymbol{\xi}, \boldsymbol{\Omega}}})$ is the cumulative distribution function of $\text{EC}_1(0, 1, g_{Q_{\mathbf{x}}^{\boldsymbol{\xi}, \boldsymbol{\Omega}}})$ with the generator function $g_{Q_{\mathbf{x}}^{\boldsymbol{\xi}, \boldsymbol{\Omega}}}(u) = g^{(d+1)}(u + Q_{\mathbf{x}}^{\boldsymbol{\xi}, \boldsymbol{\Omega}})/g^{(d)}(Q_{\mathbf{x}}^{\boldsymbol{\xi}, \boldsymbol{\Omega}})$. Although $\boldsymbol{\alpha}$ is referred to as a shape parameter, the shape of the distribution (3.1) is regulated in a more complex way. In what follows we use the notation $\text{SE}_d(\boldsymbol{\xi}, \boldsymbol{\Omega}, \boldsymbol{\alpha}, g^{(d+1)})$ to refer to a family of skew-elliptical distributions with density (3.1).

We also consider two special cases of multivariate skew-elliptical distributions: the skew-normal and the skew- t . The density of the multivariate skew-normal distribution is

$$f_{\text{SN}_d}(\mathbf{x}; \Theta) = 2 \phi_d(\mathbf{x}; \boldsymbol{\xi}, \boldsymbol{\Omega}) \Phi\{\boldsymbol{\alpha}^\top \boldsymbol{\omega}^{-1}(\mathbf{x} - \boldsymbol{\xi})\}, \quad \mathbf{x} \in \mathbb{R}^d, \quad (3.2)$$

where $\Theta = (\boldsymbol{\xi}, \boldsymbol{\Omega}, \boldsymbol{\alpha})$, $\phi_d(\mathbf{x}; \boldsymbol{\xi}, \boldsymbol{\Omega})$ is the density of a d -variate normal distribution with location $\boldsymbol{\xi}$ and dispersion matrix $\boldsymbol{\Omega}$, and $\Phi(\cdot)$ is the cumulative distribution function of the standard normal distribution. We denote a multivariate skew-normal random vector with density (3.2) as $\mathbf{X} \sim \text{SN}_d(\boldsymbol{\xi}, \boldsymbol{\Omega}, \boldsymbol{\alpha})$. For more details and applications of the multivariate skew-normal distribution, see Azzalini and Dalla Valle (1996), Azzalini and Capitanio (1999), Capitanio et al. (2003), and Azzalini (2005).

The density of the multivariate skew- t distribution is

$$f_{\text{ST}_d}(\mathbf{x}; \Theta) = 2 t_d(\mathbf{x}; \boldsymbol{\xi}, \boldsymbol{\Omega}, \nu) T \left\{ \boldsymbol{\alpha}^\top \boldsymbol{\omega}^{-1}(\mathbf{x} - \boldsymbol{\xi}) \left(\frac{\nu + d}{\nu + Q_{\mathbf{x}}^{\boldsymbol{\xi}, \boldsymbol{\Omega}}} \right)^{1/2}; \nu + d \right\}, \quad \mathbf{x} \in \mathbb{R}^d, \quad (3.3)$$

where $\Theta = (\boldsymbol{\xi}, \boldsymbol{\Omega}, \boldsymbol{\alpha}, \nu)$, $T(\cdot; \nu + d)$ is the cumulative distribution function of a univariate Student's t distribution with $\nu + d$ degrees of freedom, and $t_d(\mathbf{x}; \boldsymbol{\xi}, \boldsymbol{\Omega}, \nu) = \Gamma\{(\nu + d)/2\}(1 + Q_{\boldsymbol{\xi}, \boldsymbol{\Omega}}^{\mathbf{x}}/\nu)^{-(\nu+d)/2}/\{|\boldsymbol{\Omega}|^{1/2}(\nu\pi)^{d/2}\Gamma(\nu/2)\}$ is the density of a d -variate Student's t distribution with ν degrees of freedom. We denote a multivariate skew- t random vector with density (3.3) as $\mathbf{X} \sim \text{ST}_d(\boldsymbol{\xi}, \boldsymbol{\Omega}, \boldsymbol{\alpha}, \nu)$. For more details and applications of the multivariate skew- t distribution see, for example, Azzalini and Capitanio (2003) and Azzalini and Genton (2008), among others.

The chapter is organized as follows. The multivariate log-skew-elliptical distributions and their properties are defined and studied in Section A, with emphasis on two special cases: the log-skew-normal and log-skew- t distributions. The relevant proofs of the results are given in the Appendix B. Numerical applications of the log-skew-normal and log-skew- t distributions to U.S. monthly precipitation data are presented in Section B. The chapter ends with a discussion in Section C.

A. Multivariate Log-Skew-Elliptical Distributions

In this section we provide a formal definition of a family of multivariate log-skew-elliptical distributions and present their probabilistic properties including stochastic representations, conditional and marginal distributions, and moments, as well as inferential properties.

1. Definitions

Let $\mathbf{ln}(\mathbf{X}) = \{\ln(X_1), \dots, \ln(X_d)\}^\top$, $X_i > 0$, $i = 1, \dots, d$ be the component-wise logarithm of the positive random vector $\mathbf{X} = (X_1, \dots, X_d)^\top$ and $\mathbf{exp}(\mathbf{Y}) = \{\exp(Y_1), \dots, \exp(Y_d)\}^\top$ be the component-wise exponential of the random vector $\mathbf{Y} = (Y_1, \dots, Y_d)^\top$.

Definition 1. A positive random vector \mathbf{X} has a multivariate log-skew-elliptical distribution, denoted as $\mathbf{X} \sim \text{LSE}_d(\boldsymbol{\xi}, \boldsymbol{\Omega}, \boldsymbol{\alpha}, g^{(d+1)})$, if $\ln(\mathbf{X})$ is a multivariate skew-elliptical random vector, $\ln(\mathbf{X}) \sim \text{SE}_d(\boldsymbol{\xi}, \boldsymbol{\Omega}, \boldsymbol{\alpha}, g^{(d+1)})$, with density (3.1). Likewise, if $\mathbf{X} \sim \text{SE}_d(\boldsymbol{\xi}, \boldsymbol{\Omega}, \boldsymbol{\alpha}, g^{(d+1)})$, then $\exp(\mathbf{X}) \sim \text{LSE}_d(\boldsymbol{\xi}, \boldsymbol{\Omega}, \boldsymbol{\alpha}, g^{(d+1)})$.

If the multivariate log-skew-elliptical density exists, it is of the form

$$f_{\text{LSE}_d}(\mathbf{x}; \Theta) = 2 \left(\prod_{i=1}^d x_i^{-1} \right) f_d\{\ln(\mathbf{x}); \boldsymbol{\xi}, \boldsymbol{\Omega}, g^{(d)}\} F[\boldsymbol{\alpha}^\top \boldsymbol{\omega}^{-1} \{\ln(\mathbf{x}) - \boldsymbol{\xi}\}; g_{Q_{\ln(\mathbf{x})}^{\boldsymbol{\xi}, \boldsymbol{\Omega}}}], \quad \mathbf{x} > \mathbf{0}, \quad (3.4)$$

where $\Theta = (\boldsymbol{\xi}, \boldsymbol{\Omega}, \boldsymbol{\alpha}, g^{(d)})$, and all other terms are as defined in (3.1). Here, the term $\left(\prod_{i=1}^d x_i^{-1} \right)$ is the Jacobian associated with the transformation $\ln(\mathbf{x}) \rightarrow \mathbf{x}$. Notice that when $\boldsymbol{\alpha} = \mathbf{0}$, the multivariate log-skew-elliptical density (3.4) reduces to a multivariate log-elliptical density as defined in Fang et al. (1990, p. 56).

Clearly, the interpretation of the parameters in Θ is not the same for \mathbf{X} as for $\ln(\mathbf{X})$. For example, $\boldsymbol{\xi}$, $\boldsymbol{\Omega}$, and $\boldsymbol{\alpha}$ do not regulate strictly location, scale, and skewness, respectively, on the original scale compared with the log scale. From the definition, $\mathbf{X} = \exp\{\ln(\mathbf{X})\} = \exp\{\boldsymbol{\xi} + \boldsymbol{\omega} \ln(\mathbf{Z})\} = \text{diag}\{\exp(\boldsymbol{\xi})\} \exp\{\boldsymbol{\omega} \ln(\mathbf{Z})\}$, where $\ln(\mathbf{Z}) \sim \text{SE}_d(\mathbf{0}, \bar{\boldsymbol{\Omega}}, \boldsymbol{\alpha}, g^{(d+1)})$ and $\bar{\boldsymbol{\Omega}} = \boldsymbol{\omega}^{-1} \boldsymbol{\Omega} \boldsymbol{\omega}^{-1}$ is the correlation matrix. We can see that $\boldsymbol{\xi}$ affects the scale of the distribution of \mathbf{X} , and $\boldsymbol{\omega}$ (and more generally $\boldsymbol{\Omega}$) together with the ‘‘shape’’ parameter $\boldsymbol{\alpha}$ regulate the shape of the distribution. We investigate how these parameters affect the shape of the distribution in more detail using the log-skew-normal and log-skew- t distributions which we define next.

From Definition 1 and from (3.2) and (3.3), the density of a multivariate log-skew-normal distribution, denoted as $\text{LSN}_d(\boldsymbol{\xi}, \boldsymbol{\Omega}, \boldsymbol{\alpha})$, is

$$f_{\text{LSN}_d}(\mathbf{x}; \Theta) = 2 \left(\prod_{i=1}^d x_i^{-1} \right) \phi_d\{\ln(\mathbf{x}); \boldsymbol{\xi}, \boldsymbol{\Omega}\} \Phi[\boldsymbol{\alpha}^\top \boldsymbol{\omega}^{-1} \{\ln(\mathbf{x}) - \boldsymbol{\xi}\}], \quad \mathbf{x} > \mathbf{0}, \quad (3.5)$$

and the density of a multivariate log-skew- t distribution, denoted as $\text{LST}_d(\boldsymbol{\xi}, \boldsymbol{\Omega}, \boldsymbol{\alpha}, \nu)$,

is for $\mathbf{x} > \mathbf{0}$:

$$f_{\text{LST}_d}(\mathbf{x}; \Theta) = \frac{2 t_d\{\ln(\mathbf{x}); \boldsymbol{\xi}, \boldsymbol{\Omega}, \nu\} T \left[\boldsymbol{\alpha}^\top \boldsymbol{\omega}^{-1} \{\ln(\mathbf{x}) - \boldsymbol{\xi}\} \left(\frac{\nu+d}{\nu+Q_{\boldsymbol{\xi}, \boldsymbol{\Omega}}^{\ln(\mathbf{x})}} \right)^{1/2}; \nu+d \right]}{\prod_{i=1}^d x_i}. \quad (3.6)$$

It is easily seen that the densities (3.5) and (3.6) reduce to the density of a multivariate log-normal distribution, when $\boldsymbol{\alpha} = \mathbf{0}$ and $\nu = \infty$. Also, the density (3.6) reduces to the density of a multivariate log- t distribution, when $\boldsymbol{\alpha} = \mathbf{0}$. For $d = 1$, the densities (3.5) and (3.6) correspond to the distributions in Azzalini et al. (2003), although they did not give the densities explicitly.

We illustrate the shapes of univariate log-skew-normal and log-skew- t distributions, and bivariate log-skew-normal distributions next.

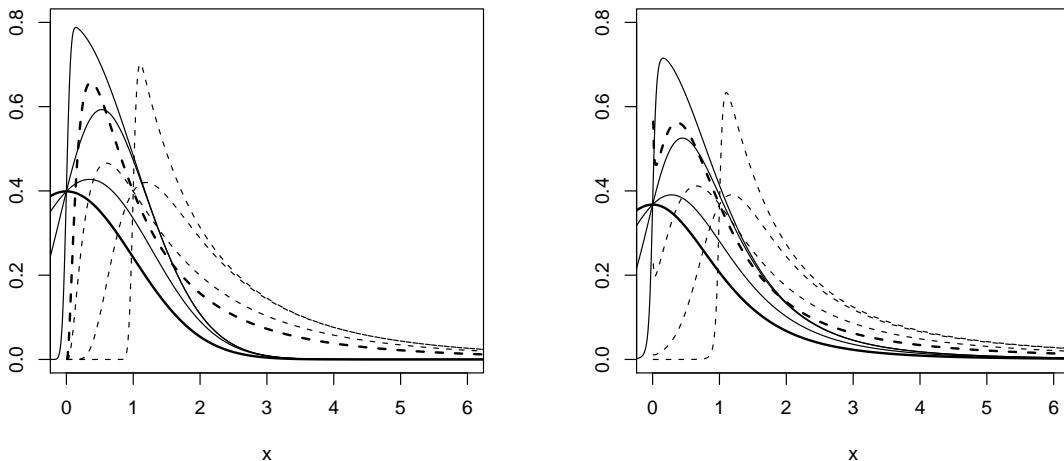


Figure 3. Left panel: skew-normal (solid curves) and log-skew-normal (dashed curves) densities with $\alpha = 0$ (thick curves) and $\alpha = 0.5, 2, 20$. Right panel: skew- t (solid curves) and log-skew- t (dashed curves) densities with $\alpha = 0$ (thick curves) and $\alpha = 0.5, 2, 20$ for fixed $\nu = 3$.

Figure 3 presents univariate log-skew-normal (left panel) and log-skew- t (right panel) densities for $\xi = 0$, $\omega = 1$, $\nu = 3$, and varying $\alpha = 0, 0.5, 2, 20$ (dashed curves). The respective skew-normal and skew- t densities are depicted for comparison as solid curves. The respective reference distributions are log-normal (log- t) and normal (Student's t) (thick curves). The additional parameter α allows the log-skew-normal and log-skew- t densities to have more flexible shapes than the reference log-normal and log- t distributions. The spike in the shape of the log- t density (right panel, thick dashed curve) is explained by the fact that this density has two stationary points for small values of ν . We observe a similar behavior for the log-skew- t density with $\alpha = 0.5$, but as in the case of the log- t distribution, it vanishes as α (or ν) increases.

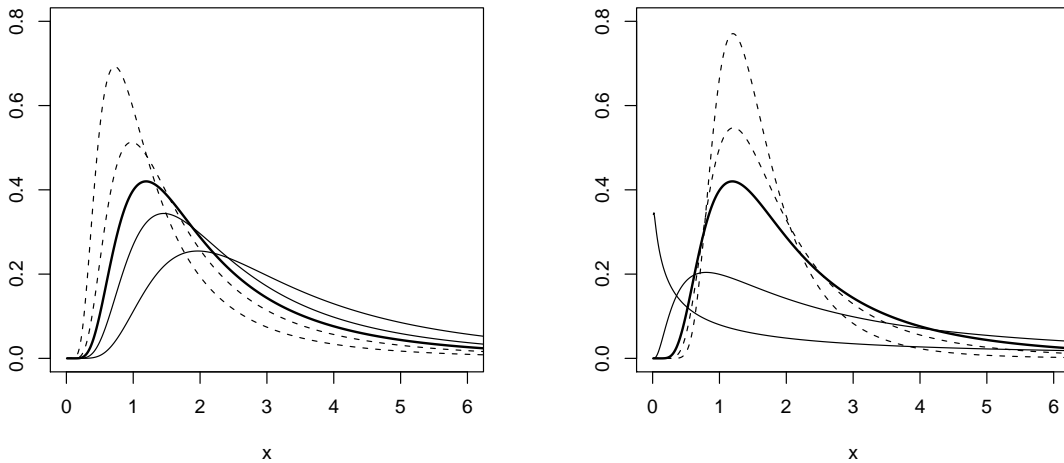


Figure 4. Log-skew-normal densities $LSN(\xi, \omega^2, 2)$ for varying ξ (left panel) and varying ω (right panel). Left panel: $\xi = 0$ (solid thick curve), $\xi = -0.2, -0.5$ (dashed curves), and $\xi = 0.2, 0.5$ (solid curves). Right panel: $\omega = 1$ (solid thick curve), $\omega = 0.6, 0.8$ (dashed curves), and $\omega = 2, 5$ (solid curves).

Figure 4 depicts how the shape of a log-skew-normal density changes as a function of ξ and ω . The location ξ affects the shape in a multiplicative fashion (left panel). For positive values of ξ the density is stretched compared with the reference density with $\xi = 0$. For negative values of ξ , the density contracts toward the mode. Varying values of ω (right panel) change the look of the distribution, especially for large values of ω .

Figure 5 depicts bivariate log-skew-normal densities for various values of $\boldsymbol{\alpha} = (\alpha_1, \alpha_2)^\top$. The shape $\boldsymbol{\alpha} = \mathbf{0}$ corresponds to the reference bivariate log-normal distribution. For negative values of α_1 and α_2 , the density is skewed to the left in each direction, whereas for positive values of α_1 and α_2 , it is more skewed to the right in each direction.

2. Stochastic representations

As in the case of the multivariate skew-elliptical distributions, there are other equivalent ways of defining the log-skew-elliptical distribution by using different stochastic representations. These representations may be used for simulation purposes. In this subsection we formulate three stochastic representations for log-skew-elliptical random vectors.

Proposition 1 (selection representation 1). Consider a $d+1$ -dimensional random vector $(\tilde{U}_0, \mathbf{U}^\top)^\top$ that follows a multivariate elliptical distribution $\text{EC}_{d+1}(\mathbf{0}, \bar{\boldsymbol{\Omega}}^*, g^{(d+1)})$ with

$$\bar{\boldsymbol{\Omega}}^* = \begin{pmatrix} 1 & \mathbf{0}^\top \\ \mathbf{0} & \bar{\boldsymbol{\Omega}} \end{pmatrix}.$$

Let $\mathbf{X} = \exp\{\boldsymbol{\xi} + \boldsymbol{\omega} \ln(\mathbf{Z})\}$, $\mathbf{V} = \exp(\mathbf{U})$ and $\mathbf{Z} \stackrel{d}{=} (\mathbf{V} | \tilde{U}_0 < \boldsymbol{\alpha}^\top \mathbf{U})$. Then $\mathbf{Z} \sim \text{LSE}_d(\mathbf{0}, \bar{\boldsymbol{\Omega}}, \boldsymbol{\alpha}, g^{(d+1)})$ and, so, $\mathbf{X} \sim \text{LSE}_d(\boldsymbol{\xi}, \boldsymbol{\Omega}, \boldsymbol{\alpha}, g^{(d+1)})$.

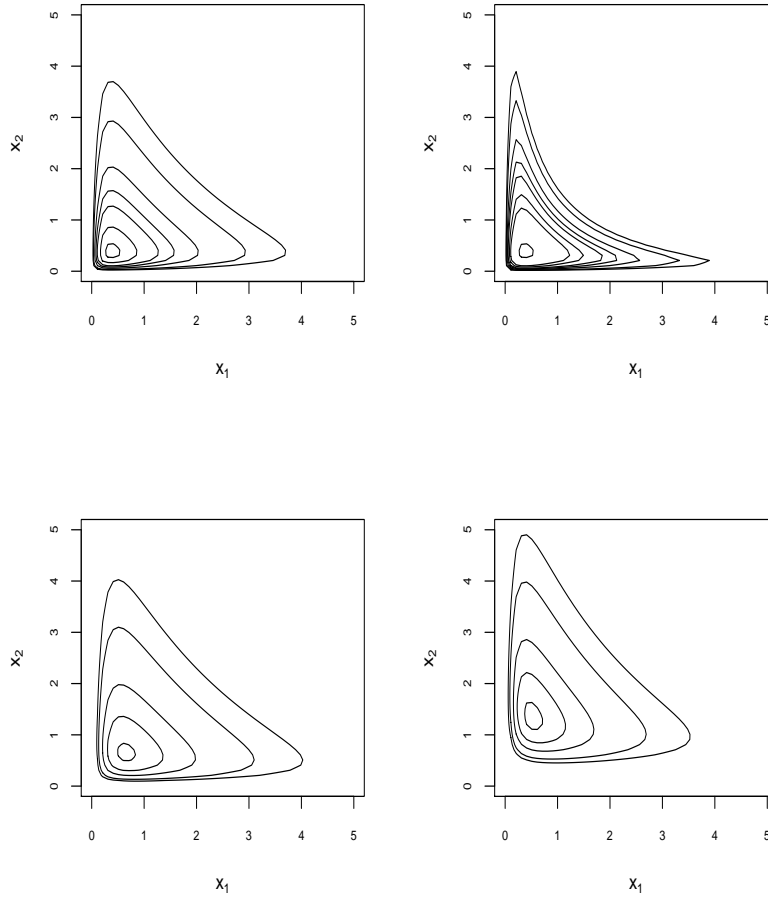


Figure 5. Contour plots of the standard bivariate log-skew-normal density plotted at levels 0.03, 0.05, 0.1, 0.15, 0.2, 0.3, 0.4, 0.8 for varying values of α . $\alpha = (0, 0)^\top$ (top left panel), $\alpha = (-2, -2)^\top$ (top right panel), $\alpha = (0.5, 0.5)^\top$ (bottom left panel), and $\alpha = (0.5, 2)^\top$ (bottom right panel).

Proposition 2 (selection representation 2). Consider a $d+1$ -dimensional random vector $(U_0, \mathbf{U}^\top)^\top$ that follows a multivariate elliptical distribution $\text{EC}_{d+1}(\mathbf{0}, \bar{\boldsymbol{\Omega}}^*, g^{(d+1)})$ with

$$\bar{\boldsymbol{\Omega}}^* = \begin{pmatrix} 1 & \boldsymbol{\delta}^\top \\ \boldsymbol{\delta} & \bar{\boldsymbol{\Omega}} \end{pmatrix},$$

and skewness parameter $\boldsymbol{\delta} = (\delta_1, \dots, \delta_d)^\top \in (-1, 1)^d$. Let $\mathbf{X} = \mathbf{exp}\{\boldsymbol{\xi} + \boldsymbol{\omega} \ln(\mathbf{Z})\}$, $\mathbf{V} = \mathbf{exp}(\mathbf{U})$ and $\mathbf{Z} \stackrel{d}{=} (\mathbf{V}|U_0 > 0)$. Then $\mathbf{Z} \sim \text{LSE}_d(\mathbf{0}, \bar{\boldsymbol{\Omega}}, \boldsymbol{\alpha}, g^{(d+1)})$ and, so, $\mathbf{X} \sim \text{LSE}_d(\boldsymbol{\xi}, \boldsymbol{\Omega}, \boldsymbol{\alpha}, g^{(d+1)})$ with $\boldsymbol{\alpha} = \bar{\boldsymbol{\Omega}}^{-1} \boldsymbol{\delta} / (1 - \boldsymbol{\delta}^\top \bar{\boldsymbol{\Omega}}^{-1} \boldsymbol{\delta})^{1/2}$.

Azzalini and Dalla Valle (1996) refer to the selection representation 2 as a *conditioning method* and apply it to define a multivariate skew-normal distribution. This stochastic representation corresponds to the so-called $\boldsymbol{\delta}$ -parameterization of the shape parameter $\boldsymbol{\alpha}$ from \mathbb{R}^d to $(-1, 1)^d$. The selection representation 1 may be obtained from the selection representation 2 by setting $U_0 = (1 + \boldsymbol{\alpha}^\top \bar{\boldsymbol{\Omega}} \boldsymbol{\alpha})^{-1/2} (\boldsymbol{\alpha}^\top \mathbf{U} - \tilde{U}_0)$.

Branco and Dey (2001) present a special subclass of the skew-elliptical distributions, the scale mixture of a skew-normal distribution, denoted as $\text{SMSN}_d\{\boldsymbol{\xi}, \boldsymbol{\Omega}, \boldsymbol{\alpha}, K(\eta), H(\eta)\}$, with a density of the form

$$f_{\text{SMSN}_d}(\mathbf{x}; \Theta) = 2 \int_0^\infty \phi_d\{\mathbf{x}; \boldsymbol{\xi}, K(\eta) \boldsymbol{\Omega}\} \Phi\{\boldsymbol{\alpha}^\top \boldsymbol{\omega}^{-1}(\mathbf{x} - \boldsymbol{\xi}) K^{-1/2}(\eta)\} dH(\eta), \quad \mathbf{x} \in \mathbb{R}^d, \quad (3.7)$$

where $\Theta = \{\boldsymbol{\xi}, \boldsymbol{\Omega}, \boldsymbol{\alpha}, K(\eta), H(\eta)\}$, η is a random variable (a so-called mixing variable) with cumulative distribution function $H(\eta)$ and $K(\eta)$ is a weight function. We denote $\mathbf{X} \sim \text{LSMSN}_d\{\boldsymbol{\xi}, \boldsymbol{\Omega}, \boldsymbol{\alpha}, K(\eta), H(\eta)\}$, if $\ln(\mathbf{X}) \sim \text{SMSN}_d\{\boldsymbol{\xi}, \boldsymbol{\Omega}, \boldsymbol{\alpha}, K(\eta), H(\eta)\}$ with density (3.7).

Proposition 3 (log-skew-normal mixture). Let $\mathbf{Z} \sim \text{LSN}_d(\mathbf{0}, \boldsymbol{\Omega}, \boldsymbol{\alpha})$. Suppose that η is a random variable with cumulative distribution function $H(\eta)$ and $K(\eta)$ is a weight function. If $\mathbf{X} = \mathbf{exp}(\boldsymbol{\xi}) \mathbf{Z}^{K^{1/2}(\eta)}$, then $\mathbf{X} \sim \text{LSMSN}_d\{\boldsymbol{\xi}, \boldsymbol{\Omega}, \boldsymbol{\alpha}, K(\eta), H(\eta)\}$.

For example, the multivariate log-skew-normal and log-skew- t distributions are special cases of the LSMSN distribution. The multivariate log-skew-normal distribution arises when $K(\eta) = 1$ and $H(\eta)$ is degenerate. The multivariate log-skew- t distribution arises when $K(\eta) = 1/\eta$ and η is distributed as $Gamma(\nu/2, \nu/2)$ with density $f(\eta) = (\nu/2)^{\nu/2} \exp(-\nu\eta/2)/\Gamma(\nu/2)$.

Other stochastic representations of the skew-elliptical random vectors can be used to generate the log-skew-elliptical random vectors. For example, similarly to Proposition 2 we can formulate the convolution-type stochastic representation (Arellano-Valle and Genton 2010a) for the log-skew-elliptical random vector. Also, the skew-elliptical random variates can be viewed as linear combinations of order statistics of elliptical exchangeable random variates (Arellano-Valle and Genton 2007). Coupled with Definition 1, this presents yet another stochastic representation of the log-skew-elliptical random variates.

3. Marginal and conditional distributions

From Definition 1 it can be inferred that all marginal distributions of \mathbf{X} are univariate log-skew-elliptical. Since $\ln(\mathbf{X})$ is multivariate skew-elliptical, then each component $\ln(X_i)$ is univariate skew-elliptical (e.g. Branco and Dey 2001). Therefore, by definition X_i is univariate log-skew-elliptical.

Proposition 4 (marginal distribution). Let $\mathbf{X} \sim \text{LSE}_d(\boldsymbol{\xi}, \boldsymbol{\Omega}, \boldsymbol{\alpha}, g^{(d+1)})$. Consider the following partition of $\mathbf{X}^\top = (\mathbf{X}_1^\top, \mathbf{X}_2^\top)$, $\boldsymbol{\xi}^\top = (\boldsymbol{\xi}_1^\top, \boldsymbol{\xi}_2^\top)$, and $\boldsymbol{\alpha}^\top = (\boldsymbol{\alpha}_1^\top, \boldsymbol{\alpha}_2^\top)$ into q and $d - q$ components, respectively. Let $\boldsymbol{\Omega} = (\omega_{ij})_{i,j=1}^d$ have the following partition:

$$\begin{pmatrix} \boldsymbol{\Omega}_{11} & \boldsymbol{\Omega}_{12} \\ \boldsymbol{\Omega}_{21} & \boldsymbol{\Omega}_{22} \end{pmatrix},$$

with $\boldsymbol{\Omega}_{21} = \boldsymbol{\Omega}_{12}^\top$. Define $\boldsymbol{\omega}_1 = \text{diag}(w_{11}^{1/2}, \dots, w_{qq}^{1/2})$ of dimension $q \times q$ and $\boldsymbol{\omega}_2 = \text{diag}(w_{q+1,q+1}^{1/2}, \dots, w_{dd}^{1/2})$ of dimension $(d-q) \times (d-q)$. Then the marginal distributions of \mathbf{X}_1 and \mathbf{X}_2 are $\text{LSE}_q(\boldsymbol{\xi}_1, \boldsymbol{\Omega}_{11}, \boldsymbol{\alpha}_1^*, g^{(d+1)})$ and $\text{LSE}_{d-q}(\boldsymbol{\xi}_2, \boldsymbol{\Omega}_{22}, \boldsymbol{\alpha}_2^*, g^{(d+1)})$, respectively, where

$$\begin{aligned} \boldsymbol{\alpha}_1^* &= \frac{\boldsymbol{\alpha}_1 + \boldsymbol{\omega}_1 \boldsymbol{\Omega}_{11}^{-1} \boldsymbol{\Omega}_{12} \boldsymbol{\omega}_2^{-1} \boldsymbol{\alpha}_2}{(1 + \boldsymbol{\alpha}_2^\top \bar{\boldsymbol{\Omega}}_{22.1} \boldsymbol{\alpha}_2)^{1/2}}, & \boldsymbol{\alpha}_2^* &= \frac{\boldsymbol{\alpha}_2 + \boldsymbol{\omega}_2 \boldsymbol{\Omega}_{22}^{-1} \boldsymbol{\Omega}_{21} \boldsymbol{\omega}_1^{-1} \boldsymbol{\alpha}_1}{(1 + \boldsymbol{\alpha}_1^\top \bar{\boldsymbol{\Omega}}_{11.2} \boldsymbol{\alpha}_1)^{1/2}}, \\ \bar{\boldsymbol{\Omega}}_{22.1} &= \boldsymbol{\omega}_2^{-1} (\boldsymbol{\Omega}_{22} - \boldsymbol{\Omega}_{21} \boldsymbol{\Omega}_{11}^{-1} \boldsymbol{\Omega}_{12}) \boldsymbol{\omega}_2^{-1}, & \bar{\boldsymbol{\Omega}}_{11.2} &= \boldsymbol{\omega}_1^{-1} (\boldsymbol{\Omega}_{11} - \boldsymbol{\Omega}_{12} \boldsymbol{\Omega}_{22}^{-1} \boldsymbol{\Omega}_{21}) \boldsymbol{\omega}_1^{-1} \end{aligned} \quad (3.8)$$

As in the case of the skew-elliptical family, the log-skew-elliptical family is closed under marginalization but not under conditioning. To present a conditional distribution of $\mathbf{X}_2 | \mathbf{X}_1$ for a log-skew-elliptical random vector \mathbf{X} , we first need to define the so-called extended skew-elliptical family (Arellano-Valle and Genton 2010a; Arellano-Valle and Azzalini 2006; Arellano-Valle et al. 2006). A random vector \mathbf{X} has a multivariate extended skew-elliptical distribution, denoted by $\mathbf{X} \sim \text{ESE}_d(\boldsymbol{\xi}, \boldsymbol{\Omega}, \boldsymbol{\alpha}, \tau, g^{(d+1)})$, if its density is of the form

$$f_{\text{ESE}_d}(\mathbf{x}; \Theta) = \frac{f_d(\mathbf{x}; \boldsymbol{\xi}, \boldsymbol{\Omega}, g^{(d)}) F\{\boldsymbol{\alpha}^\top \boldsymbol{\omega}^{-1}(\mathbf{x} - \boldsymbol{\xi}) + \tau; g_{Q_{\boldsymbol{\xi}, \boldsymbol{\Omega}}}\}}{F\left(\tau / \sqrt{1 + \boldsymbol{\alpha}^\top \boldsymbol{\Omega} \boldsymbol{\alpha}}; g^{(d)}\right)}, \quad \mathbf{x} \in \mathbb{R}^d, \quad (3.9)$$

where $\Theta = (\boldsymbol{\xi}, \boldsymbol{\Omega}, \boldsymbol{\alpha}, g^{(d)}, \tau)$, $\tau \in \mathbb{R}$ is the extension parameter, and other parameters are as defined in (3.1). Note that (3.9) reduces to (3.1) if $\tau = 0$.

Similar to Definition 1, we define \mathbf{X} to be a random vector from a log-extended-skew-elliptical family, denoted as $\mathbf{X} \sim \text{LESE}_d(\boldsymbol{\xi}, \boldsymbol{\Omega}, \boldsymbol{\alpha}, \tau, g^{(d+1)})$, if $\ln(\mathbf{X}) \sim \text{ESE}_d(\boldsymbol{\xi}, \boldsymbol{\Omega}, \boldsymbol{\alpha}, \tau, g^{(d+1)})$ as defined in (3.9). The density function of the LESE random vector may be obtained similarly to (3.4) using rules for transformation of random vectors.

Proposition 5 (conditional distribution). Let $\mathbf{X} \sim \text{LSE}_d(\boldsymbol{\xi}, \boldsymbol{\Omega}, \boldsymbol{\alpha}, g^{(d+1)})$. Con-

sider the partitions defined in Proposition 4. Then

$$(\mathbf{X}_2|\mathbf{X}_1) \sim \text{LESE}_{d-q}(\boldsymbol{\xi}_{2\cdot 1}, \boldsymbol{\Omega}_{22\cdot 1}, \boldsymbol{\alpha}_2, \tau, g_{Q_{\mathbf{x}_1}^{\boldsymbol{\xi}_1, \boldsymbol{\Omega}_{11}}}}^{(d+1)}),$$

where $\boldsymbol{\xi}_{2\cdot 1} = \boldsymbol{\xi}_2 + \boldsymbol{\Omega}_{21}\boldsymbol{\Omega}_{11}^{-1}(\mathbf{x}_1 - \boldsymbol{\xi}_1)$, $\boldsymbol{\Omega}_{22\cdot 1} = \boldsymbol{\Omega}_{22} - \boldsymbol{\Omega}_{21}\boldsymbol{\Omega}_{11}^{-1}\boldsymbol{\Omega}_{12}$, and the extension parameter is $\tau = (\boldsymbol{\alpha}_1 + \boldsymbol{\omega}_1\boldsymbol{\Omega}_{11}^{-1}\boldsymbol{\Omega}_{12}\boldsymbol{\omega}_2^{-1}\boldsymbol{\alpha}_2)^\top \boldsymbol{\omega}_1^{-1}(\mathbf{x}_1 - \boldsymbol{\xi}_1)$.

4. Moments

Similar to the log-elliptical distribution (Fang et al. 1990), if the mixed moments of a log-skew-elliptical distribution exist, they can be expressed conveniently using the characteristic function (or the moment generating function if it exists) of the skew-elliptical distribution.

Proposition 6 (mixed moments). Let $\mathbf{X} \sim \text{LSE}_d(\boldsymbol{\xi}, \boldsymbol{\Omega}, \boldsymbol{\alpha}, g^{(d+1)})$ and $\mathbf{n} = (n_1, n_2, \dots, n_d)^\top$, $n_i \in \mathbb{N}$, $i = 1, \dots, d$. If the mixed moments $E\left(\prod_{i=1}^d X_i^{n_i}\right)$ exist, then

$$E\left(\prod_{i=1}^d X_i^{n_i}\right) = M_{\ln(\mathbf{X})}(\mathbf{n}),$$

where $M_{\ln(\mathbf{X})}(\cdot)$ is the moment generating function of the skew-elliptical random vector $\ln(\mathbf{X}) \sim \text{SE}_d(\boldsymbol{\xi}, \boldsymbol{\Omega}, \boldsymbol{\alpha}, g^{(d+1)})$.

The mixed moment of the multivariate log-skew-normal random vector $\mathbf{X} \sim \text{LSN}_d(\boldsymbol{\xi}, \boldsymbol{\Omega}, \boldsymbol{\alpha})$ can be expressed as

$$E\left(\prod_{i=1}^d X_i^{n_i}\right) = 2 \exp(\boldsymbol{\xi}^\top \mathbf{n} + \mathbf{n}^\top \boldsymbol{\Omega} \mathbf{n} / 2) \Phi\{\boldsymbol{\alpha}^\top \bar{\boldsymbol{\Omega}} \boldsymbol{\omega} \mathbf{n} / (1 + \boldsymbol{\alpha}^\top \bar{\boldsymbol{\Omega}} \boldsymbol{\alpha})^{1/2}\},$$

where the right-hand-side term in the above is the moment generating function of the skew-normal random vector (e.g., Genton 2004, p. 17). For example, the first four

moments of the univariate log-skew-normal random variate $X \sim \text{LSN}(\xi, \omega^2, \alpha)$ are

$$E(X) = 2 \exp(\xi + \omega^2/2) \Phi\{\alpha\omega/(1 + \alpha^2)^{1/2}\}, \quad (3.10)$$

$$E(X^2) = 2 \exp(2\xi + 2\omega^2) \Phi\{2\alpha\omega/(1 + \alpha^2)^{1/2}\},$$

$$E(X^3) = 2 \exp(3\xi + 4.5\omega^2) \Phi\{3\alpha\omega/(1 + \alpha^2)^{1/2}\}, \quad (3.11)$$

$$E(X^4) = 2 \exp(4\xi + 8\omega^2) \Phi\{4\alpha\omega/(1 + \alpha^2)^{1/2}\}.$$

We can use these expressions to estimate the mean and skewness (and also kurtosis) of the fitted log-skew-normal distribution by substituting the parameters with the respective maximum likelihood estimates (MLEs).

As for the log- t distribution, positive moments of the log-skew- t distribution do not exist.

Proposition 7 (log-skew- t moments). The mixed moments $E\left(\prod_{i=1}^d X_i^{n_i}\right)$ of the log-skew- t random vector $\mathbf{X} \sim \text{LST}_d(\boldsymbol{\xi}, \boldsymbol{\Omega}, \boldsymbol{\alpha}, \nu)$ are infinite for any $n_i \geq 0$, $i = 1, \dots, d$, such that $\sum_{i=1}^d n_i > 0$, and any $\boldsymbol{\xi} \in \mathbb{R}^d$, positive definite matrix $\boldsymbol{\Omega}$, $\boldsymbol{\alpha} \in \mathbb{R}^d$, and $\nu > 0$.

5. Inference with log-skew-elliptical distributions

Let $l_{\text{LSE}}(\boldsymbol{\Theta}|\mathbf{x})$ be the log-likelihood function for the log-skew-elliptical model with density (3.4):

$$\begin{aligned} l_{\text{LSE}}(\boldsymbol{\Theta}|\mathbf{x}) &= \ln\{f_{\text{LSE}_d}(\mathbf{x}; \boldsymbol{\Theta})\} = - \sum_{j=1}^d \ln(x_j) + \ln\{f_{\text{SE}_d}(\mathbf{ln}(\mathbf{x}); \boldsymbol{\Theta})\} \\ &= - \sum_{j=1}^d \ln(x_j) + l_{\text{SE}}(\boldsymbol{\Theta}|\mathbf{ln}(\mathbf{x})). \end{aligned} \quad (3.12)$$

From relation (3.12), the LSE log-likelihood differs from the SE log-likelihood only by the term $-\sum_{j=1}^d \ln(x_j)$, which is free of any unknown parameters $\boldsymbol{\Theta}$. As such,

inferences about Θ may be based on the SE log-likelihood. This allows us to use existing estimation methods developed for the skew-elliptical models to estimate parameters of the log-skew-elliptical model. To estimate the parameters, we simply fit the desired skew-elliptical model to the log-transformed original data. Since the LSE and the SE log-likelihoods are equivalent, we briefly summarize the inferential aspects for two important special cases, the log-skew-normal and the log-skew- t .

Inferential properties of skewed distributions received much attention in the literature, especially for the case of skew-normal and skew- t distributions (e.g., Azzalini and Capitanio 1999, 2003; Sartori 2006; Pewsey 2000, 2006; Azzalini and Genton 2008; Arellano-Valle and Azzalini 2008). The two important inferential aspects are the existence of a stationary point at $\boldsymbol{\alpha} = \mathbf{0}$ of the profile log-likelihood function and the unboundedness of the log-likelihood function in some regions of the parameter space. In the case of the univariate skew-normal distribution, Pewsey (2006) proved the existence of a stationary point at $\alpha = 0$ of the profile log-likelihood function. Azzalini and Genton (2008) extended his argument to the multivariate case and showed that the Fisher information matrix of the profile log-likelihood of the skew-normal model is singular at $\boldsymbol{\alpha} = \mathbf{0}$. Azzalini (1985) proposed an alternative (centered) parameterization that alleviates the singularity of the resulting reparametrized information matrix for the univariate skew-normal distribution. Arellano-Valle and Azzalini (2008) extended this centered parametrization to the multivariate case. This unfortunate property seems to vanish in the case of the skew- t distribution. Azzalini and Capitanio (2003) noted that the behavior of the profile log-likelihood function of the skew- t distribution is more regular and demonstrate it numerically with several datasets. Although there is no rigorous proof of it, Azzalini and Genton (2008) presented a theoretical insight into why the Fisher information matrix is not singular at $\boldsymbol{\alpha} = \mathbf{0}$ in the case of the multivariate skew- t distribution.

The unboundedness of the MLEs of the shape and degrees-of-freedom parameters of the skew- t distributions was first discussed by Azzalini and Capitanio (2003). In the case of the univariate standard skew- t distribution with fixed degrees of freedom, the infinite estimate of the shape parameter is encountered when either all observations are positive or all observations are negative which can happen with positive probability. In other more general cases, such as unknown degrees of freedom and the multivariate case, the conditions under which the log-likelihood is unbounded are more complicated and, thus, more difficult to describe. Sartori (2006) and Azzalini and Genton (2008) presented ways of dealing with the unbounded estimates. Sartori (2006) proposed a bias correction to the maximum likelihood estimates. Azzalini and Genton (2008) suggested a deviance-based approach according to which the unbounded MLEs of $(\boldsymbol{\alpha}, \nu)$ are replaced by the smallest values $(\boldsymbol{\alpha}_0, \nu_0)$ such that the likelihood ratio test of $H_0: (\boldsymbol{\alpha}, \nu) = (\boldsymbol{\alpha}_0, \nu_0)$ is not rejected at a fixed level, say 0.1.

B. Precipitation Data Analysis

U.S. national and regional precipitation data are publicly available from the National Climatic Data Center (NCDC), the largest archive of weather data. We use monthly precipitation data measured in inches to hundreds for the period of 1895 through 2007 (113 observations per month). Monthly (divisional) precipitation data are obtained as monthly equally-weighted averages of values reported by all stations within a climatic division. The regional values are computed from the statewide values (which are obtained from the divisional values weighted by area) weighted by area for each of the nine U.S. climatic regions: Northeast, East North Central, Central, Southeast, West North Central, South, Southwest, Northwest, and West (see Figure 6¹). National

¹The map of U.S. climatic regions was made available by the National Oceanic and Atmospheric Administration/Department of Commerce.

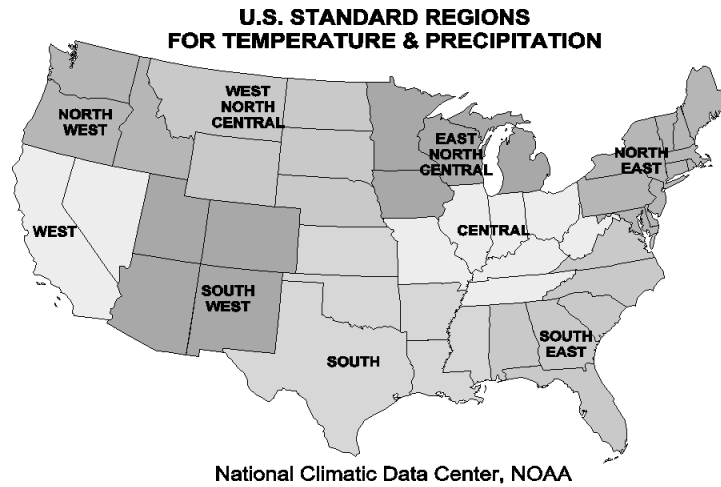


Figure 6. U.S. climatic regions from NCDC.

values are obtained from the regional values weighted by area.

Table 6 and Table 7 present basic summary statistics of the data for each month and climatic region (including national values). The analyses were performed in R (R Development Core Team 2008) and Stata (StataCorp 2009) using, among other capabilities, the R package *sn* developed by Azzalini (2006) and a suite of Stata commands to be presented in Marchenko and Genton (2010a).

1. U.S. national scale

We analyze U.S. national precipitation data by fitting univariate log-skew-normal and log-skew- t models and compare their fits to the conventional log-normal model (e.g., Crow and Shimizu 1988). Separate analyses are carried out for each month. As discussed in Subsection 5 of Section A of this chapter, we estimate the parameters of the log-skew-normal and log-skew- t distributions by fitting the skew-normal and skew- t distributions, respectively, to the log-transformed data.

The attractiveness of the skew-normal and, more generally, the skew-elliptical

Table 6. Monthly (January-June) precipitation minimum, mean, median, and maximum values for nine U.S. climatic regions. Region labels: Northeast (NE), East North Central (ENC), Central (C), Southeast (SE), West North Central (WNC), South (S), Southwest (SW), Northwest (NW), and West (W). The last column records summaries of monthly national values.

Month	U.S. regions									National
	NE	ENC	C	SE	WNC	S	SW	NW	W	
January	0.87	0.32	0.72	0.92	0.16	0.53	0.23	0.43	0.28	0.92
	3.07	1.16	3.09	3.83	0.63	2.29	0.91	3.77	3.08	2.21
	2.93	1.04	2.87	3.61	0.61	2.18	0.79	3.83	2.59	2.17
	7.22	2.47	9.61	7.73	1.25	5.34	2.90	7.81	11.69	3.95
February	0.70	0.31	0.67	1.36	0.30	0.69	0.14	0.51	0.21	0.96
	2.71	1.06	2.71	3.95	0.60	2.30	0.81	2.94	2.42	2.01
	2.53	1.00	2.54	3.94	0.58	2.34	0.73	2.94	2.00	2.03
	5.43	2.40	5.47	7.45	1.07	5.63	2.07	5.75	7.57	3.20
March	0.71	0.23	0.55	1.45	0.39	0.89	0.20	0.58	0.25	0.97
	3.40	1.70	3.82	4.51	0.96	2.71	1.02	2.60	2.23	2.39
	3.33	1.66	3.70	4.55	0.93	2.79	0.93	2.47	1.97	2.35
	6.56	3.50	6.91	8.89	2.10	6.28	2.63	5.25	6.62	3.89
April	1.40	1.04	1.55	0.85	0.50	1.08	0.26	0.61	0.17	1.41
	3.39	2.56	3.90	3.62	1.57	3.22	0.94	1.90	1.27	2.42
	3.32	2.50	3.73	3.51	1.53	3.03	0.84	1.86	1.07	2.41
	6.81	4.85	6.82	7.06	2.83	6.92	2.50	3.81	3.33	3.56
May	0.98	1.15	1.65	0.97	0.65	1.53	0.19	0.36	0.08	1.78
	3.57	3.49	4.38	3.80	2.51	4.08	1.01	1.89	0.87	2.86
	3.43	3.32	4.25	3.62	2.43	4.09	1.01	1.79	0.69	2.87
	7.25	6.85	8.03	7.61	4.63	7.33	2.31	4.19	3.24	4.15
June	1.60	1.41	1.03	2.20	1.25	0.98	0.25	0.36	0.01	1.43
	3.76	4.01	4.24	4.98	2.86	3.66	0.90	1.46	0.41	2.90
	3.67	3.96	4.21	4.78	2.76	3.56	0.84	1.33	0.31	2.93
	8.53	6.68	9.11	8.37	5.27	7.05	1.94	3.02	1.14	4.21

Table 7. Monthly (July-December) precipitation minimum, mean, median, and maximum values for nine U.S. climatic regions. Region labels: Northeast (NE), East North Central (ENC), Central (C), Southeast (SE), West North Central (WNC), South (S), Southwest (SW), Northwest (NW), and West (W). The last column records summaries of monthly national values.

Month	U.S. regions									National
	NE	ENC	C	SE	WNC	S	SW	NW	W	
July	2.02	0.85	1.47	2.94	0.84	1.34	0.90	0.16	0.01	1.81
	3.86	3.54	4.09	5.81	2.00	3.22	1.89	0.62	0.29	2.77
	3.74	3.41	4.02	5.74	1.89	3.17	1.81	0.54	0.25	2.79
	6.57	6.18	8.27	11.56	5.56	6.04	3.58	2.05	1.18	3.85
August	1.78	1.35	1.55	2.71	0.77	0.70	0.56	0.10	0.01	1.77
	3.84	3.54	3.61	5.26	1.72	2.93	1.92	0.74	0.34	2.59
	3.71	3.58	3.57	5.15	1.65	2.89	1.90	0.58	0.25	2.59
	8.01	6.27	6.30	9.77	3.03	6.06	3.25	2.98	2.01	3.55
September	1.25	0.95	0.71	1.91	0.47	0.88	0.09	0.12	0.03	1.45
	3.64	3.36	3.47	4.57	1.53	3.19	1.29	1.19	0.47	2.46
	3.42	3.28	3.34	4.42	1.46	3.15	1.34	1.11	0.30	2.47
	8.04	7.21	6.94	9.68	3.42	6.87	3.07	3.42	2.00	3.57
October	0.45	0.25	0.53	0.53	0.13	0.12	0.02	0.14	0.04	0.54
	3.36	2.33	2.90	3.18	1.14	2.81	1.10	2.22	0.97	2.14
	3.08	2.28	2.67	3.04	1.08	2.46	0.90	2.16	0.91	2.15
	9.43	4.66	7.15	7.33	2.95	7.07	3.67	5.20	3.24	3.72
November	0.88	0.24	0.72	0.83	0.06	0.20	0.09	0.30	0.04	0.88
	3.46	1.80	3.20	2.92	0.75	2.47	0.78	3.68	1.86	2.13
	3.36	1.73	3.13	2.73	0.71	2.36	0.70	3.64	1.57	2.09
	6.34	4.03	7.71	8.39	1.63	6.48	2.33	7.84	5.79	3.76
December	0.98	0.37	0.90	1.18	0.19	0.68	0.16	1.17	0.09	1.22
	3.26	1.27	3.13	3.71	0.62	2.55	0.88	3.91	2.50	2.23
	3.18	1.24	3.14	3.49	0.58	2.47	0.79	3.77	2.08	2.27
	6.74	2.62	7.58	7.05	1.20	5.51	2.29	8.42	7.05	3.60

distributions is that they preserve some pleasant properties of their respective elliptical counterparts. One of them is that the distribution of quadratic forms of skew-elliptical random vectors does not depend on the skewness parameter (and is chi-squared for the skew-normal model). This property is useful, for example, for investigating how well the specified skew-elliptical model fits the data. Specifically, we can compare the probabilities of quantiles of the squared standardized residuals from a univariate skew-normal fit to the probabilities of quantiles of the chi-squared distribution with 1 degree of freedom. We present an example of such a probability plot (PP-plot) for the skew-normal fit for January in Figure 7 (top panel).

From Figure 7, the skew-normal distribution fits the log-precipitation data slightly better than the normal distribution for January. In fact, PP-plots for the skew-normal and skew- t fits revealed that these distributions fit the log-precipitation data slightly better than the normal distribution for most months. For other months all distributions provided good fit.

Figures 8 and 9 present estimated log-skew-normal means and skew-normal skewness indexes, respectively, for each month. Mean estimates are computed from the formula for the first moment of the log-skew-normal distribution given in (3.10) with all parameters being replaced by the obtained maximum likelihood estimates. We present the skewness indexes on the log scale. If desired, skewness indexes on the original scale can be obtained similarly to means from (3.11) using the recursive relationship between moments and central moments. The skew-normal model failed to converge when fitted to the June and October data due to unboundedness of the MLE for the shape parameter α . To alleviate this problem, we used the approaches of Sartori (2006) (labeled on the graph as MMLE for modified MLE) and of Azzalini and Genton (2008) (labeled on the graph as BMLE for bounded MLE), briefly mentioned in Subsection 5 of Section A of this chapter. The obtained two estimates (BMLE and

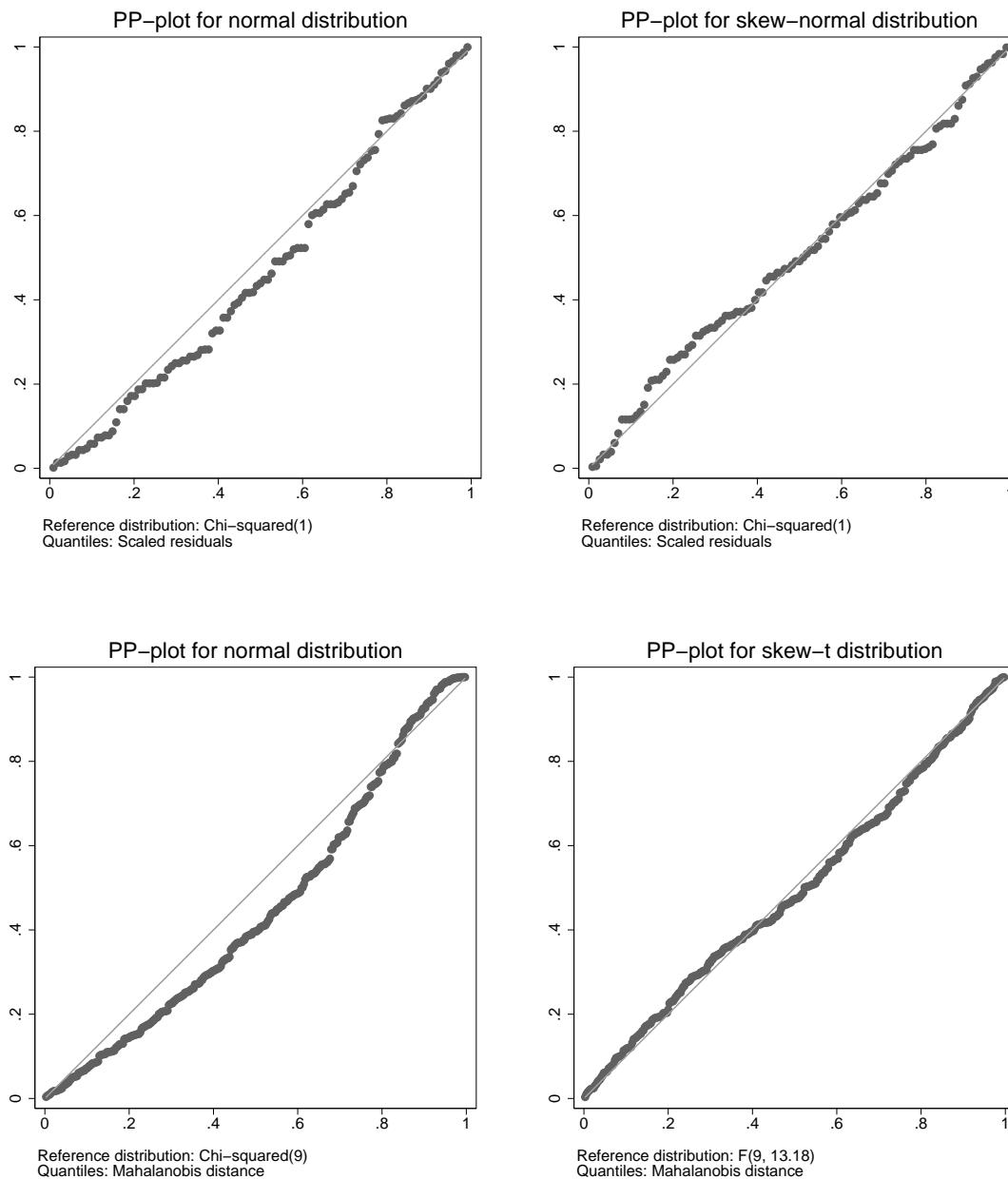


Figure 7. PP-plots for the univariate normal model (top left panel) and the univariate skew-normal model (top right panel) fitted to the log-precipitation for January. PP-plots for the multivariate normal model (bottom left panel) and the multivariate skew- t model (bottom right panel) fitted to the log-precipitation for November.

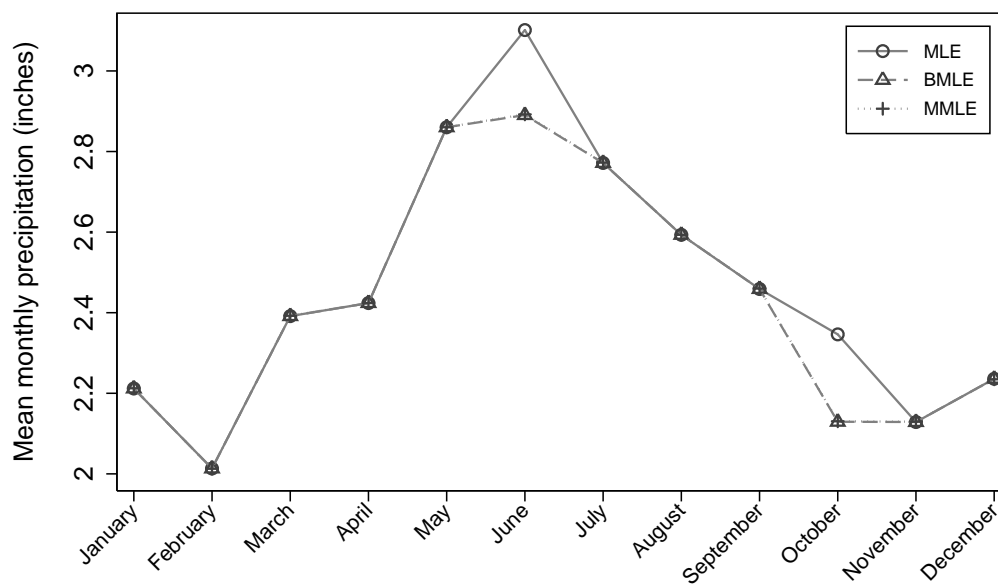


Figure 8. Estimated mean of the log-skew-normal distribution fitted to precipitation for each month.

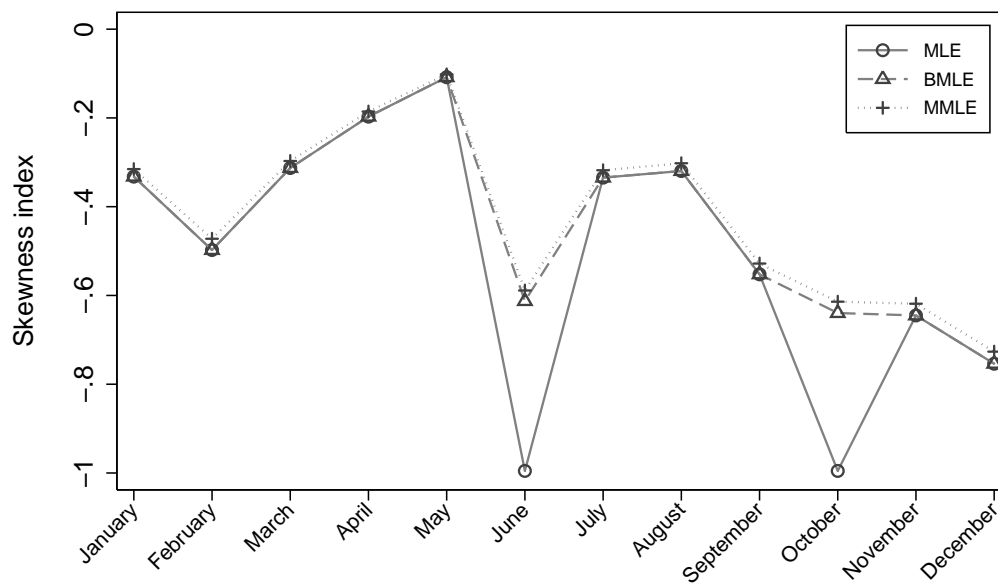


Figure 9. Estimated skewness indexes (log scale) of the log-skew-normal distribution fitted to precipitation for each month.

MMLE) of the shape parameter are similar and their magnitude is comparable to the shape parameter estimates from other months.

The estimated mean precipitation ranges between 2.01 and 2.89 inches (after correcting for the unboundedness of the shape parameter). From Figure 8, a particular trend is visible for the mean precipitation at the national level. The mean precipitation increases in Spring and Summer and decreases in Fall and Winter. The lowest mean precipitation of 2 inches is observed in February and the highest of 2.9 is observed in June. From Figure 9, the estimated skewness index ranges between -0.7 and -0.1 and is negative for all months suggesting that the log-precipitation data are left-skewed. The values of the skewness indexes are not too far from zero for some months which explains only a slight improvement in the fit of the skew-normal distribution. However for other months based on the estimated skewness indexes the shape of the precipitation distribution deviates from symmetry and the fitted model allows to capture such deviations. More specifically, the normality assumption of the log-precipitation data was rejected by the likelihood ratio test of the null hypothesis of $\alpha = 0$ at the 1% significance level for June, September, October, November, and December.

Figure 10 displays the estimated degrees-of-freedom parameter ν , controlling the heaviness of the distributional tails of the fitted log-skew- t models. According to this picture, the degrees of freedom ν is estimated to be 10 and higher for most months, including December, not depicted in the picture, for which the estimated degrees of freedom are very large. For these months the log-skew- t fit may be comparable to that of the log-skew-normal (or log-normal if the skewness index or the estimated shape parameter is not far from zero). For March and, especially, June and July, the estimated degrees of freedom are small suggesting heavier tails of the distribution of the precipitation compared with the log-skew-normal and the log-normal.

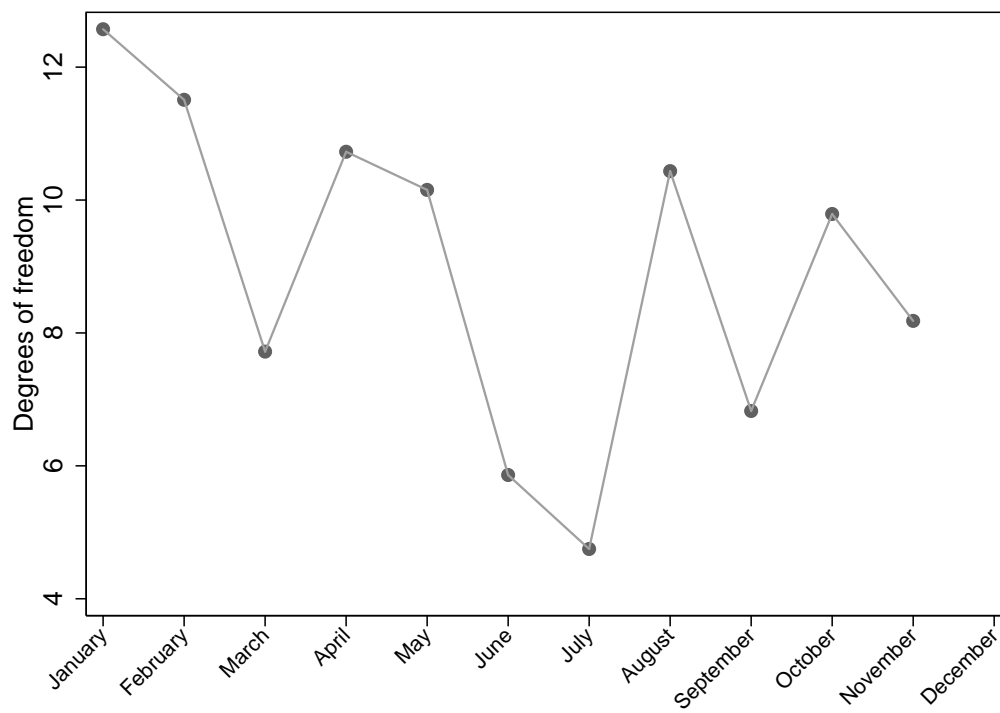


Figure 10. Estimated degrees of freedom of the log-skew- t distribution fitted to the national precipitation data for each month. (December is not included because of the very large estimate).

Since positive moments of the log-skew- t distribution are infinite (Proposition 7), we cannot estimate the mean and marginal skewness on the original scale as for the skew-normal distribution above. However, if estimation of the median (or any other quantile) is of interest, the log-skew- t model may be preferable, especially for estimating extreme events, than the skew-normal model due to its robustness properties (Azzalini and Genton 2008).

If the significance or the precision of the considered quantities is of interest, one can perform a formal significance test (see, for example, Dalla Valle 2007) or compute confidence intervals. Since all considered quantities are functions of MLEs, a classical Delta method can be used to make inferences about them. Another alternative is to use the bootstrap method to obtain confidence intervals. The latter approach may have better finite sample properties than the former, which relies on large sample sizes.

2. U.S. regional scale

We apply multivariate log-skew-normal and log-skew- t models to fit monthly precipitation data over the nine U.S. climatic regions. The multivariate aspect of the model accounts for possible dependence between the precipitation measurements from the regions. More data are usually required to reliably estimate parameters of the multivariate models, compared with their univariate analogs. For multivariate analysis of the regional precipitation data, we consider a 3-months moving window $(t-1, t, t+1)$, resulting in a total of 339 observations per region, to analyze precipitation data for month t .

Bivariate scatter and fitted contour plots visually confirmed a satisfactory fit of the multivariate log-skew- t model to the precipitation data. The multivariate log-skew-normal model exhibits some lack of fit for several months but still fits the

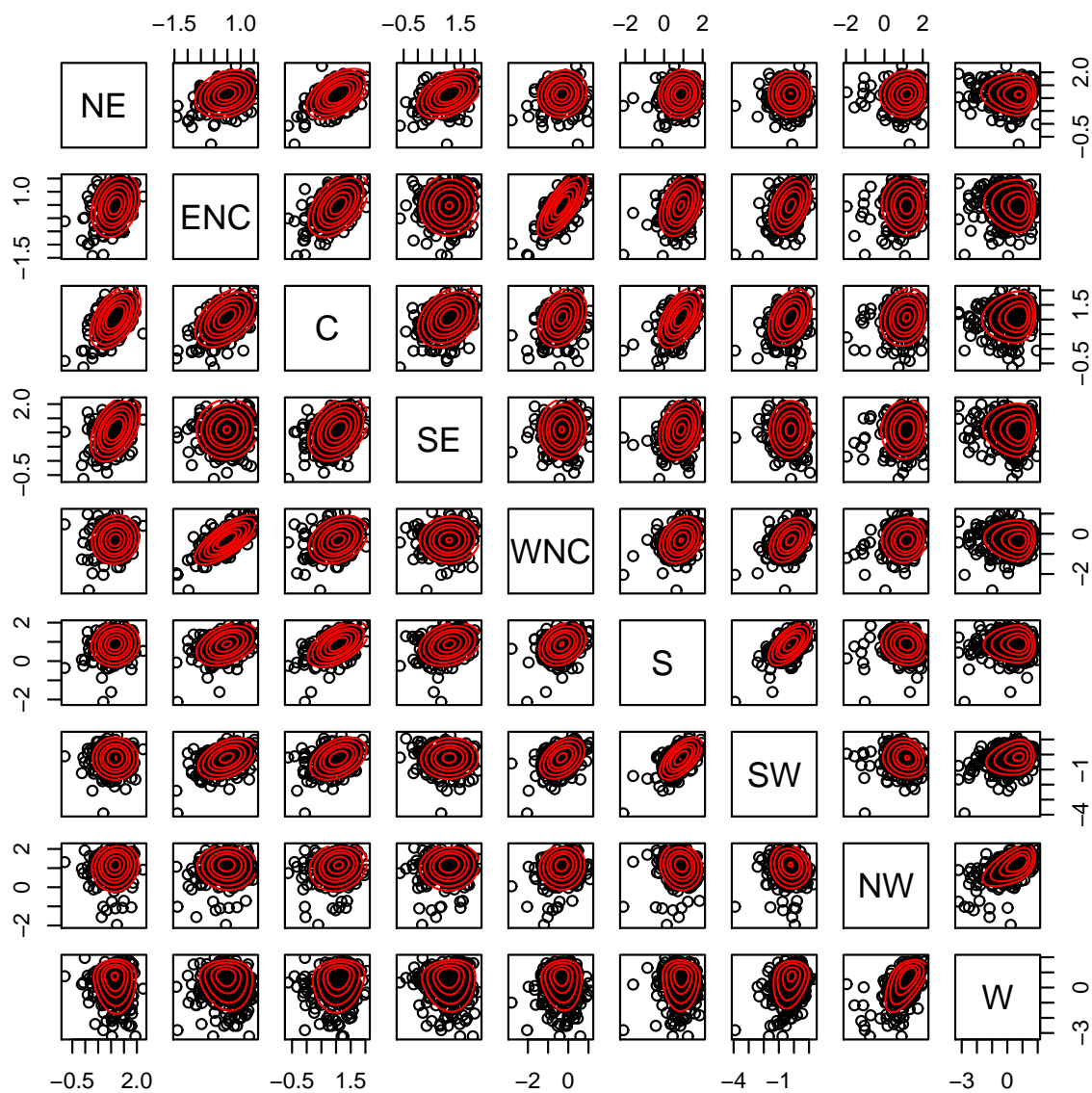


Figure 11. Bivariate scatter plots overlaid with contour plots for the multivariate skew- t model fitted to log-precipitation for November.

data slightly better than the multivariate log-normal model. In this subsection we concentrate on the multivariate log-skew- t model. We present the contour and PP-plots for the multivariate skew- t fit to the log-precipitation for November in Figures 11 and 7 (bottom panel), respectively. The bivariate scatter plot suggests that bivariate distributions for some pairs of regions, such as bivariate distributions involving the West region, may deviate from the bivariate normal distribution. The PP-plot in Figure 7 confirms this by demonstrating some lack of fit of the multivariate normal distribution (left panel) and the improved fit by the multivariate skew- t distribution (right panel). The flexibility of the skew- t model in capturing both the skewness and heavier tails of the data results in a better fit. As with any richer model, however, such flexibility comes with the price of having to estimate more parameters and may also lead to the problem of overfitting.

The estimated degrees of freedom from the multivariate skew- t models fitted to each month (window) is presented in Figure 12. From the graph, there is a noticeable separation in the tail behavior of the observed precipitation distribution over the seasons. The estimates of the degrees of freedom are around 40 and higher for the Winter, Spring, and Summer months and they drop to under 20 for the Fall and early Winter months suggesting somewhat heavier-tailed distributions of the precipitation in these months. We can also see that for November from the bivariate scatter plot depicted in Figure 11: a fair number of precipitation values are observed in the tails of the distribution for some regions. The estimates of the degrees of freedom are fairly large. This may be explained by the fact that only a single parameter, ν , controls the tails of the whole multivariate (9-dimensional) distribution.

We cannot plot the skew- t skewness indexes for the data in the original scale. However we can still infer the information about the changes in skewness from the skewness indexes obtained for the log-transformed data. Mardia (1970) defines the

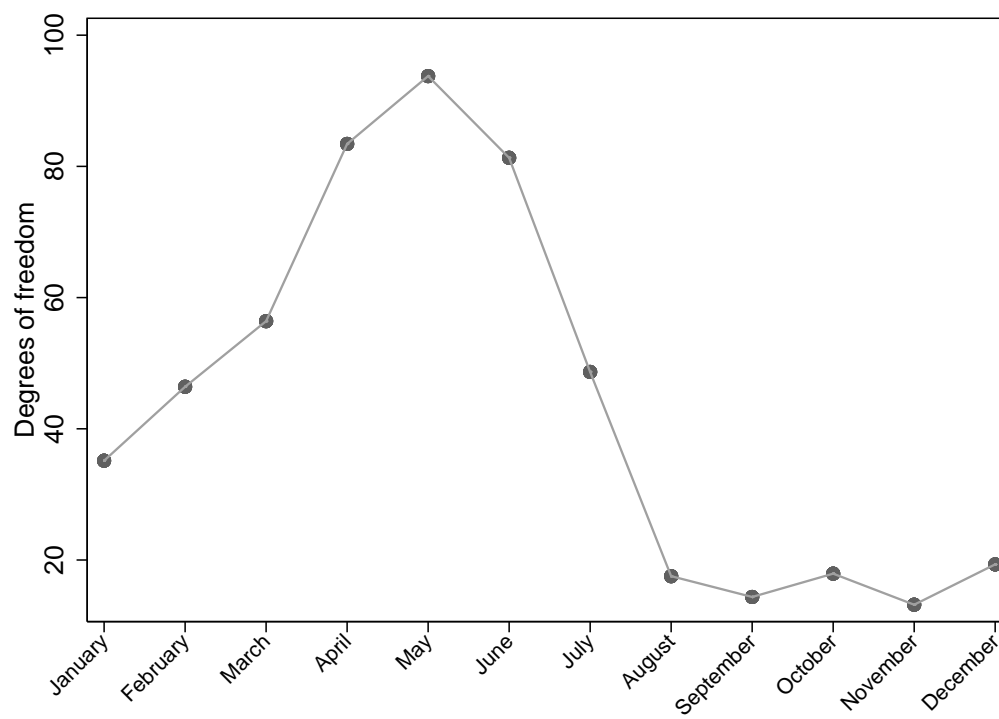


Figure 12. Estimated degrees of freedom of the multivariate log-skew- t distribution fitted to the regional precipitation data for each month.

measures of multivariate skewness and kurtosis for multivariate data and Arellano-Valle and Genton (2010a) present these measures for the multivariate extended skew- t distribution. For the purpose of this exposition we consider marginal skewness indexes. We compute them as follows. First, we compute the parameters of the marginal distributions using the property of linear transformations of the skew-elliptical random vectors (e.g., Capitanio et al. 2003). Then, we use these parameters to compute the univariate skewness index (see Subsection 1 of Section B of this chapter), which we refer to as a marginal skewness index. Plots of these marginal skewness indexes computed for each region based on the log-transformed data are given in Figure 13.

As for the national-level precipitation data, the estimated marginal skewness indexes are close to zero for most months in all regions. This suggests that the marginal distributions of the log-precipitation corresponding to the regions are symmetric for these months. In some regions, a negative estimate of skewness is observed for some months. For example, the skewness values of -0.83 , -0.89 , -0.98 , and -0.44 are observed for July, August, September, and October in the Southwest region. So, the use of the skew-normal or skew- t model may be justified for these months.

As we briefly mentioned in Subsection 1 of Section B of this chapter, using a skew- t distribution may be preferable to using a skew-normal distribution when modeling of the tails of the distribution is of interest. This is important, for example, to obtain accurate estimates of extreme quantiles (say $p > 0.95$). The estimated quantiles may then be used to make inferences about extreme events and their magnitude (Beirlant et al. 2004). The definition of a univariate quantile is straightforward but a concept of a multivariate quantile is more difficult. Various definitions of a multivariate quantile have been proposed and studied in the literature (e.g., Chaudhuri 1996; Chakraborty 2001). As for the skewness index above, we concentrate on the marginal quantiles, i.e. the quantiles obtained using the marginal distributions of a random vector.

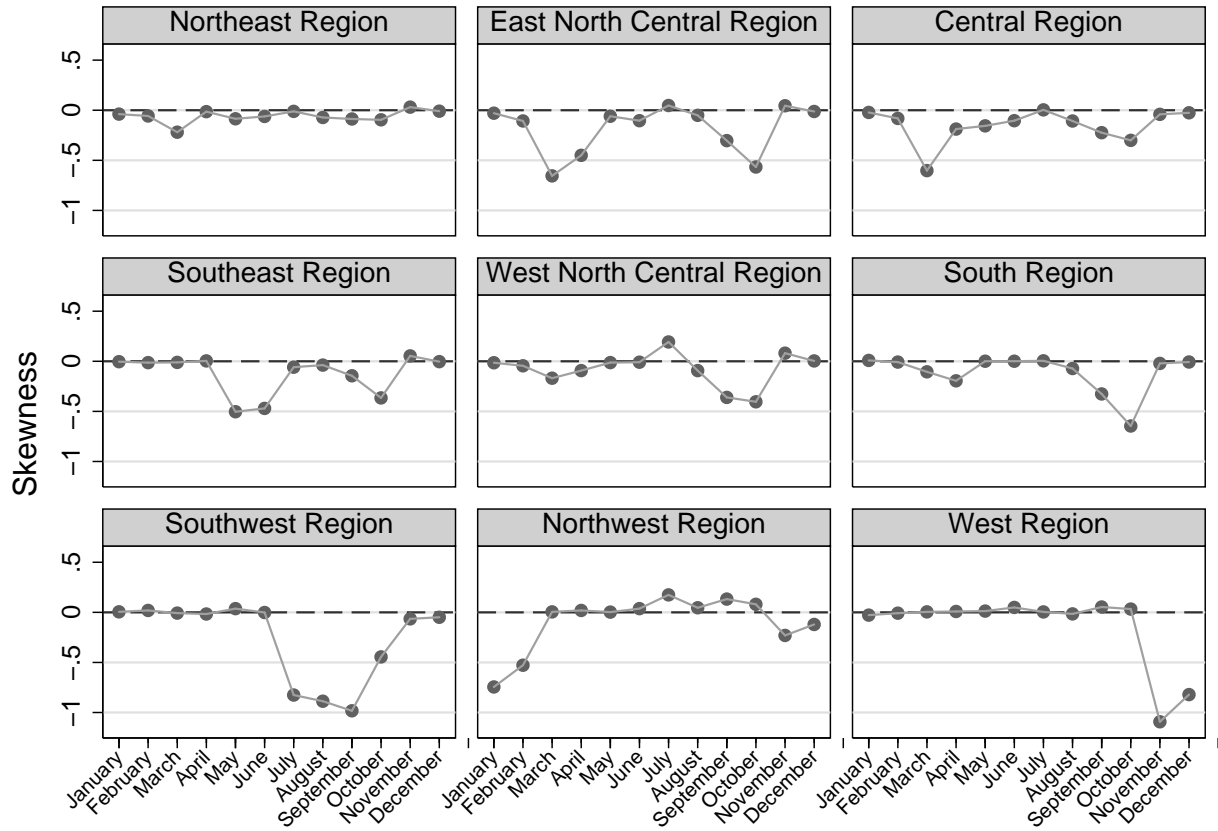


Figure 13. Estimated marginal skewness indexes of the multivariate skew- t distributions fitted to log-precipitation for each month.

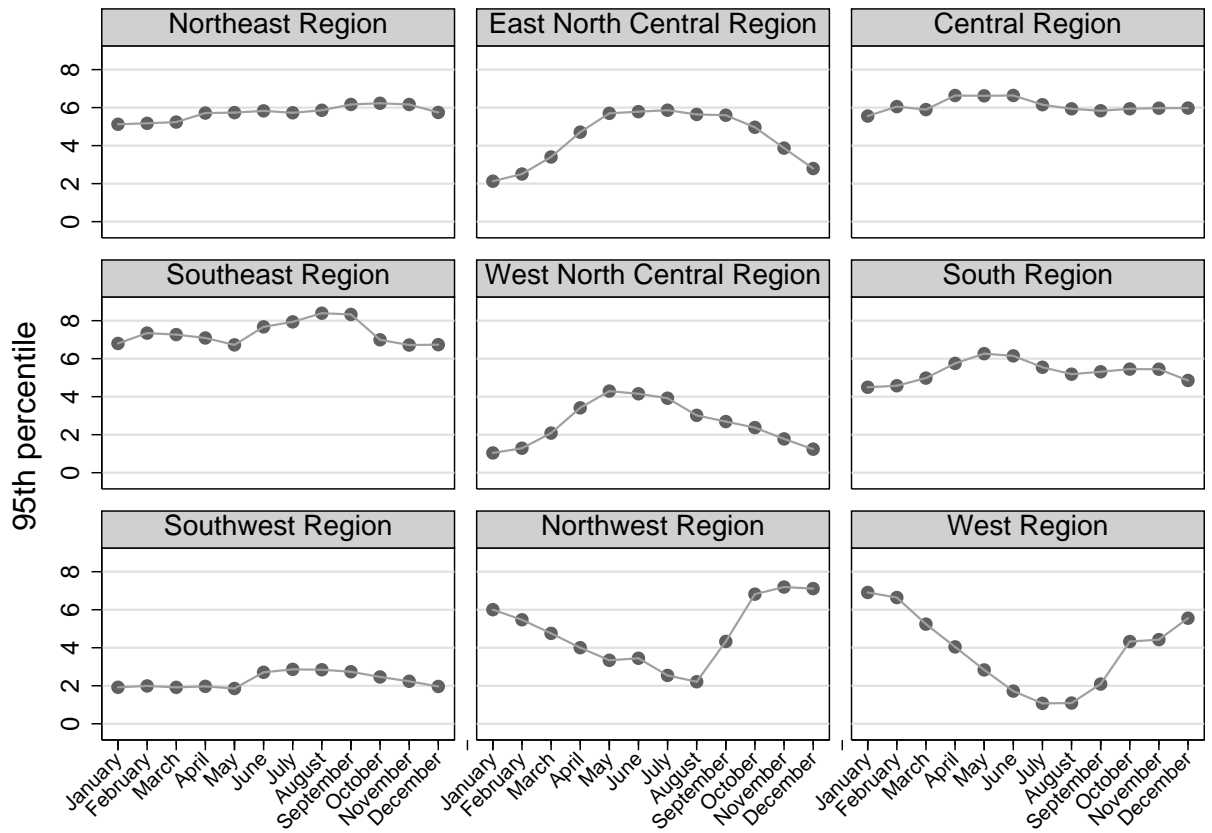


Figure 14. The estimated 95th marginal percentiles of the multivariate log-skew- t distributions fitted to the precipitation for each month.

Figure 14 presents the 95th marginal percentiles corresponding to the multivariate log-skew- t distributions fitted to precipitation data over U.S. climatic regions for all months. We compute the quantiles of the log-skew- t distribution by exponentiating the quantiles obtained from the respective skew- t distribution.

There are only slight variations in the estimated values of the 95th marginal percentile values over months in the Northeast, Central, Southeast, South, and Southwest regions. The variations are much more pronounced in the East North Central, West North Central, Northwest, and West regions. For example, in the West region the estimated values are declining rapidly from 6 inches in January and February to 1 inch in July and August and increase to 5 inches in December. This suggests that observing a monthly average precipitation of 6 inches is very unlikely for July or August whereas there is a 5% chance of observing it in January in the West region.

C. Discussion

The introduced family of multivariate log-skew-elliptical distributions enlarges the family of multivariate distributions with positive support. As for the classical log-elliptical families, its definition is based on the component-wise log transformation, traditionally used to map positive values onto a real line, applied to a skew-elliptical random vector. Although various approaches may be pursued to define multivariate log-skew-elliptical distributions, the considered definition is the most natural one. Its attractiveness is that existing techniques, developed for the skew-elliptical family, can be used to estimate the parameters of and to generate from the defined log-skew-elliptical distribution.

Other extensions to the log-skew-elliptical family, similar to those suggested in the literature for the log-elliptical family, may be considered. One of them is, for

example, the addition of the location parameter $\boldsymbol{\theta} \in \mathbb{R}^d$, by considering $\mathbf{X} = \boldsymbol{\theta} + \mathbf{Z}$, where $\mathbf{Z} \sim \text{LSE}_d(\boldsymbol{\xi}, \boldsymbol{\Omega}, \boldsymbol{\alpha}, g^{(d+1)})$. The corresponding density will be of the form (3.4), where \mathbf{x} and $\ln(\mathbf{x})$ are replaced with $(\mathbf{x} - \boldsymbol{\theta})$ and $\ln(\mathbf{x} - \boldsymbol{\theta})$, respectively, and the constraint $\mathbf{x} > \boldsymbol{\theta}$ is placed on \mathbf{x} . However, this would require the development of specialized maximum-likelihood estimation techniques or at the very least incorporation of the profile-likelihood estimation approach to estimate the extra parameter, $\boldsymbol{\theta}$.

Two considered special cases, the log-skew-normal and log-skew- t distributions, provide more flexibility in capturing various shapes of the distribution compared with the conventional log-normal model, while introducing only d and $d+1$ additional parameters, respectively. The d -dimensional shape parameter $\boldsymbol{\alpha}$ controls the skewness and the degrees-of-freedom parameter ν controls the heaviness of the tails of the underlying distribution on the log-transformed scale. Of course, such extra complexity may not be worthwhile in some applications due to limited availability of the data relative to the dimensionality of the model, in which case difficulties in estimating the model parameters and the problem of overfitting may occur. In the case of the multivariate log-skew-normal and log-skew- t distributions, a sample size sufficient for estimation of the multivariate log-normal model should often suffice for estimation of these models as well.

We stated and proved some properties, known to hold for the skew-elliptical family and the log-elliptical family, in the case of the log-skew-elliptical family. As for the log- t distribution, the positive moments of the log-skew- t distribution do not exist. This introduces some limitations to its use in applications for which the estimation of mean (or other moments and their functions) is the goal. However, it may be preferred to the log-skew-normal and the log-normal distributions in applications for which the tails of the distribution are of interest (e.g., the analysis of extreme events).

We also presented the numerical application of both the univariate and the mul-

tivariate log-skew-normal and log-skew- t models to U.S. monthly precipitation data. The application has some promising results demonstrating that, although for most months and U.S. climatic regions the log-normal distribution provides a satisfactory fit (possibly due to data being averaged over time and space), there are months and regions for which the use of a more flexible parametric model, such as the log-skew-normal and the log-skew- t , is beneficial. Also, in the case of the multivariate log-skew- t (or log- t) models, it would be interesting to see if extending these models to allow for component-specific degrees of freedom results in an improved fit, although the construction of such a generalization is difficult, see the discussion in Azzalini and Genton (2008).

The log-skew-elliptical distributions may also be used to model daily precipitation data. In this case we should see more improvement in the fit of the log-skew-elliptical models over the log-normal model since no averaging over time is done. Since daily precipitation often have zeros, a mixture model, which is a linear combination of continuous distributions and distributions with point mass at zero, can be considered. A log-skew-elliptical distribution can be used for the continuous component of the mixture model. For example, Chai and Baily (2008) investigated such mixture modeling using the univariate log-skew-normal distribution.

CHAPTER IV

A SUITE OF COMMANDS FOR FITTING THE SKEW-NORMAL AND
SKEW- T MODELS

Non-normal data arise often in practice. One common way of dealing with non-normal data is to find a suitable transformation that makes the data more normal-like and apply standard normal-based methods for the transformed data. Finding a suitable transformation can be difficult with multivariate data. Also, for the ease of interpretation, it is often preferable to work with data in the original scale. This motivated a search for more flexible parametric families of distributions to model non-normal data. Because real data often deviate from normality in the tails and/or asymmetry of the distribution, there has been a growing interest in distributions with additional parameters allowing to regulate asymmetry and tails. For example, for heavier-tailed data, the Student's t distribution is often considered. To accommodate asymmetry, skew-normal and skew- t distributions can be considered, which are “skewed” versions of the respective Normal and Student's t distributions. More generally, the family of skew-elliptical distributions is proposed by Branco and Dey (2001) to allow for asymmetry in a class of elliptically symmetric distributions.

The simplest representative of the skew-elliptical family, as defined by Azzalini (1985), is the *skew-normal* distribution. Compared with the Normal distribution, in addition to the location and scale parameters, the skew-normal distribution has a shape parameter, regulating the asymmetry of the distribution. Another commonly used representative is the *skew- t distribution* (Azzalini and Capitanio 2003) that extends the normal distribution to allow for both asymmetry and heavier tails with two additional parameters, the shape and the degrees-of-freedom parameters. These extra parameters allow to capture the features of the data more adequately. Azzalini

and Dalla Valle (1996), Branco and Dey (2001), and Azzalini and Capitanio (2003) introduce multivariate analogs of these distributions.

What makes these distributions appealing for use in practice is that they are simple extensions of their more commonly used counterparts, the Normal and Student's t distributions, and that they share some of their properties. For example, the distribution of the quadratic forms of skew-normal and skew- t random vectors does not depend on the shape parameter (and is chi-square for the skew-normal model as it is for the normal model). This property is useful for evaluating model fit. These distributions are closed under linear transformations and multivariate versions are closed under marginalization (but not conditioning). Similar to the Normal and Student's t distributions, the skew-normal and skew- t distributions can also be adapted to handle nonnegative data by considering their log versions (Azzalini et al. 2003, Marchenko and Genton 2010*b*).

A more detailed description of these and other skewed distributions can be found in the book edited by Genton (2004) and in the review by Azzalini (2005).

The structure of this chapter is the following. We start with a motivating example in Section A and proceed to describe the skewed distributions in more detail in Section B. We present Stata (StataCorp 2009) commands for fitting the skewed regressions in Section C. In Sections D, E, and F we demonstrate more examples of using skew-normal and skew- t models. We conclude the chapter with Section G.

In what follows we refer to Stata concepts, such as command names, option names, datasets, variable names, etc, using the `typewriter` font. All of the presented output is obtained using Stata 11. The commands used are shown with a dot (`.`).

A. Motivating Example

We consider the Australian Institute of Sport data, `ais`, (Cook and Weisberg 1994), repeatedly used in the literature about skewed distributions. The `ais.dta` dataset contains 202 observations (100 females and 102 males) recording 13 biological characteristics of Australian athletes (Table 8).

Table 8. Description of the Australian Institute of Sport data.

```
. use ais
(Biological measures from athletes at the Australian Institute of Sport)
. describe
Contains data from ais.dta
  obs:          202
                                Biological measures from
                                athletes at the Australian
                                Institute of Sport
  vars:          13
  size:         18,584 (99.9% of memory free)
                                30 Mar 2010 11:25
```

variable name	storage type	display format	value label	variable label
<code>female</code>	byte	%9.0g	sex	Gender
<code>lbm</code>	double	%9.0g		Lean body mass (kg)
<code>bmi</code>	double	%9.0g		Body mass index (kg/m ²)
<code>weight</code>	double	%9.0g		Weight (kg)
<code>height</code>	double	%9.0g		Height (m)
<code>bfat</code>	double	%9.0g		Body fat percentage
<code>rcc</code>	double	%9.0g		Red blood cell count
<code>wcc</code>	double	%9.0g		White blood cell count
<code>hc</code>	double	%9.0g		Hematocrit
<code>hg</code>	double	%9.0g		Hemoglobin
<code>fe</code>	int	%9.0g		Plasma ferritin concentration
<code>ssf</code>	double	%9.0g		Sum of skin folds
<code>sport</code>	byte	%9.0g	sport	Sport activity

Sorted by:

For the purpose of illustration, we consider a simple regression model relating lean body mass, `lbm`, to weight and height. To adjust for likely differences in the relationship due to gender we interact `weight` and `height` with `female`. (Or, we could have fit separate regressions for males and females to also allow the variability in the measurements to differ across gender.)

First, we fit a Normal regression to the data using `regress` and specify the

noconstant option to force regression through the origin (Table 9).

Table 9. Regression of lean body mass on weight and height.

```
. regress lbm i.female#c.(weight height), noconstant
```

Source	SS	df	MS	Number of obs = 202		
Model	883259.002	4	220814.75	F(4, 198) =35991.71		
Residual	1214.76071	198	6.13515512	Prob > F = 0.0000		
Total	884473.763	202	4378.58298	R-squared = 0.9986		
				Adj R-squared = 0.9986		
				Root MSE = 2.4769		

lbm	Coef.	Std. Err.	t	P> t	[95% Conf. Interval]	
female#						
c.weight						
0	.7316397	.0237056	30.86	0.000	.6848918	.7783875
1	.5268039	.027829	18.93	0.000	.4719247	.5816831
female#						
c.height						
0	7.706318	1.065338	7.23	0.000	5.605454	9.807183
1	11.11977	1.086061	10.24	0.000	8.978036	13.2615

We examine the distribution of the residuals from the regression model:

```
. predict resid, residuals
. kdensity resid, normal
```

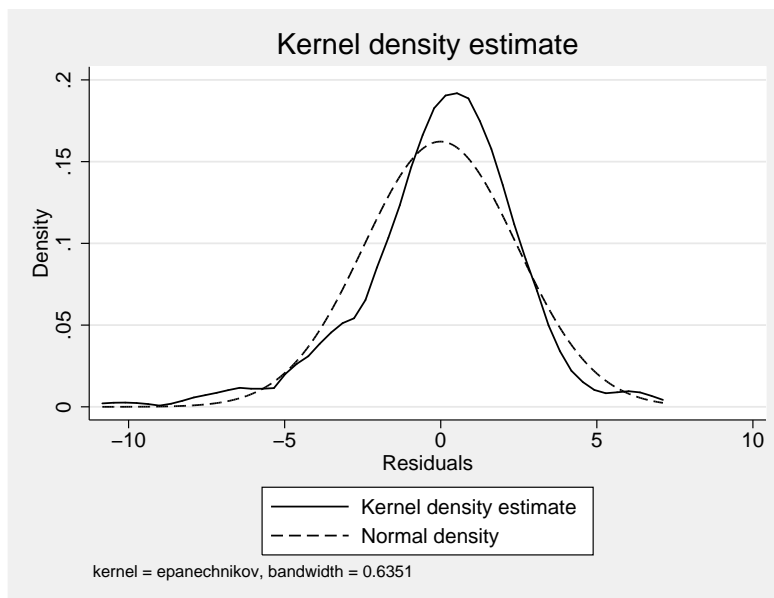


Figure 15. Normal residuals density estimate.

Figure 15 demonstrates a slight skewness (longer left tail) in the distribution of residuals compared with the assumed underlying Normal distribution.

To capture the asymmetry in the data, we fit the skew-normal regression using the new command `skewnreg` and obtain results as shown in Table 10. A quick examination of the output confirms the existence of the negative skewness in the distribution of the data — the estimated skewness (labeled as `gamma` in the output) is -0.66 with a 95% confidence interval of $(-0.88, -0.43)$.

Table 10. Skew-normal regression of lean body mass on weight and height.

```
. skewnreg lbm i.female#c.(weight height), noconstant nolog
```

Skew-normal regression	Number of obs	=	202
	Wald chi2(4)	=	153403.46
Log likelihood = -458.68244	Prob > chi2	=	0.0000

lbm	Coef.	Std. Err.	z	P> z	[95% Conf. Interval]	
female#						
c.weight						
0	.7941643	.0273495	29.04	0.000	.7405602	.8477683
1	.5725823	.0252675	22.66	0.000	.5230589	.6221057
female#						
c.height						
0	4.777069	1.25218	3.82	0.000	2.322841	7.231298
1	9.431404	.9825755	9.60	0.000	7.505591	11.35722
gamma	-.6564894	.1143292	-5.74	0.000	-.8805706	-.4324082
sigma2	5.960411	.6520583			4.810124	7.385776

LR test vs normal regression: chi2(1) = 18.28 Prob > chi2 = 0.0000

We also obtain the residual fit plot in Figure 16 by using command `skewrplot` after `skewnreg`. Figure 16 demonstrates an improved fit to the distribution of residuals.

```
. skewrplot, kdensity
```

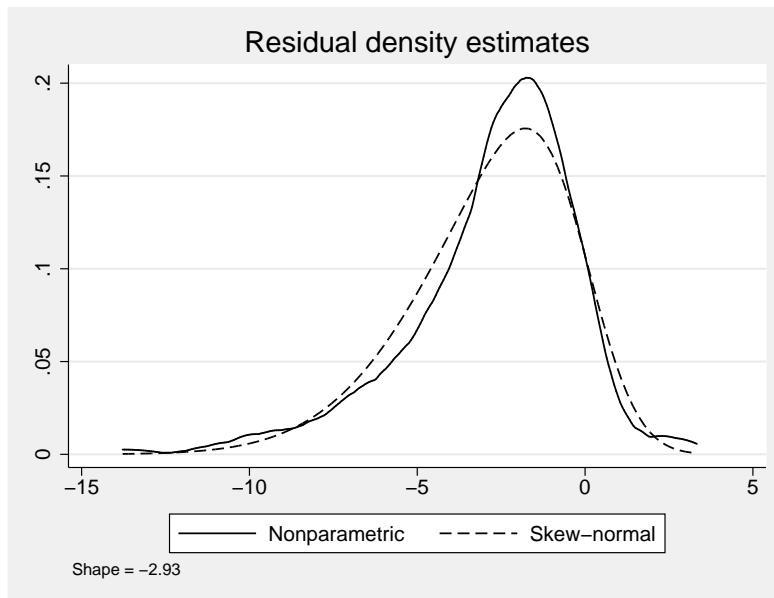


Figure 16. Skew-normal residuals density estimate.

Table 11. Skew- t regression of lean body mass on weight and height.

```
. skewtreg lbm i.female#c.(weight height), noconstant nolog
Skew-t regression                               Number of obs   =       202
                                                Wald chi2(4)    =   82096.51
Log likelihood = -452.19315                    Prob > chi2     =    0.0000
```

lbm	Coef.	Std. Err.	z	P> z	[95% Conf. Interval]
female#					
c.weight					
0	.7959185	.029476	27.00	0.000	.7381466 .8536904
1	.5802067	.0204974	28.31	0.000	.5400326 .6203809
female#					
c.height					
0	5.972125	1.238966	4.82	0.000	3.543796 8.400454
1	10.50348	.7941971	13.23	0.000	8.946881 12.06008
alpha	-2.285011	.8234053	-2.78	0.006	-3.898856 -.6711667
omega2	6.544033	1.876612			2.865941 10.22213
DF	3.940011	1.314396			2.048966 7.576351

```
LR test vs normal regression:      chi2(2) =    31.26   Prob > chi2 = 0.0000
```

Similarly, we can fit a skew- t regression in Table 11. In addition to negative skewness, the estimated degrees of freedom (labeled as DF in the output) of 3.94 suggests that the distribution of residuals has heavier tails than the Normal and skew-normal distributions. By examining the residuals from the skew- t regression fit in Figure 17, we see that the fitted skew- t density line follows the nonparametric estimate very closely.

```
. skewrplot, kdensity
```

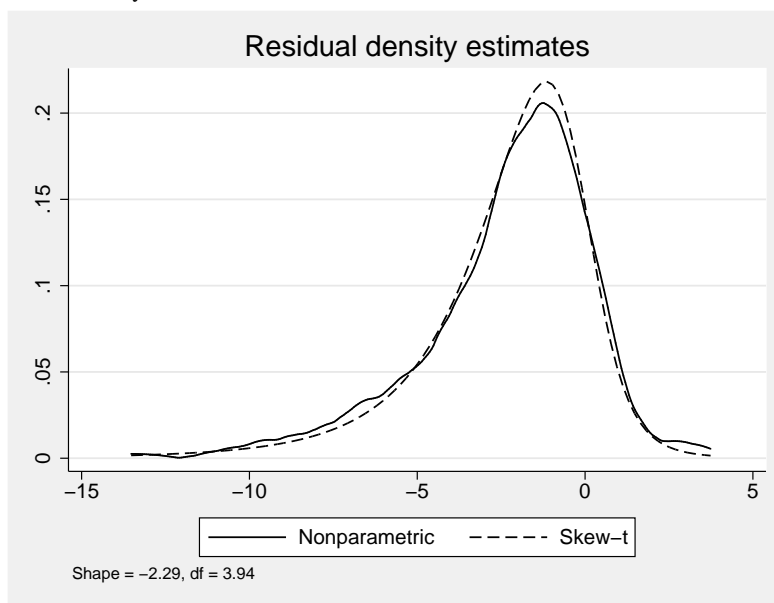


Figure 17. Skew- t residuals density estimate.

As expected, all three regressions find weight and height adjusted for gender to be strong predictors of LBM. However, the coefficients for the height of females and, especially, of males from the skew-normal and skew- t regressions differ greatly from those from the Normal regression.

B. The Skew-Normal and Skew- t Regression Models

1. Definition and some properties

The density of the univariate skew-normal (SN) distribution, $\text{SN}(\xi, \omega, \alpha)$, is

$$f_{\text{SN}}(x; \xi, \omega, \alpha) = 2\omega^{-1}\phi(z)\Phi(\alpha z), \quad x \in \mathbb{R}, \quad (4.1)$$

where $z = \omega^{-1}(x - \xi)$, $\phi(\cdot)$ is the density of a univariate standard normal distribution, and $\Phi(\cdot)$ is the cumulative distribution function of the standard normal distribution. The additional multiplier $2\Phi\{\alpha z\}$ is a “skewness” factor and it is controlled by the shape parameter $\alpha \in \mathbb{R}$. When $\alpha > 0$, the distribution is skewed to the right, when $\alpha < 0$ the distribution is skewed to the left, and when $\alpha = 0$, the skew-normal distribution (4.1) reduces to the Normal distribution.

The univariate skew- t distribution, $\text{ST}(\xi, \omega, \alpha, \nu)$, is defined in a similar manner by introducing a multiplier to the Student’s t density which is a heavier-tailed distribution than the Normal distribution:

$$f_{\text{ST}}(x; \xi, \omega, \alpha, \nu) = 2\omega^{-1}t(z; \nu)T\left\{\alpha z\sqrt{(\nu+1)/(\nu+z^2)}; \nu+1\right\}, \quad x \in \mathbb{R}, \quad (4.2)$$

where $t(x; \xi, \omega^2, \nu)$ is the density of a univariate Student’s t distribution with degrees of freedom ν , and $T(\cdot; \nu+1)$ is the cumulative distribution function of a univariate Student’s t distribution with $\nu+1$ degrees of freedom. Here, again, the shape parameter α regulates asymmetry of the distribution and the degrees-of-freedom parameter $\nu > 0$ regulates the tails of the distribution. When $\alpha = 0$, the density (4.2) reduces to the Student’s t density and when $\alpha = 0$ and the degrees of freedom becomes very large (ν tends to ∞), the skew- t density reduces to the Normal density. By introducing an extra parameter for regulating the tails, the skew- t distribution accommodates outlying observations and, thus, can be viewed as a more robust model than the

skew-normal model; see Azzalini and Genton (2008) for details.

As we mentioned in the introduction, one of the useful properties of the skew-normal and skew- t distributions is that their quadratic forms do not depend on the shape parameter. In the univariate case, if $X \sim \text{SN}(\xi, \omega, \alpha)$, then $(X - \xi)^2/\omega^2 \sim \chi_1^2$. If $X \sim \text{ST}(\xi, \omega, \alpha, \nu)$, then $(X - \xi)^2/\omega^2 \sim F_{1,\nu}$. These properties provide a way of evaluating model fit using quantile-quantile (QQ) or probability-probability (PP) plots.

Multivariate analogs of the skew-normal and skew- t distributions are constructed in a similar manner for the corresponding multivariate Normal and multivariate Student's t distributions. The density of the multivariate skew-normal distribution, $\text{SN}_d(\boldsymbol{\xi}, \boldsymbol{\Omega}, \boldsymbol{\alpha})$, is

$$f_{\text{SN}_d}(\mathbf{x}; \Theta) = 2 \phi_d(\mathbf{x}; \boldsymbol{\xi}, \boldsymbol{\Omega}) \Phi\{\boldsymbol{\alpha}^\top \mathbf{z}\}, \quad \mathbf{x} \in \mathbb{R}^d, \quad (4.3)$$

where $\Theta = (\boldsymbol{\xi}, \boldsymbol{\Omega}, \boldsymbol{\alpha})$, $\mathbf{z} = \boldsymbol{\omega}^{-1}(\mathbf{x} - \boldsymbol{\xi}) \in \mathbb{R}^d$, $\phi_d(\mathbf{x}; \boldsymbol{\xi}, \boldsymbol{\Omega})$ is the density of a d -variate Normal distribution with location $\boldsymbol{\xi}$ and covariance matrix $\boldsymbol{\Omega}$, and $\boldsymbol{\omega}$ is the $d \times d$ diagonal matrix containing the square roots of the diagonal elements of $\boldsymbol{\Omega}$. Similarly to the univariate case, when all d components of $\boldsymbol{\alpha}$ are zero, the multivariate skew-normal density (4.3) reduces to the multivariate Normal density $\phi_d(\cdot)$.

The density of the multivariate skew- t distribution, $\text{ST}_d(\boldsymbol{\xi}, \boldsymbol{\Omega}, \boldsymbol{\alpha}, \nu)$, is

$$f_{\text{ST}_d}(\mathbf{x}; \Theta) = 2 t_d(\mathbf{x}; \boldsymbol{\xi}, \boldsymbol{\Omega}, \nu) T \left\{ \boldsymbol{\alpha}^\top \mathbf{z} \left(\frac{\nu + d}{\nu + Q_{\boldsymbol{\xi}^\Omega}^{\boldsymbol{\xi}}} \right)^{1/2}; \nu + d \right\}, \quad \mathbf{x} \in \mathbb{R}^d, \quad (4.4)$$

where $\Theta = (\boldsymbol{\xi}, \boldsymbol{\Omega}, \boldsymbol{\alpha}, \nu)$, $\mathbf{z} = \boldsymbol{\omega}^{-1}(\mathbf{x} - \boldsymbol{\xi})$, $Q_{\boldsymbol{\xi}^\Omega}^{\boldsymbol{\xi}} = (\mathbf{x} - \boldsymbol{\xi})^\top \boldsymbol{\Omega}^{-1}(\mathbf{x} - \boldsymbol{\xi})$, $t_d(\mathbf{x}; \boldsymbol{\xi}, \boldsymbol{\Omega}, \nu) = \Gamma\{(\nu + d)/2\}(1 + Q_{\boldsymbol{\xi}^\Omega}^{\boldsymbol{\xi}}/\nu)^{-(\nu+d)/2}/\{|\boldsymbol{\Omega}|^{1/2}(\nu\pi)^{d/2}\Gamma(\nu/2)\}$ is the density of a d -variate Student's t distribution with ν degrees of freedom, and $T(\cdot; \nu + d)$ is the cumulative distribution function of a univariate Student's t distribution with $\nu + d$ degrees of freedom. When all d components of $\boldsymbol{\alpha}$ are zero, the multivariate skew- t density (4.4)

reduces to the multivariate Student's t density $t_d(\cdot)$ and to the multivariate Normal density $\phi_d(\cdot)$ when in addition ν tends to ∞ .

Similarly to the univariate case, if $\mathbf{X} \sim \text{SN}_d(\boldsymbol{\xi}, \boldsymbol{\Omega}, \boldsymbol{\alpha})$, then the Mahalanobis distance $(\mathbf{X} - \boldsymbol{\xi})^\top \boldsymbol{\Omega}^{-1}(\mathbf{X} - \boldsymbol{\xi}) \sim \chi_d^2$. If $\mathbf{X} \sim \text{ST}_d(\boldsymbol{\xi}, \boldsymbol{\Omega}, \boldsymbol{\alpha}, \nu)$, then $\frac{1}{d}(\mathbf{X} - \boldsymbol{\xi})^\top \boldsymbol{\Omega}^{-1}(\mathbf{X} - \boldsymbol{\xi}) \sim F_{d,\nu}$.

2. Regression models

Consider a sample $\mathbf{Y} = (y_1, y_2, \dots, y_N)^\top$ of N observations. In linear regression,

$$y_i = \beta_0 + \beta_1 x_{1i} + \dots + \beta_p x_{pi} + \epsilon_i, \quad i = 1, \dots, N, \quad (4.5)$$

where x_{1i}, \dots, x_{pi} define covariate values, β_0, \dots, β_p are the unknown regression coefficients, and ϵ_i is an error term. In Normal linear regression, the errors are assumed to be normally distributed, $\epsilon_i \stackrel{\text{iid}}{\sim} \text{Normal}(0, \sigma^2)$. The skew-normal regression is a linear regression (4.5) with errors from the skew-normal distribution, $\epsilon_i \stackrel{\text{iid}}{\sim} \text{SN}(0, \omega^2, \alpha)$. Similarly, the skew- t regression is defined by (4.5) with $\epsilon_i \stackrel{\text{iid}}{\sim} \text{ST}(0, \omega^2, \alpha, \nu)$. Equivalently, the sample \mathbf{Y} is assumed to follow the skew-normal distribution, $y_i \sim \text{SN}(\xi_i, \omega^2, \alpha)$ or the skew- t distribution, $y_i \sim \text{ST}(\xi_i, \omega^2, \alpha, \nu)$, respectively, where $\xi_i = \beta_0 + \beta_1 x_{1i} + \dots + \beta_p x_{pi}$. However, because the mean μ of a skewed random variate is not the same as the location parameter ξ , $E(\epsilon_i) \neq 0$ (unless $\alpha = 0$) unlike the Normal linear regression. The mean $E(\epsilon_i) = \sqrt{2/\pi}\omega\delta$ for the skew-normal regression and $E(\epsilon_i) = \omega\delta\sqrt{\nu/\pi}\Gamma\{(\nu-1)/2\}/\Gamma(\nu/2)$ when $\nu > 1$ for the skew- t regression, where $\delta = \alpha/\sqrt{1+\alpha^2}$. Then, $E(y_i) = \xi + E(\epsilon_i)$.

Under the multivariate regression setting, \mathbf{Y} becomes an $N \times d$ data matrix, $\boldsymbol{\beta}$ becomes a $p \times d$ matrix of unknown coefficients, and the errors follow the multivariate skew-normal distribution, $\text{SN}_d(\mathbf{0}, \boldsymbol{\Omega}, \boldsymbol{\alpha})$, or the multivariate skew- t distribution, $\text{ST}_d(\mathbf{0}, \boldsymbol{\Omega}, \boldsymbol{\alpha}, \nu)$, respectively.

The method of maximum likelihood is used to obtain estimates of regression coefficients β and other parameters, Ω, α, ν . There are two issues arising with likelihood inference for the skew-normal and skew- t models: 1) the existence of the stationary point at $\alpha = \mathbf{0}$ of the profile log-likelihood function for the skew-normal model; and 2) unbounded maximum likelihood estimates. We discuss each issue in more detail below.

The existence of the stationary point at $\alpha = \mathbf{0}$ for the skew-normal model leads to the singularity of the Fisher information matrix of the profile log-likelihood for the shape parameter α (Azzalini 1985, Azzalini and Genton 2008). This violates standard assumptions underlying the asymptotic properties of the maximum likelihood estimators, and, consequently, leads to slower convergence and possibly bimodal limiting distribution of the estimates (Arellano-Valle and Azzalini 2008). All model parameters $\xi, \Omega,$ and α are identifiable so the issue is really due to the chosen parametrization. To alleviate this, Azzalini (1985) suggested an alternative *centered parametrization* for the univariate skew-normal model. Arellano-Valle and Azzalini (2008) extended this parametrization to the multivariate case. We will discuss the centered parametrization in more detail in Subsection 3 of Section B of this chapter. This unfortunate property seems to vanish in the case of the skew- t distribution, unless the degrees of freedom are large enough so that the skew- t distribution essentially becomes the skew-normal distribution; see Azzalini and Capitanio (2003) and Azzalini and Genton (2008) for details. More generally, the issue of the singularity of multivariate skew-symmetric models was investigated by Ley and Paindaveine (2010) and Hallin and Ley (2010).

Both the skew-normal and skew- t models suffer from the problem of unboundedness of the maximum likelihood estimates for the shape and degrees-of-freedom parameters, that is, the maximum likelihood estimator can be infinite with positive

probability for the finite true value of the parameter. For example, in the cases of the univariate standard skew-normal distribution and the univariate standard skew- t distribution with fixed degrees of freedom, when all observations are positive (or negative), which can happen with positive probability, the likelihood function is monotone increasing and, thus, an infinite estimate of the shape parameter is encountered. In other more general cases, such as unknown degrees of freedom and the multivariate case, the conditions under which the log-likelihood is unbounded are more complicated and, thus, more difficult to describe. Sartori (2006) and Azzalini and Genton (2008) presented ways of dealing with the unbounded estimates. Sartori (2006) proposed a bias correction to the maximum likelihood estimates. Azzalini and Genton (2008) suggested a deviance-based approach according to which the unbounded MLEs of $(\boldsymbol{\alpha}, \nu)$ are replaced by the smallest values $(\boldsymbol{\alpha}_0, \nu_0)$ such that the likelihood ratio test of $H_0: (\boldsymbol{\alpha}, \nu) = (\boldsymbol{\alpha}_0, \nu_0)$ is not rejected at a fixed level, say 0.1. Within a Bayesian framework, Liseo and Loperfido (2006) showed that the estimate of the posterior mode of the shape parameter is finite for the skew-normal model under the Jeffreys prior and Bayes and Branco (2007) considered an alternative noninformative uniform prior for the shape parameter.

3. Centered parametrization

Here, we briefly describe the centered parametrization for the univariate skew-normal distribution as proposed by Azzalini (1985) and outline the points made in Arellano-Valle and Azzalini (2008). More details and the extension to the multivariate case can be found in Arellano-Valle and Azzalini (2008).

Let Y be distributed as $\text{SN}(\xi, \omega^2, \alpha)$. Consider the following decomposition of Y :

$$Y = \xi + \omega Z = \mu + \sigma(Y - \mu_z)/\sigma_z,$$

where $\mu_z = E(Y) = \sqrt{2/\pi}\delta$, $\sigma_z^2 = \text{Var}(Y) = 1 - 2\delta^2/\pi$, and $\delta = \alpha/\sqrt{1 + \alpha^2}$. Let $\gamma = (4 - \pi) \text{sign}(\alpha) (\mu_z/\sigma_z^2)^3/2$ denote the skewness index of Y . Then, mean, standard deviation, and skewness index, (μ, σ, γ) , form the centered parametrization. They are referred to as the *centered parameters* (CP) because they are obtained via centering Y . The set of parameters (ξ, ω, α) are referred to as the *direct parameters* (DP).

The use of CP is advantageous from both estimation and interpretation standpoints. The sampling distributions of the maximum likelihood estimates of CP are closer to quadratic forms, and the profile log-likelihood for γ does not have a stationary point at $\gamma = 0$. Although the shape parameter α can be used as a guidance to whether the normal model is sufficient for analysis, it is easier to infer the actual magnitude of the departure from normality based on the skewness index γ . Also, in the multivariate case, components of a skewness vector $\boldsymbol{\gamma}$ correspond to the skewness indexes of the marginal distributions whereas individual components of $\boldsymbol{\alpha}$ cannot be used to infer the direction or the magnitude of the skewness but only if asymmetry is not present. DP is useful for direct interpretation in original model.

From the above formulas, there is a one-to-one correspondence between CP and DP. So, after obtaining estimates in the CP metric, one can use the formulas above and the delta method to obtain respective estimates and their standard errors in the DP metric, if desired.

The centered parameterization for the skew- t distribution is not yet available; it is currently being developed by Adelchi Azzalini and Reinaldo B. Arellano-Valle.

C. A Suite of Commands for Fitting Skewed Regressions

1. Syntax

The syntaxes of the commands for fitting univariate and multivariate skew-normal regressions are presented in Figures 18 and 19. The syntaxes of the commands for fitting univariate and multivariate skew- t regressions are presented in Figures 20 and 21. The syntaxes for postestimation commands are presented in Figures 22 and Figures 23.

```
skewnreg depvar [indepvars] [if] [in] [weight] [, noconstant constraints(constraints)  
collinear vce(vcetype) level(#) postdp estmetric dpmetric display_options  
maximize_options coeflegend ]
```

Figure 18. The syntax for fitting univariate skew-normal regression.

```
mskewnreg depvars [= indepvars] [if] [in] [weight] [, noconstant fullml dp  
constraints(constraints) collinear vce(vcetype) level(#) postdp estmetric  
dpmetric noshowsigma display_options maximize_options coeflegend ]
```

Figure 19. The syntax for fitting multivariate skew-normal regression.

```
skewtreg depvar [indepvars] [if] [in] [weight] [, noconstant df(#)  
constraints(constraints) collinear vce(vcetype) level(#) postdp estmetric  
display_options maximize_options coeflegend ]
```

Figure 20. The syntax for fitting univariate skew- t regression.

```
mskewtreg depvars [= indepvars] [if] [in] [weight] [, noconstant df(#)
  constraints(constraints) collinear vce(vcetype) level(#) postdp estmetric
  noshowsigma display_options maximize_options coeflegend ]
```

Figure 21. The syntax for fitting multivariate skew- t regression.

```
predict [type] newvar [if] [in] [, xb residuals stdp score ]
```

Figure 22. The syntax for obtaining predictions.

```
skewrplot [, kdensity[(kden_options)] qq[(qq_options)] pp[(pp_options)] ]
```

Figure 23. The syntax for producing goodness-of-fit plots.

2. Description

The `skewnreg` and `skewtreg` commands fit skew-normal and skew- t regression models to univariate data. The `mskewnreg` and `mskewtreg` commands fit skew-normal and skew- t regression models to multivariate data. The skew-normal regression supports both the CP metric (the default) and the DP metric (option `dpmetric`), whereas the skew- t regression supports only the DP metric. In the skew- t regression the degrees-of-freedom parameter can optionally be set to a fixed value via the `df()` option.

The postestimation features include predictions and residual plots. The `predict` command can be used after any of the four estimation commands to obtain linear predictions and their standard errors, residual estimates, and the score estimates. The `skewrplot` command can be used to obtain the residual density plots (option `kdensity()`, assumed by default), and QQ and PP goodness-of-fit plots via options `qq()` and `pp()`, respectively.

3. Options

a. Common estimation options

`noconstant` suppresses the constant term (intercept) in the model.

`constraints(constraints)` specifies the linear constraints to be applied during estimation. The default is to perform unconstrained estimation. See [R] **estimation options** for details.

`collinear` specifies that the estimation command not omit collinear variables. See [R] **estimation options** for details.

`vce(vcetype)` specifies the type of standard error reported, which includes types that are derived from asymptotic theory, that are robust to some kinds of misspecification, that allow for intragroup correlation, and that use bootstrap or jackknife methods; see [R] **vce_option**.

`level(#)` specifies the confidence level, as a percentage, for confidence intervals. The default is `level(95)` or as set by `set level`. This option may be specified either at estimation or upon replay.

`postdp` stores direct parameter estimates and their VCE in `e(b)` and `e(V)`, respectively.

`estmetric` displays results in the estimation metric. The estimation metric used is specific to each estimation command. This option may be specified either at estimation or upon replay.

display_options: `noomitted`, `vsquish`, `noemptycells`, `baselevels`, `allbaselevels`; see [R] **estimation options**. These options may be specified either at estimation or upon replay.

maximize_options: `difficult`, `technique(algorithm_spec)`, `iterate(#)`, `[no]log`, `init(init_spec)`, `trace`, `gradient`, `showstep`, `hessian`, `showtolerance`, `tolerance(#)`, `ltolerance(#)`, `nrtolerance(#)`, `nonrtolerance`;

see [R] **maximize**.

`coeflegend` specifies that the legend of the coefficients and how to specify them in an expression be displayed rather than the coefficient table. This option may be specified either at estimation or upon replay.

b. Other options of `skewnreg`

`dpmetric` specifies to display results in the DP metric instead of the default CP metric. This option may be specified either at estimation or upon replay.

c. Other options of `mskewnreg`

`fullml` specifies to estimate parameters using full maximum likelihood instead of the default profile likelihood estimation.

`dp` specifies to fit the model under the direct parameterization instead of the default centered parameterization.

`dpmetric` specifies to display results in the DP metric instead of the default CP metric. This option may be specified either at estimation or upon replay.

`noshowsigma` specifies to suppress the display of the covariance parameter estimates.

d. Other options of `skewtreg`

`df(#)` specifies to fix the degrees-of-freedom parameter at `#` during estimation. This is equivalent to the constrained estimation using the `constraints()` option when the degrees-of-freedom parameter is set to `#`.

e. Other options of `mskewtreg`

`df(#)` specifies to fix the degrees-of-freedom parameter at `#` during estimation. This

is equivalent to the constrained estimation using the `constraints()` option when the degrees-of-freedom parameter is set to `#`.

`noshowsigma` specifies to suppress the display of the covariance parameter estimates.

f. Other options of `predict`

`xb`, the default, calculates the linear prediction.

`residuals` calculates the residuals.

`score` calculates first derivative of the log likelihood with respect to `xb`.

`stdp` calculates the standard error of the linear prediction.

g. Options of `skewrplot`

`kdensity(kdens_options)` specifies to produce the residual fit plot; the default after univariate regressions. `kdens_options` specifies options as allowed by [G] **twoway kdensity**. This option is not allowed after `mskewnreg` or `mskewtreg`.

`qq(qq_options)` specifies to produce the quantile-quantile plots of the observed residuals versus the residuals obtained from the fitted parametric model. `qq_options` specifies options as allowed by `qqplot` in [R] **diagnostic plots**.

`pp(pp_options)` specifies to produce the probability plots of the observed residuals versus the residuals obtained from the fitted parametric model.

D. Analyses of Australian Institute of Sport Data

1. Univariate analysis

Recall our motivating example using the Australian Institute of Sport data. Now we describe each command in more detail. We start with the skew-normal regression in Table 12.

Table 12. Skew-normal regression analysis of the AIS data.

```

. use ais
(Biological measures from athletes at the Australian Institute of Sport)
. skewnreg lbm i.female#c.(weight height), noconstant nolog
Skew-normal regression           Number of obs   =       202
                                Wald chi2(4)    =    153403.46
Log likelihood = -458.68244      Prob > chi2    =       0.0000

```

lbm	Coef.	Std. Err.	z	P> z	[95% Conf. Interval]	
female#						
c.weight						
0	.7941643	.0273495	29.04	0.000	.7405602	.8477683
1	.5725823	.0252675	22.66	0.000	.5230589	.6221057
female#						
c.height						
0	4.777069	1.25218	3.82	0.000	2.322841	7.231298
1	9.431404	.9825755	9.60	0.000	7.505591	11.35722
gamma	-.6564894	.1143292	-5.74	0.000	-.8805706	-.4324082
sigma2	5.960411	.6520583			4.810124	7.385776

LR test vs normal regression: chi2(1) = 18.28 Prob > chi2 = 0.0000

By default, `skewnreg` estimates and displays model parameters in the CP metric, as discussed in Subsection 3 of Section B of this chapter. From the output, both weight and height are strong predictors of lean body mass measurements and the relationship is different between males and females. The estimated skewness index, labeled as `gamma` in the output, is -0.66 which suggests that the distribution of `lbm` is skewed to the left. According to the reported test of $H_0: \gamma = 0$ with the test statistic of -5.74 , we have strong evidence that there is asymmetry in the distribution of `lbm` and, thus, the skew-normal regression may be more appropriate for the analysis than the Normal regression. The likelihood ratio test for the skew-normal regression versus the Normal linear regression, reported at the bottom of the table, also favors the skew-normal model.

We can redisplay results in the DP metric by using the `dpmetric` option (Table 13). Notice that all regression coefficients remain the same: the transformation from the CP to DP metric only changes the intercept. The estimate of the shape parameter,

labeled as `alpha` in the output, is -2.93 with the 95% confidence interval of $(-4.31, -1.56)$. The confidence interval does not include 0, corresponding to the Normal regression, which agrees with our earlier findings. Also, note that the squared scale parameter `omega2` is now reported instead of the variance `sigma2`.

Table 13. Skew-normal regression of lean body mass on weight and height in the `DP` metric.

```
. skewnreg, dpmetric
Skew-normal regression           Number of obs   =       202
                                Wald chi2(4)    =  153403.46
Log likelihood = -458.68244      Prob > chi2     =       0.0000
```

lbm	Coef.	Std. Err.	z	P> z	[95% Conf. Interval]	
female#						
c.weight						
0	.7941643	.0273495	29.04	0.000	.7405602	.8477683
1	.5725823	.0252675	22.66	0.000	.5230589	.6221057
female#						
c.height						
0	4.777069	1.25218	3.82	0.000	2.322841	7.231298
1	9.431404	.9825755	9.60	0.000	7.505591	11.35722
alpha	-2.933908	.7030481			-4.311857	-1.555959
omega2	13.87301	2.038646			9.877334	17.86868

```
LR test vs normal regression:      chi2(1) =   18.28   Prob > chi2 = 0.0000
```

In our earlier example, we used the `skewrplot` command to graphically evaluate model fit by comparing residual density estimates obtained nonparametrically with those from the skew-normal distribution evaluated at the MLES of the model parameters (Figure 15). Alternatively, we can obtain QQ or PP plots by using the respective options. For example, the `skewrplot, pp name(snpp, replace)` command specification produces the PP plot of the probabilities of quantiles of the squared standardized residuals from the fitted skew-normal model to the probabilities of quantiles of the chi-squared distribution with 1 degree of freedom (Figure 24). The `name(snpp)` option is used to store our graph in memory under the name `snpp` for later use. Ac-

According to the PP plot, the skew-normal model does not seem to fit data well in the middle of the distribution, as we also observed in Figure 16. See Dalla Valle (2007) for a formal test of the skew-normality in a population.

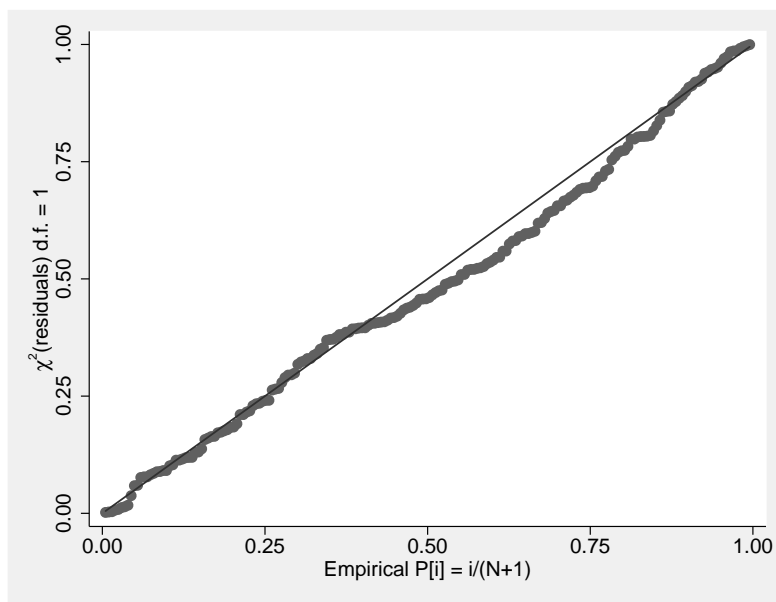


Figure 24. Probability plot for the skew-normal model.

Next, we store estimation results from the skew-normal regression for later comparison with other models:

```
. estimates store skewn
```

Now we fit the skew- t model (Table 14). The centered parametrization for the skew- t model is not yet available although it is currently being investigated by researchers. Thus, the `skewtreg` command reports results only in the DP metric. Compared with the output of direct parameters from `skewnreg`, the `skewtreg` command reports an additional estimate of the degrees of freedom. The estimate of the degrees of freedom is 3.94 with a 95% confidence interval of (2.05, 7.58) which implies heavier tails for the distribution of `lbm`. The estimate for the shape parameter `alpha`

Table 14. Skew- t regression analysis of the AIS data.

```

. skewtreg lbm i.female#c.(weight height), noconstant nolog
Skew-t regression                Number of obs   =       202
                                Wald chi2(4)    =    82096.51
Log likelihood = -452.19315      Prob > chi2   =       0.0000

```

lbm	Coef.	Std. Err.	z	P> z	[95% Conf. Interval]	
female#						
c.weight						
0	.7959185	.029476	27.00	0.000	.7381466	.8536904
1	.5802067	.0204974	28.31	0.000	.5400326	.6203809
female#						
c.height						
0	5.972125	1.238966	4.82	0.000	3.543796	8.400454
1	10.50348	.7941971	13.23	0.000	8.946881	12.06008
alpha	-2.285011	.8234053	-2.78	0.006	-3.898856	-.6711667
omega2	6.544033	1.876612			2.865941	10.22213
DF	3.940011	1.314396			2.048966	7.576351

LR test vs normal regression: chi2(2) = 31.26 Prob > chi2 = 0.0000

is -2.29 and is larger than the estimate obtained earlier from the skew-normal regression -2.93 , although the skew-normal estimate is still within the reported 95% confidence interval $(-3.90, -.67)$. Again, the reported likelihood ratio test rejects the hypothesis of normality, although the results from this test should be interpreted with caution because it does not account for the fact that the degrees of freedom ν are tested at the boundary value $\nu = \infty$ (DiCiccio and Monti 2009).

We can also perform the likelihood ratio test of the skew- t model versus the skew-normal model using the `lrtest` command (Table 15). (The above comment about boundary correction applies here as well.) Because `skewnreg` and `skewtreg` are two different estimation commands, we need to specify the `force` option to obtain results. Although using this option is generally not recommended, it is safe in our case because we know that the skew-normal model is nested within the skew- t model. According to the results in Table 15, the likelihood ratio test favors the skew- t model over the skew-normal model.

Table 15. Likelihood-ratio test of the skew- t regression versus the skew-normal regression.

```
. lrtest skewn ., force
Likelihood-ratio test
(Assumption: skewn nested in .)
LR chi2(1) = 12.98
Prob > chi2 = 0.0003
```

We can also compare the two fits visually using, for example, PP plots. We use `skewrplot`, `pp` to obtain the PP plot after `skewtreg`. We then combine the PP plot obtained earlier for the skew-normal regression with the one for the skew- t regression using `graph combine` to obtain Figure 25.

```
. skewrplot, pp name(stpp) nodraw
. graph combine snpp stpp
```

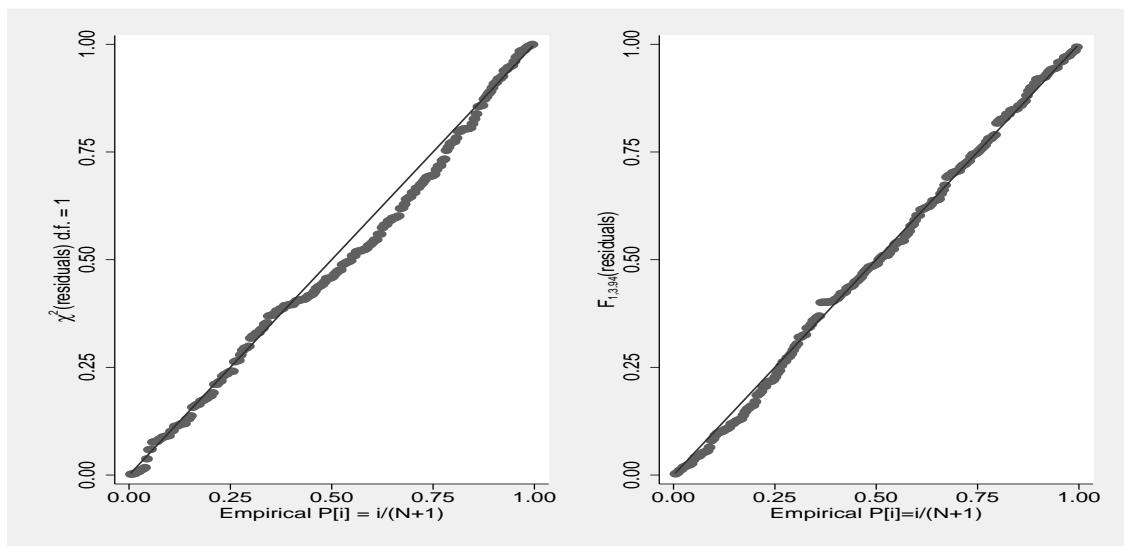


Figure 25. Probability plots for the skew-normal model (left panel) and the skew- t model (right panel).

According to Figure 25, the skew- t model fits the `lbm` regression better than the skew-normal model, although it may suffer from possible data overfitting.

Alternatively, we can use information criteria to compare the two models (Ta-

ble 16). Both AIC and BIC are smaller for the skew- t model suggesting that it is preferable to the skew-normal model.

Table 16. AIC and BIC for the skew-normal and skew- t models.

```
. estimates stats skewn .
```

Model	Obs	ll(null)	ll(model)	df	AIC	BIC
skewn	202	.	-458.6824	6	929.3649	949.2145
.	202	.	-452.1932	7	918.3863	941.5442

Note: N=Obs used in calculating BIC; see [R] BIC note

We can also compare results from all three regressions, including the Normal regression, side-by-side by using `estimates table`. To use `estimates table`, we must first obtain estimation results from all three models.

Because there is no CP parametrization for the skew- t regression, we can compare results only in the DP metric. Although `skewtreg` displays results in the DP metric, the results are saved in the estimation metric. To save results in the DP metric, we use the `postdp` option.

```
. skewtreg, postdp
. estimates store skewt_dp
```

Next, we repeat the same for the skew-normal regression. Prior to reposting results to the DP metric, we use `estimates restore` to make the estimation results of the previously fitted `skewnreg` active.

```
. estimates restore skewn
(results skewn are active now)
. skewnreg, postdp
. estimates store skewn_dp
```

Finally, we obtain estimation results from the Normal linear regression:

```
. regress lbm i.female#c.(weight height), noconstant
(output omitted)
. estimates store reg
```

We now combine all three estimation results in one Table 17 using `estimates table`.

Table 17. Combined regression results.

```
. estimates table reg skewn_dp skewt_dp, equation(1) star(0.05 0.01 0.005)
```

Variable	reg	skewn_dp	skewt_dp
#1			
female#			
c.weight			
0	.73163967***	.79416426***	.79591845***
1	.52680392***	.5725823***	.58020674***
female#			
c.height			
0	7.7063182***	4.7770693***	5.9721247***
1	11.119767***	9.4314038***	10.503479***
alpha			
_cons		-2.9339079***	-2.2850114**
omega2			
_cons		13.873007***	6.5440334***
df			
_cons			3.9400106***

legend: * p<.05; ** p<.01; *** p<.005

According to the three regression models, all considered covariates are strong predictors of lean body mass measurements. Coefficient estimates are highly significant at a 0.005 level of a test against zero. Both `skewnreg` and `skewtreg` report larger estimates of the coefficients for `weight` than the Normal regression. The estimates of the coefficients for `height` from the skew- t regression fall between the estimates reported by the other two regressions. The estimates of the shape parameter provide significant evidence against the null hypothesis $H_0: \alpha = 0$ for both models. Because tests against zero are not appropriate for the scale and degrees-of-freedom parameters, the significance levels, reported automatically by `estimates table`, should be ignored.

2. Multivariate analysis

Suppose we are interested in the distribution of `lbm` and `bmi`, recording the body mass index. The scatter plot in Figure 26 of the `lbm` and `bmi` values suggests that the two variables are related and, thus, should be analyzed jointly.

```
. use ais
(Biological measures from athletes at the Australian Institute of Sport)
. scatter lbm bmi, ytitle(Lean body mass) xtitle(Body mass index)
```

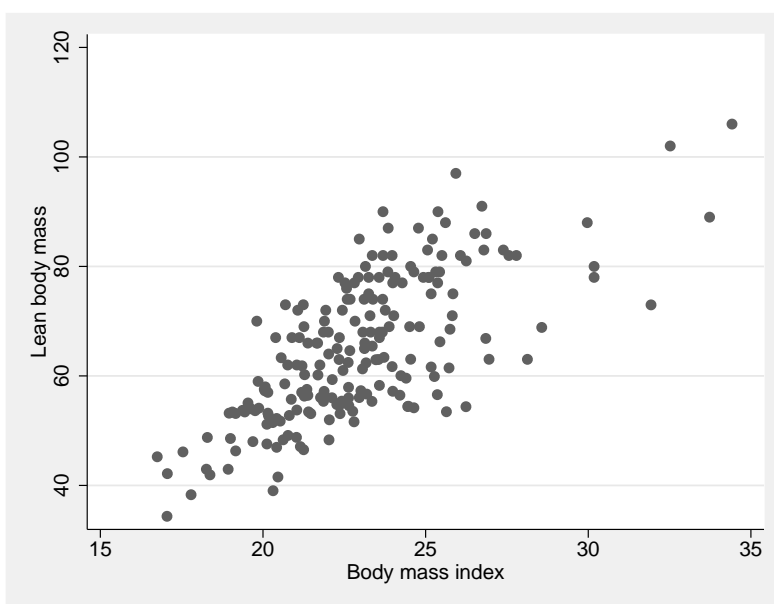


Figure 26. Scatter plot of `lbm` and `bmi`.

The scatter plot also suggests that the joint distribution of `lbm` and `bmi` is somewhat asymmetric and so we fit the bivariate skew-normal distribution to `lbm` and `bmi` using `mskewnreg` (Table 18). Here we chose to estimate all parameters by using full maximum likelihood (option `fullml`) instead of the default profile log-likelihood approach. The estimate of the skewness parameter for `lbm` is close to zero and, according to the z -test ($p = 0.520$), the hypothesis of $H_0: \gamma_1 = 0$ cannot be rejected. For `bmi`, however, there is a strong evidence that the skewness parameter is different

from zero.

Table 18. Bivariate skew-normal fit to lean body mass and body mass index.

```
. mskewnreg lbm bmi, fullml nolog
Skew-normal regression           Number of obs =      202
                                Wald chi2(0)   =      .
Log likelihood = -1213.2609      Prob > chi2    =      .
```

		Coef.	Std. Err.	z	P> z	[95% Conf. Interval]	
lbm	_cons	64.9224	.916583	70.83	0.000	63.12593	66.71887
bmi	_cons	23	.1964777	117.06	0.000	22.61491	23.38509
gamma	1	.006136	.0095483	0.64	0.520	-.0125783	.0248504
	2	.4534164	.0935876	4.84	0.000	.269988	.6368447
Sigma	11	169.681	16.86091			136.6342	202.7278
	12	26.31286	3.148478			20.14195	32.48376
	22	7.910924	.820492			6.302789	9.519059

```
LR test vs normal regression:      chi2(2) =    37.55  Prob > chi2 = 0.0000
```

The joint test of $H_0: \gamma_1 = 0, \gamma_2 = 0$ (see Table 19) and the reported LR test strongly reject the hypothesis of bivariate normality for `lbm` and `bmi`.

Table 19. Joint test of bivariate normality for lean body mass and body mass index.

```
. test [gamma1]_cons [gamma2]_cons
( 1) [gamma1]_cons = 0
( 2) [gamma2]_cons = 0
      chi2( 2) =    52.16
      Prob > chi2 =    0.0000
```

We can also redisplay the results in the DP metric in Table 20. Notice that the estimate of α_1 corresponding to the shape parameter of `lbm` in the DP metric is very far from zero compared with the skewness index, reported earlier. As we mentioned in Subsection 3 of Section B of this chapter, the individual shape parameters are poor estimates of the magnitude of the asymmetry. Although their zero values provide

evidence that the multivariate Normal model may be adequate the opposite is not necessarily true, as we witnessed in this example.

Table 20. Bivariate skew-normal fit to lean body mass and body mass index in the DP metric.

```
. mskewnreg, dpmetric
Skew-normal regression           Number of obs =      202
                                Wald chi2(0)   =      .
Log likelihood = -1213.2609      Prob > chi2    =      .
```

		Coef.	Std. Err.	z	P> z	[95% Conf. Interval]	
lbm	_cons	61.76091	1.859599	33.21	0.000	58.11616	65.40565
bmi	_cons	20.13544	.292038	68.95	0.000	19.56306	20.70782
alpha	1	-2.302208	.5772413	-3.99	0.000	-3.43358	-1.170836
	2	5.515455	1.30094	4.24	0.000	2.96566	8.065251
Omega	11	179.676	21.2868			137.9547	221.3974
	12	35.36914	7.520692			20.62885	50.10942
	22	16.11664	2.297473			11.61367	20.6196

LR test vs normal regression: chi2(2) = 37.55 Prob > chi2 = 0.0000

We again compare the fit against the Normal model using the PP plots. Figure 27 shows that the bivariate skew-normal model fits the data better than the bivariate Normal model.

We can fit the bivariate skew- t model as shown in Table 21. The estimated degrees of freedom are large and the estimates are very close to those from the skew-normal model suggesting that the skew-normal model is adequate for modeling `lbm` and `bmi`.

We can also adjust the location for gender by including variable `female` as a covariate as depicted in Table 22. (Alternatively, we could fit separate regressions for males and females to allow all parameters of the joint distribution to vary across gender.)

```
. skewrplot, pp normal
```

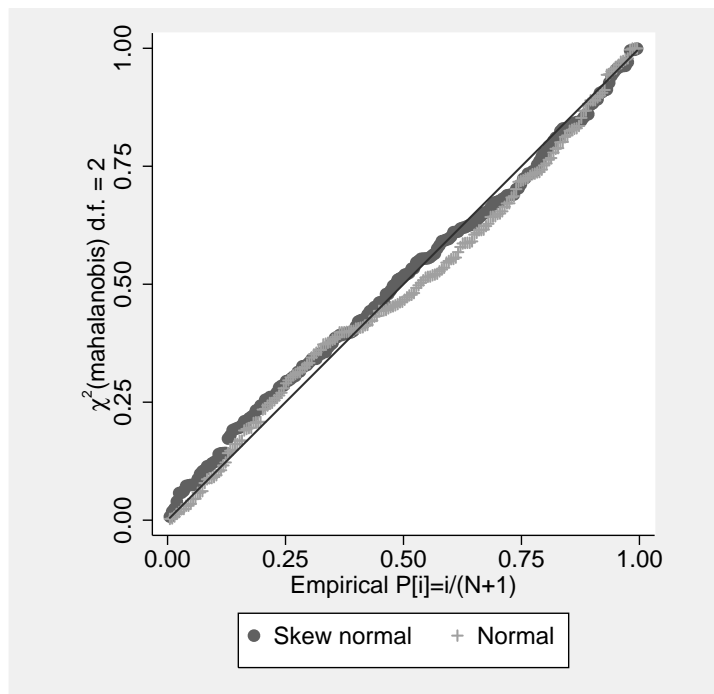


Figure 27. Probability plot for bivariate skew-normal and Normal model.

Table 21. Bivariate skew- t fit to lean body mass and body mass index.

```
. mskewtreg lbm bmi, nolog
```

Skew-t regression		Number of obs	=	202
		Wald chi2(0)	=	.
Log likelihood = -1213.1074		Prob > chi2	=	.

		Coef.	Std. Err.	z	P> z	[95% Conf. Interval]	
lbm	_cons	61.9651	1.926496	32.16	0.000	58.18923	65.74096
bmi	_cons	20.19786	.3165281	63.81	0.000	19.57748	20.81825
alpha	1	-2.234864	.5836008	-3.83	0.000	-3.378701	-1.091028
	2	5.242385	1.355909	3.87	0.000	2.584852	7.899918
Omega	11	171.7734	24.33631			130.1249	226.7522
	12	32.63322	8.546205			15.88297	49.38348
	22	14.8864	3.046903			9.967089	22.23366
	DF	51.00153	95.45759			1.301435	1998.684

LR test vs normal regression: chi2(3) = 37.86 Prob > chi2 = 0.0000

Table 22. Gender-specific bivariate skew- t fit to lean body mass and body mass index.

```

. mskewnreg lbm bmi = female, fullml nolog
Skew-normal regression
Log likelihood = -1105.0246
Number of obs = 202
Wald chi2(1) = 314.13
Prob > chi2 = 0.0000

```

		Coef.	Std. Err.	z	P> z	[95% Conf. Interval]	
lbm	female	-20.36519	1.149042	-17.72	0.000	-22.61728	-18.11311
	_cons	75.0455	.8219825	91.30	0.000	73.43444	76.65655
bmi	female	-2.267239	.3202247	-7.08	0.000	-2.894868	-1.639611
	_cons	24.13093	.2413966	99.96	0.000	23.6578	24.60406
gamma	1	.103742	.0543487	1.91	0.056	-.0027796	.2102635
	2	.6843177	.0915275	7.48	0.000	.5049271	.8637083
Sigma	11	71.51101	7.115746			57.5644	85.45761
	12	16.63505	2.002029			12.71114	20.55896
	22	6.954868	.7537917			5.477463	8.432272

```

LR test vs normal regression:      chi2(2) = 35.63  Prob > chi2 = 0.0000

```

E. Analysis of Automobile Prices

Consider the `auto.dta` dataset, distributed with Stata (see Section 1.2.2 in the User's Guide of the Stata 11 documentation), containing prices (in dollars) of 74 automobiles in 1978 (Table 23).

Suppose that we want to analyze the distribution of automobile prices. It is natural to compare the distribution to the Normal distribution first. Although we can already suspect that the normality assumption is implausible for these data from the output above — the distribution is not symmetric and, in fact, is right skewed — let us look at the Normal probability plot of `price` in Figure 28. Visual inspection of Figure 28 confirms our earlier observation — the distribution of prices is skewed to the right and also seems to have heavier tails than the Normal distribution.

Table 23. Summary statistics of automobile prices.

```
. sysuse auto, clear
(1978 Automobile Data)
. summarize price, detail
```

		Price			
	Percentiles	Smallest			
1%	3291	3291			
5%	3748	3299			
10%	3895	3667	Obs	74	
25%	4195	3748	Sum of Wgt.	74	
50%	5006.5		Mean	6165.257	
		Largest	Std. Dev.	2949.496	
75%	6342	13466			
90%	11385	13594	Variance	8699526	
95%	13466	14500	Skewness	1.653434	
99%	15906	15906	Kurtosis	4.819188	

```
. pnorm price
```

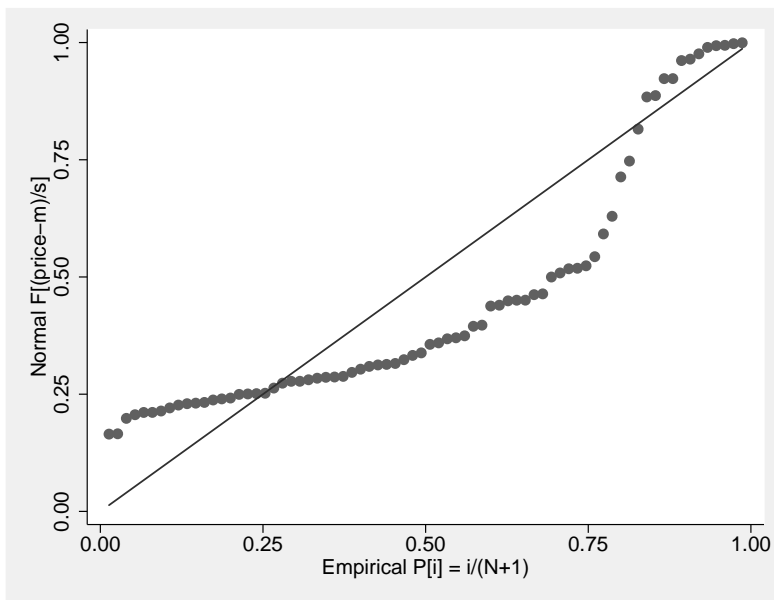


Figure 28. Normal probability plot of automobile prices.

Let us fit the skew- t distribution to the automobile prices and inspect its fit. As before, we use the `skewtreg` command to fit a skew- t distribution to `price`. From the output in Table 24, the estimate of 6.66 with 95% confidence interval of (0.36, 12.96) of the shape parameter α supports our claim that the distribution is skewed

to the right. The estimated degrees of freedom of 1.53 with 95% confidence interval of (0.86, 2.74) provides strong evidence for heavy tails of the distribution of prices.

Table 24. Skew- t model for automobile prices.

```
. skewtreg price, nolog
Skew-t regression           Number of obs =       74
                           Wald chi2(0) =       .
Log likelihood = -661.43121   Prob > chi2 =       .
```

price	Coef.	Std. Err.	z	P> z	[95% Conf. Interval]	
_cons	3702.711	117.3818	31.54	0.000	3472.647	3932.775
alpha	6.657068	3.213596	2.07	0.038	.3585365	12.9556
Omega	2502318	1168437			212222.5	4792413
DF	1.531112	.4548823			.8553057	2.740898

```
LR test vs normal regression:      chi2(2) =    68.56   Prob > chi2 = 0.0000
```

The reported results correspond to the estimates of the parameters of the skew- t distribution and, as we mentioned earlier, do not correspond to the actual moments of the distribution. For example, the estimate of the mean of the fitted skew- t distribution is not $\hat{\xi} = 3702.7$ but a function of all estimated parameters (Table 25).

Table 25. Estimated moments of the skew- t fit to automobile prices.

```
. estat summarize
note: moments missing because estimated DF <= 2
```

	Moments
mean	6784.941
variance	.
skewness	.

We used `estat summarize` after `skewtreg` to obtain the estimates of the three central moments: mean, variance, and skewness. The mean estimate is 6784.941, which is somewhat larger than the sample mean estimate of 6165.257 obtained earlier. The variance and skewness estimates cannot be computed in this example because the

estimated degrees of freedom are less than 2. The variance of the skew- t distribution is defined for $\nu > 2$ and the skewness is defined for $\nu > 3$.

We can use the `skewrplot` command to obtain a probability plot (Figure 29) to inspect how well the skew- t distribution fits the data.

```
. skewrplot, pp
```

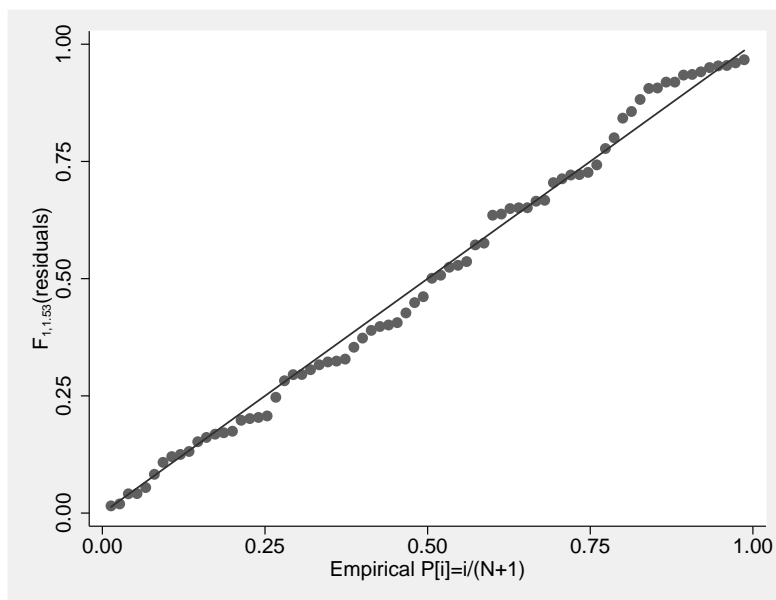


Figure 29. Skew- t probability plot of automobile prices.

Figure 29 plots empirical probabilities of the residuals from the skew- t fit above against the probabilities of the skew- t distribution with parameters $\xi = 0$, $\hat{\omega}^2 = 2502318$, $\hat{\alpha} = 6.66$, and $\hat{\nu} = 1.53$. According to this plot, the skew- t distribution fits the automobile price data very well.

F. Log-Skew-Normal and Log-Skew- t Distributions

The log-normal and log- t distributions are often used to model data with a nonnegative support such as precipitation data or income data. These distributions imply

that the distribution of the data in the log metric is symmetric. This assumption may be too restrictive in some applications. For example, here we investigate how reasonable this assumption is in the analysis of the monthly U.S. national precipitation data as described in Marchenko and Genton (2010*b*). The data are publicly available from the National Climatic Data Center (NCDC), the largest archive of weather data, and include monthly precipitation measured in inches to hundreds for the period of 1895-2007 (113 observations per month).

To fit the log-skew-normal model to the precipitation data, we follow the standard procedure and fit the skew-normal model, described previously, to the log of the precipitation. For example, we generate the new variable `lnprecip` to contain the log of the precipitation and fit the skew-normal distribution to the January (`month==1`) log-precipitation measurements over 113 years (Table 26).

Table 26. Log-skew-normal fit to U.S. national precipitation data in January.

```

. use precip07_national
(Precipitation (inches), national US data)
. gen lnprecip = ln(precip)
. skewnreg lnprecip if month==1, nolog
Skew-normal regression
Log likelihood = .71065091
Number of obs = 113
Wald chi2(0) = .
Prob > chi2 = .

```

lnprecip	Coef.	Std. Err.	z	P> z	[95% Conf. Interval]	
_cons	.7651154	.0228328	33.51	0.000	.7203639	.8098669
gamma	-.3321966	.1894123	-1.75	0.079	-.703438	.0390447
sigma2	.058959	.0081885			.0449088	.0774049

```

LR test vs normal regression:      chi2(1) = 2.96   Prob > chi2 = 0.0853

```

The skewness index is not significantly different from zero at a 5% level so the assumption of normality seems reasonable for January log precipitation.

More generally, we can obtain skewness indexes for all months. As depicted in

estimated degrees of freedom from `skewtreg`. If average precipitation is of interest, it can be computed as described in Marchenko and Genton (2010*b*) using the expressions for the moments of the log-skew-normal distribution.

G. Conclusion

In this chapter we described two flexible parametric models, the skew-normal and skew- t models, which can be used for the analysis of non-normal data. We presented a suite of commands for fitting these models in Stata to univariate and multivariate data. We also provided postestimation features to obtain linear predictions and to graphically evaluate the goodness of fit of the skewed distributions to the data. We demonstrated how to use the commands for univariate analysis of automobile-prices data and univariate and multivariate analyses of the well-known Australian Institute of Sport data. We also showed how to use the developed commands to analyze data with nonnegative support on the example of U.S. precipitation data.

CHAPTER V

SUMMARY

We presented novel approaches for analysis and modeling using multivariate skew-normal and skew- t distributions in econometrics and environmetrics.

Specifically, we established the link between sample-selection models and extended skewed distributions and used this link to introduce and study properties of the selection- t model, a robust version of the selection-normal model. We provided a numerical algorithm for estimating the model parameters and studied finite-sample performance of MLEs under various simulation scenarios. We demonstrated that with heavy-tailed data, MLEs from the selection-normal model are biased, and the asymptotic tests, such as Wald and likelihood-ratio tests, do not maintain the correct nominal level unlike MLEs from the proposed selection- t model. We applied the proposed selection- t model to analyze ambulatory expenditures and obtained statistical evidence of the existence of sample selection in these data which was not detected by the conventional selection-normal model.

We introduced a family of multivariate log-skew-elliptical distributions and studied their formal properties. We provided a number of stochastic representations of log-skew-elliptical random vectors and expressions for their marginal and conditional distributions. We also provided expressions for the moments of log-skew-normal random vectors and showed that positive moments of log-skew- t random vectors do not exist. We used log-skew-normal and log-skew- t distributions to model univariate and multivariate U.S. precipitation data.

We implemented four user-written estimation commands in Stata for fitting univariate and multivariate skew-normal and skew- t regression models: `skewnreg`, `skewtreg`, `mskewnreg`, and `mskewtreg`. We also implemented two postestimation

commands: `predict`, for obtaining predictions, and `skewrplot`, for graphical evaluation of the goodness-of-fit. As a tutorial, we demonstrated detailed use of the commands for the analyses of the well-known Australian Institute of Sport data, automobile price data, and U.S. precipitation data.

For the future, it would be interesting to investigate the sensitivity of the proposed selection- t model to misspecification of the error distribution. Also, one may look into extending the proposed selection- t model to the skew-selection- t model, where the errors are modeled according to the multivariate skew- t distribution. Such extension is appealing from a practical standpoint because it allows a more flexible model for the entire underlying population and not only the selected sample. More applications of the multivariate log-skew-normal and log-skew- t distributions to environmetrics as well as to other areas are encouraged.

REFERENCES

- Ahn, H., and Powell, J. L. (1993), “Semiparametric Estimation of Censored Selection Models with a Nonparametric Selection Mechanism,” *Journal of Econometrics*, 58, 3–29.
- Amemiya, T. (1985), *Advanced Econometrics*, Cambridge, MA: Harvard University Press.
- Arellano-Valle, R. B., and Azzalini, A. (2006), “On the Unification of Families of Skew-Normal Distributions,” *Scandinavian Journal of Statistics*, 33, 561–574.
- Arellano-Valle, R. B., and Azzalini, A. (2008), “The Centred Parametrization for the Multivariate Skew-Normal Distribution,” *Journal of Multivariate Analysis*, 99, 1362–1382.
- Arellano-Valle, R. B., Branco, M. D., and Genton, M. G. (2006), “A Unified View on Skewed Distributions Arising from Selections,” *The Canadian Journal of Statistics*, 34, 581–601.
- Arellano-Valle, R. B., and Genton, M. G. (2007), “On the Exact Distribution of Linear Combinations of Order Statistics from Dependent Random Variables,” *Journal of Multivariate Analysis*, 98, 1876–1894.
- Arellano-Valle, R. B., and Genton, M. G. (2010a), “Multivariate Extended Skew- t Distributions and Related Families,” *Metron*, to appear.
- Arellano-Valle, R. B., and Genton, M. G. (2010b), “Multivariate Unified Skew-Elliptical Distributions,” *Chilean Journal of Statistics*, 1, 17–33.

- Azzalini, A. (1985), “A Class of Distributions which Includes the Normal Ones,” *Scandinavian Journal of Statistics*, 12, 171–178.
- Azzalini, A. (2005), “The Skew-Normal Distribution and Related Multivariate Families (with discussion by Marc G. Genton and a rejoinder by the author),” *Scandinavian Journal of Statistics*, 32, 159–200.
- Azzalini, A. (2006), sn: *The Skew-Normal and Skew-t Distributions*, R Package Version 0.4-6.
- Azzalini, A., and Capitanio, A. (1999), “Statistical Applications of the Multivariate Skew-Normal Distribution,” *Journal of the Royal Statistical Society, Ser. B*, 61, 579–602.
- Azzalini, A., and Capitanio, A. (2003), “Distributions Generated by Perturbation of Symmetry with Emphasis on a Multivariate Skew- t Distribution,” *Journal of the Royal Statistical Society, Ser. B*, 65, 367–389.
- Azzalini, A., dal Cappello, T., and Kotz, S. (2003), “Log-Skew-Normal and Log-Skew- t Distributions as Models for Family Income Data,” *Journal of Income Distribution*, 11, 12–20.
- Azzalini, A., and Dalla Valle, A. (1996), “The Multivariate Skew-Normal Distribution,” *Biometrika*, 83, 715–726.
- Azzalini, A., and Genton, M. G. (2008), “Robust Likelihood Methods Based on the Skew- t and Related Distributions,” *International Statistical Review*, 76, 106–129.
- Bayes, C. L., and Branco, M. D. (2007), “Bayesian Inference for the Skewness Parameter of the Scalar Skew-Normal Distribution,” *Brazilian Journal of Probability and Statistics*, 21, 141–163.

- Beirlant, J., Goegebeur, Y., Segers, J., and Teugels, J. (2004), *Statistics of Extremes: Theory and Applications*, Chichester, England: John Wiley & Sons.
- Branco, M. D., and Dey, D. K. (2001), “A General Class of Multivariate Skew-Elliptical Distributions,” *Journal of Multivariate Analysis*, 79, 99–113.
- Cameron, A. C., and Trivedi, P. K. (2010), *Microeconometrics using Stata*, (Revised ed.), College Station, TX: Stata Press.
- Capitanio, A., Azzalini, A., and Stanghellini, E. (2003), “Graphical Models for Skew-Normal Variates,” *Scandinavian Journal of Statistics*, 30, 129–144.
- Chai, H. S., and Baily, K. R. (2008), “Use of Log-Skew-Normal Distribution in Analysis of Continuous Data with a Discrete Component at Zero,” *Statistics in Medicine*, 27, 3643–3655.
- Chakraborty, B. (2001), “On Affine Equivariant Multivariate Quantiles,” *Annals of the Institute of Statistical Mathematics*, 53, 380–403.
- Chaudhuri, P. (1996), “On a Geometric Notion of Quantiles for Multivariate Data,” *Journal of the American Statistical Association*, 91, 862–872.
- Chib, S., and Hamilton, B. H. (2000), “Bayesian Analysis of Cross-Section and Clustered Data Treatment Models,” *Journal of Econometrics*, 97, 25–50.
- Cook, R. D., and Weisberg, S. (1994), *An Introduction to Regression Graphics*, New York: Wiley.
- Copas, J. B., and Li, H. G. (1997), “Inference for Non-Random Samples,” *Journal of the Royal Statistical Society, Ser. B*, 59, 55–95.

- Crow, E. L., and Shimizu, K. (1988), *Lognormal Distributions: Theory and Applications*, New York: Marcel Dekker.
- Dalla Valle, A. (2007), “A Test for the Hypothesis of Skew-Normality in a Population,” *Journal of Statistical Computation and Simulation*, 77, 63–77.
- Das, M., Newey, W. K., and Vella, F. (2003), “Nonparametric Estimation of Sample Selection Models,” *Review of Economics Studies*, 70, 33–58.
- DiCiccio, T., and Monti, A. (2009), “Inferential Aspects of the Skew- t Distribution,” Manuscript in preparation.
- Fang, K.-T., Kotz, S., and Ng, K. W. (1990), *Symmetric Multivariate and Related Distributions*, New York: Chapman & Hall.
- Genton, M. G. (2004), *Skew-Elliptical Distributions and Their Applications: A Journey Beyond Normality*, (Edited vol.), Boca Raton, FL: Chapman & Hall / CRC.
- Greene, W. H. (2008), *Econometric Analysis*, (6th ed.), Upper Saddle River, NJ: Prentice-Hall.
- Hallin, M., and Ley, C. (2010), “Skew-Symmetric Distributions and Fisher Information — A Tale of Two Densities,” Working paper ECARES # 2010-014, Université Libre de Bruxelles.
- Heckman, J. (1974), “Shadow Prices, Market Wages, and Labor Supply,” *Econometrica*, 42, 679–694.
- Heckman, J. (1979), “Sample Selection Bias as a Specification Error,” *Econometrica*, 47, 153–161.

- Lange, K. L., Little, R. J. A., and Taylor, J. M. G. (1989), “Robust Statistical Modeling Using the t Distribution,” *Journal of the American Statistical Association*, 84, 881–896.
- Lee, L. F. (1982), “Some Approaches to the Correction of Selectivity Bias,” *Review of Economic Studies*, 49, 355–372.
- Lee, L. F. (1983), “Generalized Econometric Models with Selectivity,” *Econometrica*, 51, 507–512.
- Ley, C., and Paindaveine, D. (2010), “On the Singularity of Multivariate Skew-Symmetric Models,” *Journal of Multivariate Analysis*, to appear.
- Li, Q., and Racine, J. S. (2007), *Nonparametric Econometrics*, Princeton, NJ: Princeton University Press.
- Liseo, B., and Loperfido, N. (2006), “A Note on Reference Priors for the Scalar Skew-Normal Distribution,” *Journal of Statistical Planning and Inference*, 136, 373–389.
- Little, R. J. A., and Rubin, D. B. (2002), *Statistical Analysis with Missing Data*, (2nd ed.), Hoboken, NJ: Wiley.
- Marchenko, Y. V., and Genton, M. G. (2010a), “A Suite of Commands for Fitting the Skew-Normal and Skew- t Models,” *Stata Journal*, to appear.
- Marchenko, Y. V., and Genton, M. G. (2010b), “Multivariate Log-Skew-Elliptical Distributions with Applications to Precipitation Data,” *Environmetrics*, 21, 318–340.
- Mardia, K. V. (1970), “Measures of Multivariate Skewness and Kurtosis with Applications,” *Biometrika*, 36, 519–530.

- Muñoz-Gajardo, K. A., Arellano-Valle, R. B., and González-Farías, G. (2010), “The Student- t Censored Regression,” Manuscript in preparation.
- Newey, W. K. (1999), “Two-Step Series Estimation of Sample Selection Models,” Working paper # 99-04, Massachusetts Institute of Technology, Cambridge, MA.
- Pewsey, A. (2000), “Problems of Inference for Azzalini’s Skew-Normal Distribution,” *Journal of Applied Statistics*, 27, 859–770.
- Pewsey, A. (2006), “Some Observations on a Simple Means of Generating Skew Distributions” in *Advances in Distribution Theory, Order Statistics, and Inference*, N. Balakrishnan, E. Castillo, J.M. Sarabia (Eds.), Boston, MA: Statistics for Industry and Technology, Birkhauser, 75–84.
- Puhani, P. A. (2000), “The Heckman Correction for Sample Selection and its Critique,” *Journal of Economic Surveys*, 14, 53–68.
- R Development Core Team (2008), *R: A Language and Environment for Statistical Computing*, R Foundation for Statistical Computing, Vienna, Austria, ISBN 3-900051-07-0.
- URL:** <http://www.R-project.org>
- Rubin, D. B. (1976), “Inference and Missing Data,” *Biometrika*, 63, 581–592.
- 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.
- Scruggs, J. T. (2007), “Estimating the Cross-Sectional Market Response to an Endogenous Event: Naked vs. Underwritten Calls of Convertible Bonds,” *Journal of Empirical Finance*, 14, 220–247.

- StataCorp (2009), *Stata Statistical Software: Release 11*, College Station, TX: StataCorp LP.
- Vella, F. (1998), “Estimating Models with Sample Selection Bias: A Survey,” *Journal of Human Resources*, 33, 127–172.
- Wilks, D. S. (2006), *Statistical Methods in the Atmospheric Sciences*, (2nd ed.), New York: Academic Press.
- Wilson, P. S., and Toumi, R. (2005), “A Fundamental Probability Distribution for Heavy Rainfall,” *Geophysical Research Letters*, 32, L14812, doi:10.1029/2005GL022465.

APPENDIX A

A1. Score equations

Let $\boldsymbol{\theta} = (\boldsymbol{\beta}^\top, \boldsymbol{\gamma}^\top, \rho, \sigma, \nu)$. The log-likelihood for the selection- t model based on a single pair of observations (y, u) is

$$l(\boldsymbol{\theta}; y, u) = u \ln t(y; \mathbf{x}^\top \boldsymbol{\beta}, \sigma^2, \nu) + u \ln T(\eta_\theta; \nu + 1) + (1 - u) \ln T(-\mathbf{w}^\top \boldsymbol{\gamma}; \nu), \quad (\text{A.1})$$

where $u = 1$ if y is observed and $u = 0$ if y is unobserved, $z = (y - \mathbf{x}^\top \boldsymbol{\beta})/\sigma$,

$$\ln t(y; \mathbf{x}^\top \boldsymbol{\beta}, \sigma^2, \nu) = c(\nu, \sigma) - \frac{\nu + 1}{2} \ln \left(1 + \frac{z^2}{\nu} \right),$$

$$\eta_\theta = \sqrt{\frac{\nu + 1}{\nu + z^2}} \frac{\rho z + \mathbf{w}^\top \boldsymbol{\gamma}}{\sqrt{1 - \rho^2}},$$

and

$$c(\nu, \sigma) = \ln \Gamma \left(\frac{\nu + 1}{2} \right) - \ln \Gamma \left(\frac{\nu}{2} \right) - 0.5 \ln \pi - 0.5 \ln \nu - \ln \sigma.$$

Let

$$Q_\nu = \sqrt{\frac{\nu + 1}{\nu + \left(\frac{y - \mathbf{x}^\top \boldsymbol{\beta}}{\sigma} \right)^2}} = \sqrt{\frac{\nu + 1}{\nu + z^2}},$$

$$\zeta_{\theta-\nu} = \frac{\frac{\rho(y - \mathbf{x}^\top \boldsymbol{\beta})}{\sigma} + \mathbf{w}^\top \boldsymbol{\gamma}}{\sqrt{1 - \rho^2}} = A_{\rho\rho} z + A_\rho \mathbf{w}^\top \boldsymbol{\gamma},$$

$$\eta_\theta = \sqrt{\frac{\nu + 1}{\nu + \left(\frac{x_i - \mathbf{x}^\top \boldsymbol{\beta}}{\sigma} \right)^2}} \frac{\frac{\rho(y - \mathbf{x}^\top \boldsymbol{\beta})}{\sigma} + \mathbf{w}^\top \boldsymbol{\gamma}}{\sqrt{1 - \rho^2}} = Q_\nu \zeta_{\theta-\nu} = Q_\nu (A_{\rho\rho} z + A_\rho \mathbf{w}^\top \boldsymbol{\gamma}),$$

$$M_\nu(x) = \frac{t(x; \nu)}{T(x; \nu)} = \frac{\partial \ln T(x; \nu)}{\partial x},$$

where $A_\rho = 1/\sqrt{1 - \rho^2}$ and $A_{\rho\rho} = \rho/\sqrt{1 - \rho^2} = \rho A_\rho$.

Let $S_\alpha = \frac{\partial l(\boldsymbol{\theta})}{\partial \alpha}$ be the score function corresponding to the parameter α . For

$$\alpha \in \{\boldsymbol{\beta}^\top, \boldsymbol{\gamma}^\top, \rho, \sigma\},$$

$$\begin{aligned} S_\alpha &= u \frac{\partial \ln t(y; \mathbf{x}^\top \boldsymbol{\beta}, \sigma^2, \nu)}{\partial \alpha} + u \frac{\partial \ln T(\eta_\theta; \nu + 1)}{\partial \eta_\theta} \frac{\partial \eta_\theta}{\partial \alpha} - (1 - u) \frac{\partial \ln T(-\mathbf{w}^\top \boldsymbol{\gamma}; \nu)}{\partial (-\mathbf{w}^\top \boldsymbol{\gamma})} \frac{\partial \mathbf{w}^\top \boldsymbol{\gamma}}{\partial \alpha} \\ &= -u I(\alpha = \sigma) \frac{1}{\sigma} - u Q_\nu^2 z \frac{\partial z}{\partial \alpha} + u M_{\nu+1}(\eta_\theta) \frac{\partial \eta_\theta}{\partial \alpha} - (1 - u) I(\alpha = \gamma_k) w_k M_\nu(-\mathbf{w}^\top \boldsymbol{\gamma}). \end{aligned}$$

In particular, the scores are

$$\begin{aligned} S_{\beta_k} &= \frac{u x_k Q_\nu}{\sigma} [Q_\nu z + \{\zeta_{\theta-\nu}(\nu + z^2)^{-1} z - A_{\rho\rho}\} M_{\nu+1}(\eta_\theta)], \quad k = 1, \dots, p \\ S_{\gamma_k} &= u w_k A_\rho Q_\nu M_{\nu+1}(\eta_\theta) - (1 - u) w_k M_\nu(-\mathbf{w}^\top \boldsymbol{\gamma}), \quad k = 1, \dots, q \\ S_\rho &= u Q_\nu A_\rho^3 (z + \rho \mathbf{w}^\top \boldsymbol{\gamma}) M_{\nu+1}(\eta_\theta) \\ S_\sigma &= \frac{u}{\sigma} [-1 + Q_\nu^2 z^2 + Q_\nu z \{\zeta_{\theta-\nu}(\nu + z^2)^{-1} z - A_{\rho\rho}\} M_{\nu+1}(\eta_\theta)] \\ S_\nu &= u \left\{ \frac{\partial c(\nu, \sigma)}{\partial \nu} - \frac{1}{2} \ln \left(1 + \frac{z^2}{\nu} \right) + \frac{Q_\nu^2 z^2}{2\nu} + \frac{\partial \ln T(\eta_\theta; \nu + 1)}{\partial \nu} \right\} \\ &\quad + (1 - u_i) \frac{\partial \ln T(-\mathbf{w}^\top \boldsymbol{\gamma}; \nu)}{\partial \nu}, \end{aligned}$$

where $\frac{\partial c(\nu, \sigma)}{\partial \nu} = \frac{1}{2} \psi\left(\frac{\nu+1}{2}\right) - \frac{1}{2} \psi\left(\frac{\nu}{2}\right) - \frac{1}{2\nu}$, $\psi(\cdot)$ is the derivative of the log-gamma function, and $\frac{\partial \ln T(x; \nu)}{\partial \nu}$ must be computed numerically.

A2. Hessian matrix

Let $\mathbf{s}_i(\hat{\boldsymbol{\theta}})$ be the score vector of the selection- t model from A1 for observation i for $i = 1, \dots, N$, evaluated at the MLE $\hat{\boldsymbol{\theta}}$. Then, under appropriate regularity conditions, the Hessian matrix \mathbf{H} can be approximated using $\mathbf{H} = \frac{1}{N} \sum_{i=1}^N \mathbf{s}_i(\hat{\boldsymbol{\theta}}) \mathbf{s}_i(\hat{\boldsymbol{\theta}})^\top$.

We also provide direct computation of the Hessian matrix below. Let $S_{\alpha_1 \alpha_2} = \frac{\partial^2 l(\boldsymbol{\theta})}{\partial \alpha_1 \partial \alpha_2}$ be the second-order partial derivative of $l(\boldsymbol{\theta})$ from (A.1) with respect to α_1 and α_2 . The lower diagonal entries of the Hessian matrix are

$$\begin{aligned}
S_{\beta_k \beta_l} &= \frac{ux_k x_l Q_\nu^2}{\sigma^2} \left[\left\{ 2z^2(\nu + z^2)^{-1} - 1 \right\} + \left\{ \zeta_{\theta-\nu} \frac{z}{\nu + z^2} - A_{\rho\rho} \right\}^2 M'_{\nu+1}(\eta_\theta) \right] \\
&\quad + \frac{ux_k x_l}{\sigma^2} Q_\nu (\nu + z^2)^{-1} \left[\zeta_{\theta-\nu} \left\{ 3(\nu + z^2)^{-1} z^2 - 1 \right\} - 2A_{\rho\rho} z \right] M_{\nu+1}(\eta_\theta) \\
S_{\beta_k \gamma_m} &= u \frac{x_k w_n Q_\nu A_\rho}{\sigma} \left[(\nu + z^2)^{-1} z \left\{ M_{\nu+1}(\eta_\theta) + \zeta_{\theta-\nu} M'_{\nu+1}(\eta_\theta) \right\} - A_{\rho\rho} M'_{\nu+1}(\eta_\theta) \right] \\
S_{\beta_k \rho} &= u \frac{x_k Q_\nu A_\rho^3}{\sigma} \frac{z(z + \rho \mathbf{w}^\top \boldsymbol{\gamma})}{\nu + z^2} \left\{ M_{\nu+1}(\eta_\theta) + \eta_\theta M'_{\nu+1}(\eta_\theta) \right\} \\
&\quad - u \frac{x_k Q_\nu A_\rho^3}{\sigma} \left[Q_\nu (z + \rho \mathbf{w}^\top \boldsymbol{\gamma}) A_{\rho\rho} M'_{\nu+1}(\eta_\theta) + M_{\nu+1}(\eta_\theta) \right] \\
S_{\beta_k \sigma} &= u \frac{2x_k Q_\nu^2 z}{\sigma^2} \left(\frac{z^2}{\nu + z^2} - 1 \right) + u \frac{x_k z M'_{\nu+1}(\eta_\theta)}{\sigma^2} \left(\eta_\theta \frac{z}{\nu + z^2} - Q_\nu A_{\rho\rho} \right)^2 \\
&\quad + u \frac{x_k M_{\nu+1}(\eta_\theta)}{\sigma^2} \left\{ \eta_\theta \frac{z}{\nu + z^2} \left(3 \frac{z^2}{\nu + z^2} - 2 \right) - 2Q_\nu A_{\rho\rho} \frac{z^2}{\nu + z^2} + Q_\nu A_{\rho\rho} \right\} \\
S_{\beta_k \nu} &= \frac{ux_k Q_\nu^2 z}{\sigma} \left(\frac{1}{\nu + 1} - \frac{1}{\nu + z^2} \right) + \frac{ux_k Q_\nu}{\sigma} \left(\zeta_{\theta-\nu} \frac{z}{\nu + z^2} - A_{\rho\rho} \right) \\
&\quad \times \left\{ \frac{\partial M_{\nu+1}(\eta_\theta)}{\partial \nu} + \frac{1}{2} M_{\nu+1}(\eta_\theta) \left(\frac{1}{\nu + 1} - \frac{1}{\nu + z^2} \right) \right\} \\
&\quad - \frac{ux_k}{\sigma} M_{\nu+1}(\eta_\theta) \eta_\theta \frac{z}{(\nu + z^2)^2} \\
S_{\gamma_m \gamma_n} &= uw_m w_n Q_\nu^2 A_\rho^2 M'_{\nu+1}(\eta_\theta) + (1 - u) w_m w_n M'_{\nu+1}(-\mathbf{w}^\top \boldsymbol{\gamma}) \\
S_{\gamma_m \rho} &= uw_m Q_\nu A_\rho^3 \left\{ \rho M_{\nu+1}(\eta_\theta) + Q_\nu A_\rho (z + \rho \mathbf{w}^\top \boldsymbol{\gamma}) M'_{\nu+1}(\eta_\theta) \right\} \\
S_{\gamma_m \sigma} &= \frac{uw_m A_\rho Q_\nu z}{\sigma} \left[\frac{z}{\nu + z^2} \left\{ \eta_\theta M'_{\nu+1}(\eta_\theta) + M_{\nu+1}(\eta_\theta) \right\} - Q_\nu A_{\rho\rho} M'_{\nu+1}(\eta_\theta) \right] \\
S_{\gamma_m \nu} &= uw_m A_\rho Q_\nu \left\{ \frac{\partial M_{\nu+1}(\eta_\theta)}{\partial \nu} + \frac{1}{2} \left(\frac{1}{\nu + 1} - \frac{1}{\nu + z^2} \right) \right\} \\
&\quad - (1 - u) w_m \frac{\partial M_{\nu+1}(-\mathbf{w}^\top \boldsymbol{\gamma})}{\partial \nu} \\
S_{\rho^2} &= u Q_\nu^2 A_\rho^6 (z + \rho \mathbf{w}^\top \boldsymbol{\gamma})^2 M'_{\nu+1}(\eta_\theta) + u Q_\nu A_\rho^5 \left\{ 3\rho z + (1 + 2\rho^2) \mathbf{w}^\top \boldsymbol{\gamma} \right\} M_{\nu+1}(\eta_\theta) \\
S_{\rho\sigma} &= \frac{uz Q_\nu A_\rho^3}{\sigma} (z + \rho \mathbf{w}^\top \boldsymbol{\gamma}) \left[\frac{z}{\nu + z^2} \left\{ \eta_\theta M'_{\nu+1}(\eta_\theta) + M_{\nu+1}(\eta_\theta) \right\} - Q_\nu A_{\rho\rho} M'_{\nu+1}(\eta_\theta) \right] \\
&\quad - \frac{uz Q_\nu A_\rho^3}{\sigma} M_{\nu+1} \\
S_{\rho\nu} &= u Q_\nu A_\rho^3 (z + \rho \mathbf{w}^\top \boldsymbol{\gamma}) \left\{ \frac{\partial M_{\nu+1}(\eta_\theta)}{\partial \nu} + \frac{1}{2} \left(\frac{1}{\nu + 1} - \frac{1}{\nu + z^2} \right) M_{\nu+1}(\eta_\theta) \right\}
\end{aligned}$$

$$\begin{aligned}
S_{\sigma^2} &= \frac{u}{\sigma^2} - \frac{3uQ_\nu^2 z^2}{\sigma^2} + \frac{2uQ_\nu}{\sigma^2} \frac{z^2}{\nu+z^2} \\
&\quad \times \left[Q_\nu z^2 + \left\{ \frac{3}{2} \zeta_{\theta-\nu} \left(\frac{z^2}{\nu+z^2} - 1 \right) - Q_\nu A_{\rho\rho} \right\} M_{\nu+1}(\eta_\theta) \right] \\
&\quad + \frac{2uQ_\nu A_{\rho\rho} z}{\sigma^2} M_{\nu+1}(\eta_\theta) + \frac{uz^2}{\sigma^2} \left(\eta_\theta \frac{z}{\nu+z^2} - Q_\nu A_{\rho\rho} \right)^2 M'_{\nu+1}(\eta_\theta) \\
S_{\sigma\nu} &= \frac{uQ_\nu z^2}{\sigma} \left\{ Q_\nu \left(\frac{1}{\nu+1} - \frac{1}{\nu+z^2} \right) - \frac{M_{\nu+1}(\eta_\theta)}{(\nu+z^2)^2} \right\} \\
&\quad + \frac{uQ_\nu z}{\sigma} \left\{ \frac{\partial M_{\nu+1}(\eta_\theta)}{\partial \nu} + \frac{1}{2} \left(\frac{1}{\nu+1} - \frac{1}{\nu+z^2} \right) M_{\nu+1}(\eta_\theta) \right\} \\
S_{\nu^2} &= u \frac{\partial^2 c(\nu, \sigma)}{\partial^2 \nu} + \frac{u}{2\nu} \frac{z^2}{\nu+z^2} + \frac{uQ_\nu^2 z^2}{2\nu} \left(\frac{1}{\nu+1} - \frac{1}{\nu+z^2} \right) - \frac{uz^4}{2\nu^2} \\
&\quad + u \frac{\partial^2 \ln T(\eta_\theta; \nu+1)}{\partial^2 \nu} + (1-u) \frac{\partial^2 \ln T(-\mathbf{w}^\top \boldsymbol{\gamma}; \nu)}{\partial^2 \nu}
\end{aligned}$$

where $k, l = 1, \dots, p$ and $m, n = 1, \dots, q$, $M'_\nu(x) = \frac{\partial M_\nu(x)}{\partial x} = -M_\nu(x) \left\{ x \frac{\nu+1}{\nu+x^2} + M_\nu(x) \right\}$, and $\frac{\partial M_{\nu+1}(x)}{\partial \nu}$ must be computed numerically. Moreover

$$\frac{\partial^2 c(\nu, \sigma)}{\partial^2 \nu} = \frac{1}{4} \psi' \left(\frac{\nu+1}{2} \right) - \frac{1}{4} \psi' \left(\frac{\nu}{2} \right) + \frac{1}{2\nu^2},$$

where $\psi'(\cdot)$ is the trigamma function.

A3. Fisher and observed information matrices at $\rho = 0$ for the selection-normal model

When $\nu \rightarrow \infty$, $Q_\nu = 1$, $\eta_\theta = \zeta_{\theta-\nu} = A_{\rho\rho} z + A_\rho \mathbf{w}^\top \boldsymbol{\gamma}$, $M_\nu(x) = M(x) = \frac{\phi(x)}{\Phi(x)} = \frac{\partial \ln \Phi(x)}{\partial x}$, $M'_\nu(x) = M'(x) = -M(x) \{x + M(x)\}$, $\frac{\partial c(\nu, \sigma)}{\partial \nu} = 0$, and $\frac{\partial T(x; \nu)}{\partial \nu} = 0$.

The scores are

$$\begin{aligned}
S_{\beta_k} &= \frac{ux_k}{\sigma} (z - A_{\rho\rho}M(\eta_\theta)), \quad k = 1, \dots, p \\
S_{\gamma_k} &= uw_k A_\rho M(\eta_\theta) - (1-u)w_k M(-\mathbf{w}^\top \boldsymbol{\gamma}), \quad k = 1, \dots, q \\
S_\rho &= uA_\rho^3 (z + \rho \mathbf{w}^\top \boldsymbol{\gamma}) M(\eta_\theta) \\
S_\sigma &= \frac{u}{\sigma} \{-1 + z^2 - zA_{\rho\rho}M(\eta_\theta)\} \\
S_\nu &= 0.
\end{aligned}$$

At $\rho = 0$, $A_\rho = 1$, $A_{\rho\rho} = 0$, $\eta_\theta = \mathbf{w}^\top \boldsymbol{\gamma}$ and the scores are

$$\begin{aligned}
S_{\beta_k} &= \frac{ux_k}{\sigma} z, \quad k = 1, \dots, p \\
S_{\gamma_k} &= uw_k M(\mathbf{w}^\top \boldsymbol{\gamma}) - (1-u)w_k M(-\mathbf{w}^\top \boldsymbol{\gamma}), \quad k = 1, \dots, q \\
S_\rho &= uz M(\mathbf{w}^\top \boldsymbol{\gamma}) \\
S_\sigma &= \frac{u}{\sigma} \{-1 + z^2\} \\
S_\nu &= 0.
\end{aligned}$$

W.l.g. let $p = q = 1$, then the observed information for $(\beta_0, \beta_1, \gamma_0, \gamma_1, \rho, \sigma)$ for the sample of size N at $\rho = 0$ is

$$\left(\begin{array}{cccccc}
\sum_i \frac{u_i}{\sigma^2} & \sum_i \frac{u_i x_{1i}}{\sigma^2} & 0 & 0 & \sum_i \frac{u_i M_i}{\sigma} & \sum_i \frac{2u_i z_i}{\sigma^2} \\
\sum_i \frac{u_i x_{1i}}{\sigma^2} & \sum_i \frac{u_i x_{1i}^2}{\sigma^2} & 0 & 0 & \sum_i \frac{u_i x_{1i} M_i}{\sigma} & \sum_i \frac{2u_i z_i x_{1i}}{\sigma^2} \\
0 & 0 & \sum_i B_i & \sum_i w_{1i} B_i & -\sum_i u_i z_i M_i' & 0 \\
0 & 0 & \sum_i w_{1i} B_i & \sum_i w_{1i}^2 B_i & -\sum_i u_i z_i w_{1i} M_i' & 0 \\
\sum_i \frac{u_i M_i}{\sigma} & \sum_i \frac{u_i x_{1i} M_i}{\sigma} & -\sum_i u_i z_i M_i' & -\sum_i u_i z_i w_{1i} M_i' & O_{5,5} & \sum_i \frac{u_i z_i M_i}{\sigma} \\
\sum_i \frac{2u_i z_i}{\sigma^2} & \sum_i \frac{2u_i z_i x_{1i}}{\sigma^2} & 0 & 0 & \sum_i \frac{u_i z_i M_i}{\sigma} & O_{6,6}
\end{array} \right),$$

where

$$\begin{aligned}
M_i &= M(\mathbf{w}_i^\top \boldsymbol{\gamma}) \\
M'_i &= M'(\mathbf{w}_i^\top \boldsymbol{\gamma}) \\
B_i &= B(u_i; \mathbf{w}_i^\top \boldsymbol{\gamma}) = -u_i M'(\mathbf{w}_i^\top \boldsymbol{\gamma}) - (1 - u_i) M'(-\mathbf{w}_i^\top \boldsymbol{\gamma}) \\
\mathbf{w}_i^\top \boldsymbol{\gamma} &= \gamma_0 + \gamma_1 w_{1i} \\
O_{5,5} &= -\sum_i u_i (z_i^2 M'_i + \mathbf{w}_i^\top \boldsymbol{\gamma} M_i) \\
O_{6,6} &= -\sum_i \frac{u_i}{\sigma^2} + \sum_i \frac{3u_i z_i^2}{\sigma^2}.
\end{aligned}$$

To compute the Fisher information, consider the following expectations:

$$\begin{aligned}
E(U_i) &= \Phi(\mathbf{w}_i^\top \boldsymbol{\gamma}) \\
E(U_i Z_i) &= E\{U_i E(Z_i | U_i)\} = E\left[U_i \left\{E\left(\frac{Y_i^* - \mathbf{x}_i^\top \boldsymbol{\beta}}{\sigma} \mid U_i = 1\right) + 0\right\}\right] \\
&= \rho \frac{\phi(\mathbf{w}_i^\top \boldsymbol{\gamma})}{\Phi(\mathbf{w}_i^\top \boldsymbol{\gamma})} E(U_i) = \rho \phi(\mathbf{w}_i^\top \boldsymbol{\gamma}) \Big|_{\rho=0} = 0 \\
E(U_i Z_i^2) &= E\{U_i E(Z_i^2 | U_i)\} = \left(1 - \rho^2 \mathbf{w}_i^\top \boldsymbol{\gamma} \frac{\phi(\mathbf{w}_i^\top \boldsymbol{\gamma})}{\Phi(\mathbf{w}_i^\top \boldsymbol{\gamma})}\right) E(U_i) \\
&= \Phi(\mathbf{w}_i^\top \boldsymbol{\gamma}) - \rho^2 \mathbf{w}_i^\top \boldsymbol{\gamma} \phi(\mathbf{w}_i^\top \boldsymbol{\gamma}) \Big|_{\rho=0} = \Phi(\mathbf{w}_i^\top \boldsymbol{\gamma}).
\end{aligned}$$

Note that

$$\begin{aligned}
E\{B(U_i; \mathbf{w}_i^\top \boldsymbol{\gamma})\} &= -M'(\mathbf{w}_i^\top \boldsymbol{\gamma}) E(U_i) - (1 - E(U_i)) M'(-\mathbf{w}_i^\top \boldsymbol{\gamma}) \\
&= -\Phi(\mathbf{w}_i^\top \boldsymbol{\gamma}) M'(\mathbf{w}_i^\top \boldsymbol{\gamma}) - \Phi(-\mathbf{w}_i^\top \boldsymbol{\gamma}) M'(-\mathbf{w}_i^\top \boldsymbol{\gamma}) \\
&= \frac{\phi^2(\mathbf{w}_i^\top \boldsymbol{\gamma})}{\Phi(\mathbf{w}_i^\top \boldsymbol{\gamma}) \Phi(-\mathbf{w}_i^\top \boldsymbol{\gamma})}
\end{aligned}$$

$$\begin{aligned}
E \{ -U_i Z_i^2 M'(\mathbf{w}_i^\top \boldsymbol{\gamma}) - U_i \mathbf{w}_i^\top \boldsymbol{\gamma} M(\mathbf{w}_i^\top \boldsymbol{\gamma}) \} |_{\rho=0} &= -\Phi(\mathbf{w}_i^\top \boldsymbol{\gamma}) M'(\mathbf{w}_i^\top \boldsymbol{\gamma}) \\
&\quad - \mathbf{w}_i^\top \boldsymbol{\gamma} \Phi(\mathbf{w}_i^\top \boldsymbol{\gamma}) M(\mathbf{w}_i^\top \boldsymbol{\gamma}) \\
&= \mathbf{w}_i^\top \boldsymbol{\gamma} \phi(\mathbf{w}_i^\top \boldsymbol{\gamma}) + \frac{\phi^2(\mathbf{w}_i^\top \boldsymbol{\gamma})}{\Phi(\mathbf{w}_i^\top \boldsymbol{\gamma})} \\
&\quad - \mathbf{w}_i^\top \boldsymbol{\gamma} \phi(\mathbf{w}_i^\top \boldsymbol{\gamma}) = \frac{\phi^2(\mathbf{w}_i^\top \boldsymbol{\gamma})}{\Phi(\mathbf{w}_i^\top \boldsymbol{\gamma})}.
\end{aligned}$$

Let $r_\phi(\mathbf{w}_i^\top \boldsymbol{\gamma}) = \frac{\phi^2(\mathbf{w}_i^\top \boldsymbol{\gamma})}{\Phi(\mathbf{w}_i^\top \boldsymbol{\gamma})\Phi(-\mathbf{w}_i^\top \boldsymbol{\gamma})}$. Then, the Fisher information matrix is $\mathbf{I} = \sum_i \mathbf{I}_i$,

where \mathbf{I}_i is

$$\begin{pmatrix}
\frac{\Phi(\mathbf{w}_i^\top \boldsymbol{\gamma})}{\sigma^2} & \frac{x_{1i}\Phi(\mathbf{w}_i^\top \boldsymbol{\gamma})}{\sigma^2} & 0 & 0 & \frac{\phi(\mathbf{w}_i^\top \boldsymbol{\gamma})}{\sigma} & 0 \\
\frac{x_{1i}\Phi(\mathbf{w}_i^\top \boldsymbol{\gamma})}{\sigma^2} & \frac{x_{1i}^2\Phi(\mathbf{w}_i^\top \boldsymbol{\gamma})}{\sigma^2} & 0 & 0 & \frac{x_{1i}\phi(\mathbf{w}_i^\top \boldsymbol{\gamma})}{\sigma} & 0 \\
0 & 0 & r_\phi(\mathbf{w}_i^\top \boldsymbol{\gamma}) & w_{1i}r_\phi(\mathbf{w}_i^\top \boldsymbol{\gamma}) & 0 & 0 \\
0 & 0 & w_{1i}r_\phi(\mathbf{w}_i^\top \boldsymbol{\gamma}) & w_{1i}^2r_\phi(\mathbf{w}_i^\top \boldsymbol{\gamma}) & 0 & 0 \\
\frac{\phi(\mathbf{w}_i^\top \boldsymbol{\gamma})}{\sigma} & \frac{x_{1i}\phi(\mathbf{w}_i^\top \boldsymbol{\gamma})}{\sigma} & 0 & 0 & \frac{\phi^2(\mathbf{w}_i^\top \boldsymbol{\gamma})}{\Phi(\mathbf{w}_i^\top \boldsymbol{\gamma})} & 0 \\
0 & 0 & 0 & 0 & 0 & \frac{2\Phi(\mathbf{w}_i^\top \boldsymbol{\gamma})}{\sigma^2}
\end{pmatrix}.$$

Unlike the case of the extended skew-normal model (Arellano-Valle and Genton 2010a, p. 17), the scores, corresponding to the selection-normal model, are not linearly dependent at $\rho = 0$. Also, the observed information and Fisher information for $\boldsymbol{\theta} = (\boldsymbol{\beta}, \boldsymbol{\gamma}, \rho, \sigma)$ are nonsingular at $\rho = 0$, which is not the case for the extended skew-normal model.

A4. Stata output for the numerical results

Heckman selection-normal model:

```
. heckman lny age female edu blhisp totchr ins, sel(dy=age female educ blhisp totchr ins income)
Heckman selection model                Number of obs    =    3328
(regression model with sample selection)  Censored obs     =    526
                                           Uncensored obs   =    2802
                                           Wald chi2(6)     =    288.88
                                           Prob > chi2      =    0.0000
```

Log likelihood = -5836.219

	Coef.	Std. Err.	z	P> z	[95% Conf. Interval]	
lny						
age	0.212	0.023	9.21	0.000	0.167	0.257
female	0.348	0.060	5.79	0.000	0.230	0.466
educ	0.019	0.011	1.77	0.076	-0.002	0.039
blhisp	-0.219	0.060	-3.66	0.000	-0.336	-0.102
totchr	0.540	0.039	13.73	0.000	0.463	0.617
ins	-0.030	0.051	-0.59	0.557	-0.130	0.070
_cons	5.044	0.228	22.11	0.000	4.597	5.491
dy						
age	0.088	0.027	3.21	0.001	0.034	0.142
female	0.663	0.061	10.87	0.000	0.543	0.782
educ	0.062	0.012	5.15	0.000	0.038	0.086
blhisp	-0.364	0.062	-5.88	0.000	-0.485	-0.243
totchr	0.797	0.071	11.20	0.000	0.658	0.936
ins	0.170	0.063	2.71	0.007	0.047	0.293
income	0.003	0.001	2.06	0.040	0.000	0.005
_cons	-0.676	0.194	-3.48	0.000	-1.056	-0.296
/athrho	-0.131	0.150	-0.88	0.380	-0.425	0.162
/lnsigma	0.240	0.014	16.59	0.000	0.211	0.268
rho	-0.131	0.147			-0.401	0.161
sigma	1.271	0.018			1.236	1.308
lambda	-0.166	0.188			-0.534	0.202

LR test of indep. eqns. (rho = 0): chi2(1) = 0.91 Prob > chi2 = 0.3406

Heckman selection-*t* model:

```
. heckt lny age female edu blhisp totchr ins, sel(dy=age female educ blhisp totchr ins income)
Heckman-t selection model                Number of obs    =    3328
(regression model with sample selection)  Censored obs     =    526
                                           Uncensored obs   =    2802
                                           Wald chi2(6)     =    325.58
                                           Prob > chi2      =    0.0000
```

Log likelihood = -5822.076

	Coef.	Std. Err.	z	P> z	[95% Conf. Interval]	
lny						
age	0.207	0.023	9.16	0.000	0.163	0.251
female	0.307	0.056	5.45	0.000	0.196	0.417
educ	0.017	0.010	1.69	0.091	-0.003	0.037
blhisp	-0.193	0.058	-3.35	0.001	-0.306	-0.080
totchr	0.513	0.036	14.36	0.000	0.443	0.583
ins	-0.052	0.050	-1.04	0.298	-0.151	0.046
_cons	5.206	0.209	24.93	0.000	4.797	5.615

dy						
age	0.099	0.030	3.31	0.001	0.040	0.157
female	0.725	0.069	10.58	0.000	0.591	0.859
educ	0.065	0.013	5.06	0.000	0.040	0.090
blhisp	-0.394	0.067	-5.92	0.000	-0.524	-0.263
totchr	0.890	0.087	10.21	0.000	0.719	1.061
ins	0.180	0.068	2.65	0.008	0.047	0.313
income	0.003	0.001	2.06	0.040	0.000	0.006
_cons	-0.748	0.208	-3.60	0.000	-1.155	-0.341
/athrho	-0.334	0.128	-2.61	0.009	-0.584	-0.083
/lnsigma	0.178	0.021	8.29	0.000	0.136	0.220
/lndf	2.559	0.221	11.61	0.000	2.127	2.992
rho	-0.322	0.115			-0.526	-0.083
sigma	1.195	0.026			1.146	1.246
df	12.928	2.851			8.391	19.917
lambda	-0.385	0.140			-0.659	-0.110

Wald test of indep. eqns. (rho = 0): chi2(1) = 6.82 Prob > chi2 = 0.0090

APPENDIX B

Proof of Proposition 1 (selection representation 1): To show that

$\mathbf{Z} \sim \text{LSE}_d(\mathbf{0}, \bar{\boldsymbol{\Omega}}, \boldsymbol{\alpha}, g^{(d+1)})$, we apply a property that if $\mathbf{Z} \stackrel{d}{=} (\mathbf{V} | \tilde{U}_0 < \boldsymbol{\alpha}^\top \mathbf{U})$, then $\psi(\mathbf{Z}) \stackrel{d}{=} (\psi(\mathbf{V}) | \tilde{U}_0 < \boldsymbol{\alpha}^\top \mathbf{U})$ with $\psi(\cdot) = \ln(\cdot)$. Then, the result follows from Definition 1 and a selection representation of a skew-elliptical random vector as given by, for example, Arellano-Valle and Genton (2010a). Then, $\mathbf{X} \sim \text{LSE}_d(\boldsymbol{\xi}, \boldsymbol{\Omega}, \boldsymbol{\alpha}, g^{(d+1)})$, since $\ln(\mathbf{X}) = \boldsymbol{\xi} + \boldsymbol{\omega} \ln(\mathbf{Z}) \sim \text{SE}_d(\boldsymbol{\xi}, \boldsymbol{\Omega}, \boldsymbol{\alpha}, g^{(d+1)})$. \square

Proof of Proposition 2 (selection representation 2): Although this result can also be shown using similar arguments as in the proof for the selection representation 1, we choose a different approach here. We use properties of elliptical and log-elliptical distributions (Fang et al. 1990, pp. 45, 56) to show that the density of \mathbf{Z} is of the form (3.4). It can be written as

$$f_{\mathbf{Z}}(\mathbf{z}) = f_{\mathbf{V}|U_0>0}(\mathbf{z}) = f_{\mathbf{V}}(\mathbf{z}) \frac{P(U_0 > 0 | \mathbf{V} = \mathbf{z})}{P(U_0 > 0)}. \quad (\text{B.1})$$

Using properties of elliptical distributions, $\mathbf{U} \sim \text{EC}_d(\mathbf{0}, \bar{\boldsymbol{\Omega}}, g^{(d)})$ and $U_0 | \mathbf{U} \sim \text{EC}(\boldsymbol{\delta}^\top \bar{\boldsymbol{\Omega}}^{-1} \mathbf{u}, 1 - \boldsymbol{\delta}^\top \bar{\boldsymbol{\Omega}} \boldsymbol{\delta}, g_{Q_{\mathbf{u}}^{\mathbf{0}, \bar{\boldsymbol{\Omega}}}})$, where $Q_{\mathbf{u}}^{\mathbf{0}, \bar{\boldsymbol{\Omega}}} = \mathbf{u}^\top \bar{\boldsymbol{\Omega}} \mathbf{u}$. Then $\mathbf{V} = \exp(\mathbf{U})$ follows a d -dimensional multivariate log-elliptical distribution with density

$$f_{\mathbf{V}}(\mathbf{z}) = \left(\prod_{i=1}^d z_i^{-1} \right) f_{\mathbf{U}}\{\ln(\mathbf{z})\} = \left(\prod_{i=1}^d z_i^{-1} \right) f_d\{\ln(\mathbf{z}); \mathbf{0}, \bar{\boldsymbol{\Omega}}, g^{(d)}\}.$$

Let $U_0^* = \{U_0 - \boldsymbol{\delta}^\top \bar{\boldsymbol{\Omega}}^{-1} \ln(\mathbf{z})\} / (1 - \boldsymbol{\delta}^\top \bar{\boldsymbol{\Omega}}^{-1} \boldsymbol{\delta})^{1/2}$, then $U_0^* \sim \text{EC}(0, 1, g_{Q_{\ln(\mathbf{z})}^{\mathbf{0}, \bar{\boldsymbol{\Omega}}}})$, where $Q_{\ln(\mathbf{z})}^{\mathbf{0}, \bar{\boldsymbol{\Omega}}} = \ln(\mathbf{z})^\top \bar{\boldsymbol{\Omega}} \ln(\mathbf{z})$. Hence

$$\begin{aligned} P(U_0 > 0 | \mathbf{V} = \mathbf{z}) &= P\{U_0 > 0 | \mathbf{U} = \ln(\mathbf{z})\} = P\{U_0^* < \boldsymbol{\alpha}^\top \ln(\mathbf{z}) | \mathbf{U} = \ln(\mathbf{z})\} \\ &= F\{\boldsymbol{\alpha}^\top \ln(\mathbf{z}); g_{Q_{\ln(\mathbf{z})}^{\mathbf{0}, \bar{\boldsymbol{\Omega}}}}\}. \end{aligned}$$

Since the distribution of U_0 is symmetric, $P(U_0 > 0) = 1/2$. Substituting the terms obtained above into (B.1), we arrive at the density of \mathbf{Z} ,

$$f_{\mathbf{Z}}(\mathbf{z}) = 2 \left(\prod_{i=1}^d z_i^{-1} \right) f_d\{\mathbf{ln}(\mathbf{z}); \mathbf{0}, \bar{\boldsymbol{\Omega}}, g^{(d)}\} F\{\boldsymbol{\alpha}^\top \mathbf{ln}(\mathbf{z}); g_{Q_{\mathbf{ln}(\mathbf{z})}^{\mathbf{0}, \bar{\boldsymbol{\Omega}}}}\},$$

which is a special case of (3.4) with $\boldsymbol{\xi} = \mathbf{0}$ and $\boldsymbol{\Omega} = \bar{\boldsymbol{\Omega}}$. Then, $\mathbf{X} \sim \text{LSE}_d(\boldsymbol{\xi}, \boldsymbol{\Omega}, \boldsymbol{\alpha}, g^{(d+1)})$, since $\mathbf{ln}(\mathbf{X}) = \boldsymbol{\xi} + \boldsymbol{\omega} \mathbf{ln}(\mathbf{Z}) \sim \text{SE}_d(\boldsymbol{\xi}, \boldsymbol{\Omega}, \boldsymbol{\alpha}, g^{(d+1)})$. \square

Proof of Proposition 3 (log-skew-normal mixture): This representation arises from the definition of a log-skew-elliptical random vector and skew-scale mixture representation of a skew-elliptical random vector: if $\mathbf{X} \sim \text{SN}_d(\mathbf{0}, \boldsymbol{\Omega}, \boldsymbol{\alpha})$, then $\mathbf{Y} \stackrel{d}{=} \boldsymbol{\xi} + K(\eta)^{1/2} \mathbf{X}$ has the density of the form (3.7), $\mathbf{Y} \sim \text{SMSN}_d\{\boldsymbol{\xi}, \boldsymbol{\Omega}, \boldsymbol{\alpha}, K(\eta), H(\eta)\}$; see Azzalini and Capitanio (2003), among others. Since $\mathbf{X} = \exp(\boldsymbol{\xi}) \mathbf{Z}^{K^{1/2}(\eta)} = \exp\{\boldsymbol{\xi} + K^{1/2}(\eta) \mathbf{ln}(\mathbf{Z})\} = \exp(\mathbf{Y})$, the result follows from Definition 1. \square

Proof of Proposition 4 (marginal distribution): By definition, we have that the vector $\mathbf{ln}(\mathbf{X}) = \{\mathbf{ln}(\mathbf{X}_1)^\top, \mathbf{ln}(\mathbf{X}_2)^\top\}^\top$, $\mathbf{ln}(\mathbf{X}) \sim \text{SE}_d(\boldsymbol{\xi}, \boldsymbol{\Omega}, \boldsymbol{\alpha}, g^{(d)})$. Using the result about the marginal distribution of a skew-elliptical random vector, $\mathbf{ln}(\mathbf{X}_1) \sim \text{SE}_q(\boldsymbol{\xi}_1, \boldsymbol{\Omega}_{11}, \boldsymbol{\alpha}_1^*, g^{(d)})$ and $\mathbf{ln}(\mathbf{X}_2) \sim \text{SE}_{d-q}(\boldsymbol{\xi}_2, \boldsymbol{\Omega}_{22}, \boldsymbol{\alpha}_2^*, g^{(d)})$ with parameters as defined in (3.8). Then, the result follows from Definition 1 of a log-skew-elliptical random vector. \square

Proof of Proposition 6 (mixed moments): Provided the moment exists,

$$E \left(\prod_{i=1}^d X_i^{n_i} \right) = E \left\{ \prod_{i=1}^d e^{n_i \ln(X_i)} \right\} = E \left\{ e^{\sum_{i=1}^d n_i \ln(X_i)} \right\} = E \left\{ e^{\mathbf{n}^\top \mathbf{ln}(\mathbf{X})} \right\} = M_{\mathbf{ln}(\mathbf{X})}(\mathbf{n}).$$

\square

Proof of Proposition 7 (log-skew- t moments): Without loss of generality, let $\boldsymbol{\xi} = \mathbf{0}$ and $\boldsymbol{\Omega} = \bar{\boldsymbol{\Omega}}$. We compute the moments directly from the log-skew- t density (3.6).

Let $U_{\ln(\mathbf{x})} = \boldsymbol{\alpha}^\top \ln(\mathbf{x}) \left(\frac{\nu+d}{\nu+Q_{\ln(\mathbf{x})}} \right)^{1/2}$ and $\bar{\boldsymbol{\Omega}}^{-1} = (\tilde{\rho}_{ij})_{i,j=1}^d$, where $\tilde{\rho}_{ii} > 0$ and $\tilde{\rho}_{ij} = \tilde{\rho}_{ji}$. Let \mathbb{R}_a^d denote $(a, \infty)^d$ and $D^d \mathbf{x}$ denote $dx_1 dx_2 \dots dx_d$. For any $n_i \geq 0$, $i = 1, \dots, d$ the condition $\sum_{i=1}^d n_i > 0$ requires that at least one of the n_i 's is nonzero. Suppose that $n_d > 0$. Then, $E \left(\prod_{i=1}^d X_i^{n_i} \right)$ equals

$$\begin{aligned}
& \int_{\mathbb{R}_0^d} \left(\prod_{i=1}^d x_i^{n_i} \right) f_{\text{LST}_d}(\mathbf{x}; \mathbf{0}, \bar{\boldsymbol{\Omega}}, \boldsymbol{\alpha}, \nu) D^d \mathbf{x} \\
&= 2a_\nu b_\nu \int_{\mathbb{R}_0^d} \prod_{i=1}^d x_i^{n_i-1} \left(1 + \frac{Q_{\ln(\mathbf{x})}}{\nu} \right)^{-\frac{\nu+d}{2}} \int_{-\infty}^{U_{\ln(\mathbf{x})}} \left(1 + \frac{u^2}{\nu+d} \right)^{-\frac{\nu+d+1}{2}} du D^d \mathbf{x} \\
&= 2a_\nu b_\nu \int_{\mathbb{R}_0^{d-1}} \prod_{i=1}^{d-1} x_i^{n_i-1} \int_0^\infty x_d^{n_d-1} \left(1 + \frac{Q_{\ln(\mathbf{x})}}{\nu} \right)^{-\frac{\nu+d}{2}} \\
&\quad \times \int_{-\infty}^{U_{\ln(\mathbf{x})}} \left(1 + \frac{u^2}{\nu+d} \right)^{-\frac{\nu+d+1}{2}} dudx_d D^{d-1} \mathbf{x} \tag{B.2}
\end{aligned}$$

where a_ν and b_ν are normalization constants.

The quadratic form $Q_{\ln(\mathbf{x})}$ can be rewritten as a function of x_d , $Q_{\ln(\mathbf{x})} = A_{\ln(\mathbf{x}_{-d})} + (\ln x_d + B_{\ln(\mathbf{x}_{-d})})^2$, where $A_{\ln(\mathbf{x}_{-d})}$ and $B_{\ln(\mathbf{x}_{-d})}$ do not depend on x_d (are functions of only $\ln x_1, \dots, \ln x_{d-1}$ and $\tilde{\rho}_{ij}$, $i, j = 1, \dots, d$). Then, $U_{\ln(\mathbf{x})} \rightarrow \alpha_d \sqrt{\nu+d}$ as $x_d \rightarrow \infty$, and so $\int_{-\infty}^{U_{\ln(\mathbf{x})}} \left(1 + \frac{u^2}{\nu+d} \right)^{-\frac{\nu+d+1}{2}} du \rightarrow \int_{-\infty}^{\alpha_d \sqrt{\nu+d}} \left(1 + \frac{u^2}{\nu+d} \right)^{-\frac{\nu+d+1}{2}} du > 0$ as $x_d \rightarrow \infty$. Thus, there is a $c_{\nu, \alpha_d} > 0$ (does not depend on \mathbf{x}), such that

$$\int_{-\infty}^{U_{\ln(\mathbf{x})}} \left(1 + \frac{u^2}{\nu+d} \right)^{-\frac{\nu+d+1}{2}} du > c_{\nu, \alpha_d}, \text{ for large } x_d > x_d^*.$$

Then,

$$\begin{aligned}
(B.2) &> 2a_\nu b_\nu \int_{\mathbb{R}_0^{d-1}} \prod_{i=1}^{d-1} x_i^{n_i-1} \int_{x_d^*}^{\infty} x_d^{n_d-1} \left(1 + \frac{Q_{\ln(\mathbf{x})}}{\nu}\right)^{-\frac{\nu+d}{2}} \\
&\quad \times \int_{-\infty}^{U_{\ln(\mathbf{x})}} \left(1 + \frac{u^2}{\nu+d}\right)^{-\frac{\nu+d+1}{2}} du dx_d D^{d-1} \mathbf{x} \\
&> 2a_\nu b_\nu c_{\nu, \alpha_d} \int_{\mathbb{R}_0^{d-1}} \prod_{i=1}^{d-1} x_i^{n_i-1} \\
&\quad \times \left[\int_{x_d^*}^{\infty} x_d^{n_d-1} \left\{ 1 + \frac{A_{\ln(\mathbf{x}_{-d})}}{\nu} + \frac{(\ln x_d + B_{\ln(\mathbf{x}_{-d})})^2}{\tilde{\rho}_{dd} \nu} \right\}^{-\frac{\nu+d}{2}} dx_d \right] D^{d-1} \mathbf{x}.
\end{aligned}$$

The innermost integral over x_d (in square brackets) diverges at infinity for any $n_d > 0$, and, therefore, the d -dimensional integral from the last step diverges for any $n_i \geq 0$, $i = 1, \dots, d-1$. As such, the integral (B.2) also diverges and $E\left(\prod_{i=1}^d X_i^{n_i}\right) = \infty$ for any $n_i \geq 0$, $i = 1, \dots, d-1$ and $n_d > 0$. More generally, $E\left(\prod_{i=1}^d X_i^{n_i}\right) = \infty$ for any $n_i \geq 0$, $i = 1, \dots, d$, such that $\sum_{i=1}^d n_i > 0$. \square

VITA

Yulia V. Marchenko received a Diploma (degree equivalent to M.S.) in Applied Mathematics in August 2002 from the Belarussian State University in Minsk, Belarus, where she also minored in Statistics. In August 2004, she received her M.S. in Statistics from Texas A&M University, and her Ph.D. in Statistics at Texas A&M University was awarded in December 2010. Yulia has also been employed by Stat-aCorp LP since June 2004 and is currently a senior statistician there. Her current research interests include skewed distributions, multiple imputation, survival analysis, statistical genetics, and statistical software development. Her address is: Department of Statistics, Texas A&M University, 3143 TAMU, College Station, TX 77843-3143. Her email address is yulia@stat.tamu.edu.

The typist for this dissertation was Yulia Marchenko.