The Stata Journal (2010)
10, Number 4, pp. 507–539

A suite of commands for fitting the skew-normal and skew-t models

Yulia V. Marchenko
StataCorp
College Station, TX
ymarchenko@stata.com

Marc G. Genton
Department of Statistics
Texas A&M University
College Station, TX
genton@stat.tamu.edu

Abstract. Nonnormal data arise often in practice, prompting the development of flexible distributions for modeling such situations. In this article, we describe two multivariate distributions, the skew-normal and the skew-t, which can be used to model skewed and heavy-tailed continuous data. We then discuss some inferential issues that can arise when fitting these distributions to real data. We also consider the use of these distributions in a regression setting for more flexible parametric modeling of the conditional distribution given other predictors. We present commands for fitting univariate and multivariate skew-normal and skew-t regressions in Stata (`skewnreg`, `skewtreg`, `mskewnreg`, and `mskewtreg`) as well as some postestimation features (`predict` and `skewrplot`). We also demonstrate the use of the commands for the analysis of the famous Australian Institute of Sport data and U.S. precipitation data.

Keywords: st0207, skewnreg, skewtreg, mskewnreg, mskewtreg, skewrplot, predict, distribution, heavy tails, nonnormal, precipitation, regression, skewness, skew-normal, skew-t

1 Introduction

Nonnormal data arise often in practice. One common way of dealing with nonnormal data is to find a suitable transformation that makes the data more normal-like and to apply standard normal-based methods to 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. These difficulties motivated a search for more-flexible parametric families of distributions to model nonnormal data. A number of approaches are available for univariate outcomes. For noncontinuous data, such as binary data or count data, binomial or Poisson distributions can be used. More generally, generalized linear models can be used to accommodate a range of distributions within an exponential family. However, the choices for multivariate outcomes are rather limited.

Our focus in this article is on continuous nonnormal data. Because real data often deviate from normality in the tails or exhibit asymmetry in the distribution, there has been a growing interest in distributions with additional parameters regulating asymmetry and tails directly. For example, for heavy-tailed data, the Student's t distribution is often considered. Traditionally, lognormal or gamma distributions are used to model positive skewed data. To accommodate asymmetry for data spanning a real line, one can consider skew-normal and skew-t distributions, which are skewed versions of the respective normal and Student's t distributions. One of the appealing features of these distributions is that they have tractable multivariate versions that allow us to model multivariate outcomes. More generally, the family of skew-elliptical distributions proposed by Branco and Dey (2001) allows 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 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), which extends the normal distribution to allow for both asymmetry and heavier tails with two additional parameters, a shape parameter and a degrees-of-freedom parameter. These extra parameters allow us to capture the features of the data more adequately. Azzalini and Dalla Valle (1996), Azzalini and Capitanio (1999), Branco and Dey (2001), and Azzalini and Capitanio (2003) study 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 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-squared 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). Similarly to the normal and Student's t distributions, the skew-normal and skew-t distributions can also be adapted to handle positive data by considering their log versions (Azzalini, dal Cappello, and Kotz 2002; Marchenko and Genton 2010).

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 our article is the following: We start with a motivating example in section 2. In section 3, we proceed to describe the skewed distributions and, more generally, skewed regressions in more detail. We present commands for fitting the skewed models in section 4. In section 5, we provide more examples of using skew-normal and skew-t models in the analysis of the Australian Institute of Sport data, commonly used in the literature about skewed distributions.

2 Motivating example

We consider the Australian Institute of Sport dataset (Cook and Weisberg 1994), which is repeatedly used in the literature about skewed distributions. The `ais.dta` dataset contains 202 observations (100 females and 102 males) that record 13 biological characteristics of Australian athletes. In our examples, we use only a subset of these characteristics.

```
. use ais
(Biological measures from athletes at the Australian Institute of Sport)
. describe lbm bmi weight height fe female

              storage  display     value
variable name   type   format      label      variable label
-------------------------------------------------------------------------------------

lbm             double %9.0g                  Lean body mass (kg)
bmi             double %9.0g                  Body mass index (kg/m^2)
weight          double %9.0g                  Weight (kg)
height          double %9.0g                  Height (m)
fe              int    %9.0g                  Plasma ferritin concentration (ng/ml)
female          byte   %9.0g       gender     Gender
```

Suppose we are interested in modeling plasma ferritin concentration recorded in the fe variable. From figure 1, we can see that the distribution of plasma ferritin concentration is skewed to the right compared with the normal distribution.

```
. histogram fe, normal
(bin=14, start=8, width=16.142857)
```

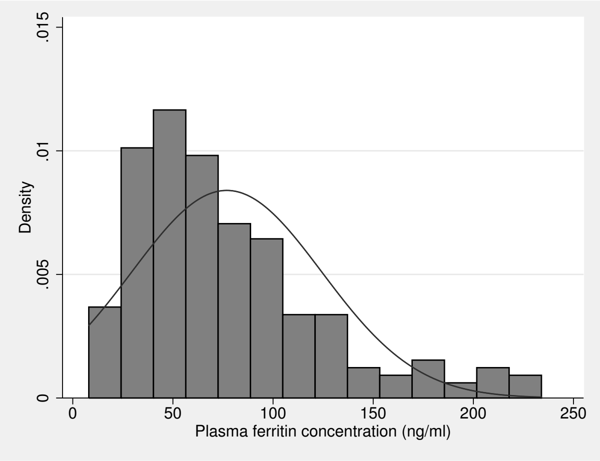

Figure 1. Histogram of plasma ferritin concentration overlaid with normal density

As mentioned in the introduction, for a univariate outcome we can choose from several options. We can use a transformation-based approach and model the fe variable in the log metric, for example. If we prefer to work with the original scale, we can use one of the univariate distributions that accommodate asymmetry. Here we demonstrate the use of the skew-normal and skew-t distributions for modeling fe.

We first fit the skew-normal distribution to plasma ferritin concentration fe using the new skewnreg command. For later comparison with the skew-t fit, we specify the dpmetric option to report results in the direct parameterization, which will be explained in section 3.3:

```
. skewnreg fe, dpmetric
initial:       log likelihood = -1033.6914
rescale:       log likelihood = -1033.6914
rescale eq:    log likelihood = -1033.6914
Iteration 0:   log likelihood = -1033.6914
Iteration 1:   log likelihood = -1032.6839
Iteration 2:   log likelihood = -1030.9463
Iteration 3:   log likelihood = -1030.9116
Iteration 4:   log likelihood = -1030.9115
Skew-normal regression                            Number of obs   =        202
                                                  Wald chi2(0)    =          .
Log likelihood = -1030.9115                       Prob > chi2     =          .

------------------------------------------------------------------------------
          fe |      Coef.   Std. Err.      z    P>|z|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
       _cons |   20.24412   2.491879     8.12   0.000     15.36012    25.12811
-------------+----------------------------------------------------------------
       alpha |   9.142567    2.56432                      4.116592    14.16854
-------------+----------------------------------------------------------------
       omega |   73.84035   4.141059                      66.15418    82.41954
------------------------------------------------------------------------------
LR test vs normal regression:       chi2(1) =    70.17   Prob > chi2 = 0.0000
```

As mentioned in the introduction, compared with the symmetric normal distribution, the skew-normal distribution has an additional shape parameter. Labeled as alpha in the output, it regulates the asymmetry of the distribution. For positive values of the shape parameter, the distribution is skewed to the right; for negative values, the distribution is skewed to the left; and the distribution is symmetric (normal) when the shape parameter is zero. From the output, we can see that alpha is estimated to be 9.14 with a 95% confidence interval of $[4.12, 14.17]$, which is evidence that the distribution of fe exhibits skewness to the right.

We can visually check how well the skew-normal distribution fits the data by using the new postestimation command, skewrplot:

```
. skewrplot, fitted
(bin=14, start=8, width=16.142857)
```

We specified the fitted option to plot the skew-normal density estimate [evaluated at the above maximum likelihood estimates (MLEs) of the model parameters] of the fitted values against the histogram of fe. From figure 2, we can see that the skew-normal density estimate closely follows the nonparametric density estimate and that it demonstrates better fit of the skew-normal distribution to fe than the normal distribution in figure 1.

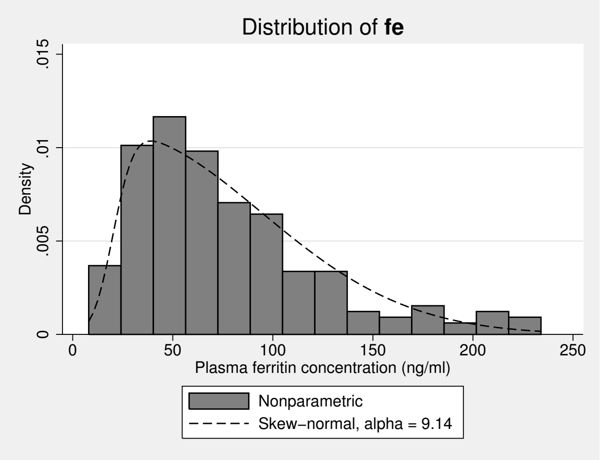

Figure 2. Histogram and skew-normal density estimate of plasma ferritin concentration

We can also fit the skew-*t* distribution to `fe` by using the `skewtreg` command:

```
. skewtreg fe
initial:       log likelihood = -1428.9045
rescale:       log likelihood = -1411.4498
rescale eq:    log likelihood = -1041.7301
Iteration 0:   log likelihood = -1041.7301
Iteration 1:   log likelihood = -1035.0139
Iteration 2:   log likelihood = -1030.6871
Iteration 3:   log likelihood =  -1029.457
Iteration 4:   log likelihood = -1029.1935
Iteration 5:   log likelihood =  -1029.186
Iteration 6:   log likelihood =  -1029.186
Skew-t regression                               Number of obs   =        202
                                                Wald chi2(0)    =          .
Log likelihood =  -1029.186                     Prob > chi2     =          .

------------------------------------------------------------------------------
          fe |      Coef.   Std. Err.      z    P>|z|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
       _cons |    22.2901   2.830001     7.88   0.000      16.7434     27.8368
-------------+----------------------------------------------------------------
       alpha |   7.244468   2.270883     3.19   0.001     2.793619    11.69532
-------------+----------------------------------------------------------------
       omega |   62.12069   7.079737                      49.68519    77.66861
-------------+----------------------------------------------------------------
          df |   7.440234   4.405123                      2.331404     23.7441
------------------------------------------------------------------------------
LR test vs normal regression:  chibar2(1_2) =   73.62 Prob >= chibar2 = 0.0000
```

In addition to the shape parameter, the skew-*t* distribution introduces a degrees-of-freedom parameter. Labeled as `df` in the output, this parameter regulates the heaviness of the tails of the distribution. The smaller the degrees of freedom, the "heavier" the tails of the distribution. (For instance, one degree of freedom yields a skew-Cauchy distribution of Arnold and Beaver [2000].) As the degrees of freedom becomes large, the skew-t distribution reduces to the skew-normal distribution or the normal distribution, when in addition the shape parameter is zero. From the output, we can see that the degrees of freedom is estimated to be 7.44 with a 95% confidence interval of $[2.33, 23.74]$, which provides evidence for heavier-than-normal tails of the distribution of `fe`. The estimate of the shape parameter `alpha` is 7.24 with a 95% confidence interval $[2.79, 11.70]$, which again confirms the existence of positive skewness in the distribution of `fe`.

As we did before, we can plot the density estimate of fitted values from the skew-t distribution estimated above against the nonparametric density estimate. The plot is shown in figure 3:

```
. skewrplot, fitted
(bin=14, start=8, width=16.142857)
```

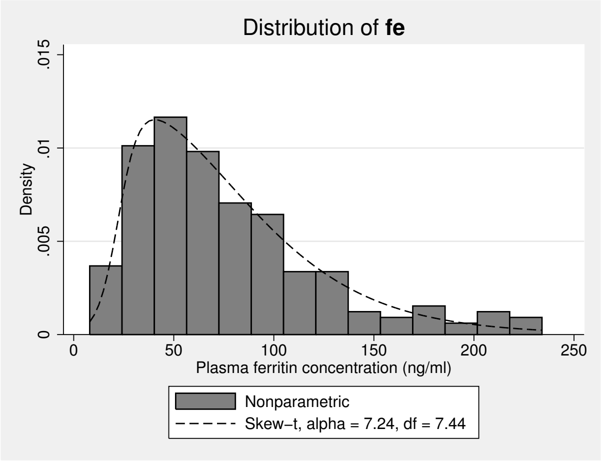

Figure 3. Histogram and skew-t density estimate of plasma ferritin concentration

From the graph, we can see that the skew-t distribution seems to fit the `fe` values better than the skew-normal distribution. We could also use probability–probability (P–P) or quantile–quantile (Q–Q) plots, as we demonstrate later, to more easily compare model fits.

Let us now describe the skew-normal and skew-t models in more detail.

3 The skew-normal and skew-t models

3.1 Definition and some properties

The density of the univariate skew-normal distribution, $\text{SN}(\xi, \omega^2, \alpha)$, is

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

where $z = \omega^{-1}(x - \xi)$, $\xi \in \mathbb{R}$ is a location parameter, $\omega > 0$ is a scale parameter, $\phi(\cdot)$ 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 a 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 (1) reduces to the normal distribution.

The univariate skew-*t* distribution, $\text{ST}(\xi, \omega^2, \alpha, \nu)$, 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:

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

where $t(z; \nu)$ is the density of a univariate standard Student's *t* distribution with degrees of freedom ν, and $T(\cdot; \nu+1)$ is the cumulative distribution function of a univariate standard Student's *t* distribution with $\nu+1$ degrees of freedom. Here again, $\xi \in \mathbb{R}$ regulates the location of the distribution, $\omega > 0$ regulates the scale of the distribution, the shape parameter $\alpha \in \mathbb{R}$ regulates asymmetry of the distribution, and the degrees-of-freedom parameter $\nu > 0$ regulates the tails of the distribution. When $\alpha = 0$, the density (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 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 $X \sim \text{SN}(\xi, \omega^2, \alpha)$, then $(X - \xi)^2/\omega^2 \sim \chi_1^2$. If $X \sim \text{ST}(\xi, \omega^2, \alpha, \nu)$, then $(X - \xi)^2/\omega^2 \sim F_{1,\nu}$. These properties provide a way of evaluating model fit using Q–Q or P–P 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, $\text{SN}_d(\boldsymbol{\xi}, \Omega, \boldsymbol{\alpha})$, is

$$f_{\text{SN}_d}(\mathbf{x}; \Theta) = 2\,\phi_d(\mathbf{x}; \boldsymbol{\xi}, \Omega)\Phi(\boldsymbol{\alpha}'\mathbf{z}), \quad \mathbf{x} \in \mathbb{R}^d \tag{3}$$

where $\Theta = (\boldsymbol{\xi}, \Omega, \boldsymbol{\alpha})$, $\mathbf{z} = \Omega_{\text{diag}}^{-1/2}(\mathbf{x} - \boldsymbol{\xi}) \in \mathbb{R}^d$, $\phi_d(\mathbf{x}; \boldsymbol{\xi}, \Omega)$ is the density of a d-variate normal distribution with location $\boldsymbol{\xi}$ and covariance matrix Ω, and Ω_{diag} is the $d \times d$

diagonal matrix containing the diagonal elements of Ω. Similarly to the univariate case, when all d components of $\boldsymbol{\alpha}$ are zero, the multivariate skew-normal density (3) reduces to the multivariate normal density $\phi_d(\cdot)$.

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

$$f_{\textsc{st}_d}(\mathbf{x}; \Theta) = 2\, t_d(\mathbf{x}; \boldsymbol{\xi}, \Omega, \nu) T \left\{ \boldsymbol{\alpha}' \mathbf{z} \left(\frac{\nu + d}{\nu + Q_{\mathbf{x}}^{\boldsymbol{\xi}, \Omega}} \right)^{1/2} ; \nu + d \right\}, \quad \mathbf{x} \in \mathbb{R}^d \tag{4}$$

where $\Theta = (\boldsymbol{\xi}, \Omega, \boldsymbol{\alpha}, \nu)$, $\mathbf{z} = \Omega_{\text{diag}}^{-1/2}(\mathbf{x} - \boldsymbol{\xi})$, $Q_{\mathbf{x}}^{\boldsymbol{\xi}, \Omega} = (\mathbf{x} - \boldsymbol{\xi})' \Omega^{-1} (\mathbf{x} - \boldsymbol{\xi})$, $t_d(\mathbf{x}; \boldsymbol{\xi}, \Omega, \nu) = \Gamma\{(\nu + d)/2\}(1 + Q_{\mathbf{x}}^{\boldsymbol{\xi}, \Omega}/\nu)^{-(\nu+d)/2} / \{|\Omega|^{1/2} (\nu\pi)^{d/2} \Gamma(\nu/2)\}$ 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) 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 $\mathbf{X} \sim \textsc{sn}_d(\boldsymbol{\xi}, \Omega, \boldsymbol{\alpha})$, then the Mahalanobis measure $(\mathbf{X} - \boldsymbol{\xi})' \Omega^{-1} (\mathbf{X} - \boldsymbol{\xi}) \sim \chi^2_d$. If $\mathbf{X} \sim \textsc{st}_d(\boldsymbol{\xi}, \Omega, \boldsymbol{\alpha}, \nu)$, then $\frac{1}{d}(\mathbf{X} - \boldsymbol{\xi})' \Omega^{-1} (\mathbf{X} - \boldsymbol{\xi}) \sim F_{d,\nu}$.

3.2 Regression models

Consider a sample $\mathbf{Y} = (y_1, y_2, \ldots, y_n)'$ of n observations. In linear regression,

$$y_i = \beta_0 + \beta_1 x_{1i} + \cdots + \beta_p x_{pi} + \epsilon_i, \ i = 1, \ldots, n \tag{5}$$

where $x_{1i}, \ldots, x_{pi}$ define covariate values, $\beta_0, \ldots, \beta_p$ are the unknown regression coefficients, and ϵ_i is an error term. In normal linear regression, the errors are assumed to be normally distributed, $\epsilon_i \overset{\text{iid}}{\sim} \text{Normal}(0, \sigma^2)$. The skew-normal regression is a linear regression (5) with errors from the skew-normal distribution, $\epsilon_i \overset{\text{iid}}{\sim} \textsc{sn}(0, \omega^2, \alpha)$. Similarly, the skew-$t$ regression is defined by (5) with $\epsilon_i \overset{\text{iid}}{\sim} \textsc{st}(0, \omega^2, \alpha, \nu)$. Equivalently, the sample $\mathbf{Y}$ is assumed to follow the skew-normal distribution, $y_i \overset{\text{iid}}{\sim} \textsc{sn}(\xi_i, \omega^2, \alpha)$, or the skew-$t$ distribution, $y_i \overset{\text{iid}}{\sim} \textsc{st}(\xi_i, \omega^2, \alpha, \nu)$, respectively, where $\xi_i = \beta_0 + \beta_1 x_{1i} + \cdots + \beta_p x_{pi}$. 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, $\textsc{sn}_d(\mathbf{0}, \Omega, \boldsymbol{\alpha})$, or the multivariate skew-t distribution, $\textsc{st}_d(\mathbf{0}, \Omega, \boldsymbol{\alpha}, \nu)$, respectively.

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

The existence of a stationary point at $\boldsymbol{\alpha} = \mathbf{0}$ for the skew-normal model leads to the singularity of the Fisher information matrix of the profile log likelihood for the shape parameter $\boldsymbol{\alpha}$ (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 a bimodal limiting distribution of the estimates (Arellano-Valle and Azzalini 2008). All model parameters $\boldsymbol{\xi}$, Ω, and $\boldsymbol{\alpha}$ are identifiable, so the issue is really due to the chosen parameterization. To alleviate this issue, Azzalini (1985) suggested an alternative centered parameterization for the univariate skew-normal model under which the sampling distributions of the new parameters are closer to the normal distribution. Arellano-Valle and Azzalini (2008) extended this parameterization to the multivariate case. We will discuss the centered parameterization in more detail in section 3.3. This unfortunate property seems to vanish in the case of the skew-t distribution, unless the degrees of freedom are large enough 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 (forthcoming).

Both the skew-normal and skew-t models suffer from the problem of unboundedness of the MLEs 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 unbound are more complicated and thus more difficult to describe. Sartori (2006) and Azzalini and Genton (2008) presented ways of dealing with the unbound estimates. Sartori (2006) proposed a bias correction to the MLEs. Azzalini and Genton (2008) suggested a deviance-based approach according to which the unbound 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.

The centered parameterization is available for `skewnreg` and `mskewnreg` to alleviate the singularity issue. The issue of unbound parameter estimates is not yet addressed in the presented commands. This issue is likely to arise when the distribution of the data (or residuals within the regression framework) is close to a half-normal distribution. If this issue occurs, one solution is to determine the iteration number after which the changes in the likelihood become very small and then to refit the model using the prespecified number of iterations in the `iterate(#)` option.

3.3 Centered parameterization

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

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

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

where $\mu_z = E(Z) = \sqrt{2/\pi}\delta$, $\sigma_z^2 = \mathrm{Var}(Z) = 1 - 2\delta^2/\pi$, and $\delta = \alpha/\sqrt{1+\alpha^2}$. Then $\mu = E(Y) = \xi + \omega\mu_z$ and $\sigma^2 = \mathrm{Var}(Y) = \omega^2(1-\mu_z^2)$. Let $\gamma = (4-\pi)\,\mathrm{sign}(\alpha)\,(\mu_z/\sigma_z^2)^3/2$ denote the skewness index of Y. (The skewness index γ is not the classical sample moment-based measure of skewness but is specific to this family of distributions.) Then, mean, standard deviation, and skewness index, (μ, σ, γ), form the centered parameterization. They are referred to as the centered parameters (CP) because they are obtained by centering Y. The set of parameters (ξ, ω, α) are referred to as the direct parameters (DP). It is worth noting that unlike the range of α, the range of γ is restricted to approximately $(-0.9953, 0.9953)$. More generally in the multivariate setting, unlike the DPs $(\boldsymbol{\xi}, \Omega, \boldsymbol{\alpha})$, the CPs $(\boldsymbol{\mu}, \Sigma, \boldsymbol{\gamma})$ cannot be chosen freely and are subject to certain constraints; see Arellano-Valle and Azzalini (2008) for details. Of course, both sets of parameters require the scale matrices to be positive definite.

In the regression setting, the CP metric affects only the estimate of the intercept and not the coefficients. Specifically, $\beta_0^{CP} = \beta_0 + \sqrt{2/\pi}\omega\delta$, $\beta_i^{CP} = \beta_i, i = 1, \ldots, p$. Consequently, $\epsilon_i^{CP} = \epsilon_i - \sqrt{2/\pi}\omega\delta$ and so the residuals in the CP metric have a mean of zero, $E(\epsilon_i^{CP}) = 0,\ i = 1, \ldots, n$. In what follows, when referring to residuals we will always assume the residuals are in the DP metric.

The use of CP is advantageous from both inferential and interpretation standpoints. The sampling distributions of the MLEs 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 guide 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 $\boldsymbol{\gamma}$ represent the skewness indexes of the marginal distributions whereas individual components of $\boldsymbol{\alpha}$, in general, cannot be used to infer the direction or the magnitude of the asymmetry in marginal distributions. Marginal skewness indexes are complicated functions of individual components of $\boldsymbol{\alpha}$. However, zero components of $\boldsymbol{\alpha}$ do imply zero marginal skewness indexes or, in other words, symmetric marginal distributions. DP is useful for direct interpretation in the original model.

From the above formulas, we can see that a one-to-one correspondence exists between CP and DP, provided CP is within its admissible range. 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, and vice versa.

The centered parameterization is implemented in `skewnreg` and `mskewnreg`. At the time of publication of this article, the centered parameterization for the skew-t distribution is yet to appear in the literature (Arellano-Valle and Azzalini 2009) and thus is not implemented in `skewtreg` and `mskewtreg`.

4 A suite of commands for fitting skewed regressions

4.1 Syntax

Skewed regression models

Univariate skew-normal regression

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

Univariate skew-t regression

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

Multivariate skew-normal regression

`mskewnreg` *depvars* [`=` *indepvars*] [*if*] [*in*] [*weight*] [`,`
`constraints(`*constraints*`) collinear vce(`*vcetype*`) level(`*#*`) dpmetric`
`estmetric noshowomega nocnsreport coeflegend postdp postcp`
display_options maximize_options]

(*Continued on next page*)

Multivariate skew-t regression

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

indepvars may contain factor variables; see [U] **Factor variables**.
`fweight`s are allowed; see [U] **weight**.

Postestimation features

Predictions

`predict` [*type*] *newvar* [*if*] [*in*] [`,` `xb` `residuals` `score` `stdp`
`equation(`*eqno*`)`]

Residual density plot over histogram (default with `skewnreg` *and* `skewtreg`*)*

`skewrplot` [`,` `histogram` `fitted` `normal` `normopts(`*norm_options*`)`
`lineopts(`*line_options*`)` `histopts(`*hist_options*`)` `addplot(`*plot*`)` *twoway_options*]

Residual density plot with kernel-density estimate (`skewnreg` *and* `skewtreg` *only)*

`skewrplot, kdensity` [`fitted` `normal` `normopts(`*norm_options*`)`
`lineopts(`*line_options*`)` `kdenopts(`*kden_options*`)` `addplot(`*plot*`)` *twoway_options*]

Residual-versus-fitted plot (`skewnreg` *and* `skewtreg` *only)*

`skewrplot, rvf` [`addplot(`*plot*`)` *scatter_options* *twoway_options*]

Probability–probability plot

`skewrplot, pp` [`normal` `normopts(`*norm_options*`)` `overlay` `addplot(`*plot*`)`
pp_options *graph_options*]

Quantile–quantile plot (default with `mskewnreg` *and* `mskewtreg`*)*

`skewrplot, qq` [`normal` `normopts(`*norm_options*`)` `addplot(`*plot*`)` *qq_options* *graph_options*]

4.2 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. `skewnreg` and `mskewnreg` support both the CP metric (the default) and the DP metric (with the `dpmetric` option), whereas `skewtreg` and `mskewtreg` support only the DP metric. Regardless of the display metric, optimization is performed in the estimation metric specific to each command; see each command's help file for details. In the skew-t regression, the degrees-of-freedom parameter can optionally be set to a fixed value with the `df()` option.

The postestimation features include predictions and residual diagnostics 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 `equation()` option can be used with multivariate regressions to obtain equation-specific predictions. The first equation is assumed by default.

The `skewrplot` command can be used after any of the four estimation commands to obtain a number of residual diagnostic plots. The default after univariate regressions is a residual density plot, where the skew-normal (or skew-t) density estimate of residuals, evaluated at MLEs from the previously fit model, is plotted together with a nonparametric residual density estimate—a histogram. Alternatively, if `kdensity` is used, a residual density plot is displayed together with a nonparametric kernel-density estimate of residuals instead of the histogram. In the absence of predictors, the `fitted` option can be used to plot density estimates of the fitted values instead of residuals. In addition, a normal density estimate can be added to the graph as a reference by specifying the `normal` option. The residual-versus-fitted plot can be obtained with the `rvf` option. The P–P and Q–Q plots are available after univariate or multivariate regressions. The Q–Q plot of residuals is the default after multivariate regressions. It can also be requested with the `qq` option. The P–P plot of residuals can be obtained with the `pp` option. If `normal` is used in combination with `pp` (or `qq`), a P–P (or Q–Q) plot of residuals from a normal regression fit is produced as a separate plot.

4.3 Options

Common estimation options

`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.

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.

nocnsreport specifies that no constraints be reported. The default is to display user-specified constraints above the coefficient table.

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.

postdp stores DP estimates and their variance–covariance estimator (VCE) in e(b) and e(V), respectively.

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, trace, gradient, showstep, hessian, showtolerance, tolerance(*#*), ltolerance(*#*), nrtolerance(*#*), nonrtolerance; see [R] **maximize**. Also, init(*ml_init_args*) can be specified; see [R] **ml**.

Other options for skewnreg

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

Other options for mskewnreg

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

noshowomega specifies that the display of the covariance (or scale) matrix be suppressed.

postcp stores CP estimates and their VCE in e(b) and e(V), respectively, instead of the estimation parameters.

Other options for skewtreg

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

Other options for mskewtreg

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

`noshowomega` specifies that the display of the covariance (or scale) matrix be suppressed.

Options for predict

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

`residuals` calculates the residuals.

`score` calculates the first derivative of the log likelihood with respect to $x_j\beta$.

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

`equation(`*eqno*`)` is allowed only when you have previously fit `mskewnreg` or `mskewtreg`. It specifies the equation to which you are referring. `equation()` is filled in with one *eqno* for the `xb`, `stdp`, and `residuals` options. `equation(#1)` means the calculation is to be made for the first equation; `equation(#2)` means the second; and so on. You could also refer to the equations by their names. `equation(lbm)` would refer to the equation named `lbm`, and `equation(bmi)` would refer to the equation named `bmi`. If you do not specify `equation()`, results are the same as if you specified `equation(#1)`.

Options for skewrplot

`histogram`, the default after `skewnreg` and `skewtreg`, requests that the histogram of residuals be plotted together with a residual density estimate from a `skewnreg` or `skewtreg` fit. This option is not allowed with `skewrplot` after `mskewnreg` or `mskewtreg`.

`kdensity` requests that the kernel-density estimate of residuals be plotted together with a residual density estimate from a `skewnreg` or `skewtreg` fit instead of the histogram. This option is not allowed with `skewrplot` after `mskewnreg` or `mskewtreg`.

`rvf` requests that the residual-versus-fitted plot be produced. This option is not allowed with `skewrplot` after `mskewnreg` or `mskewtreg`.

`pp` requests that probability–probability plots of the observed residuals versus the residuals obtained from the fitted parametric model be produced.

qq, the default after mskewnreg and mskewtreg, requests that quantile–quantile plots of the observed residuals versus the residuals obtained from the fitted parametric model be produced.

fitted requests that the density of fitted values be plotted instead of the density of residuals from a skewnreg or skewtreg fit. This option is allowed only in combination with histogram or kdensity.

normal requests that a corresponding normal plot be produced for comparison. If histogram is used, normal specifies that the histogram be overlaid with an appropriately scaled normal density. The normal will have the same mean and standard deviation as the data. If kdensity is used, normal requests that a normal density be overlaid on the density estimate of residuals from a skewed regression fit. If pp or qq is used, normal requests that an additional, separate chi-squared probability plot or chi-squared quantile plot of squared standardized residuals from a normal regression fit be produced. This option can be used in combination with overlay to overlay P–P plots on one graph. This option is not allowed in combination with rvf.

normopts(*norm_options*) specifies details about the look of normal plots produced when normal is specified. If histogram or kdensity is used, *norm_options* affect rendition of the normal curve, such as the color and style of line used, and can be any of the options documented in [G] **graph twoway line**. If pp (or qq) is used, *norm_options* affect the look of the chi-squared probability (or quantile) plot and can be any of the options documented for quantile in [R] **diagnostic plots**.

overlay specifies that the normal plot be overlaid with the main plot in one graph. This option requires normal and is not allowed in combination with qq. This option is implied with histogram and kdensity.

lineopts(*line_options*) affect rendition of the curve from the skew fit. Aspects such as the color and style of line used are affected and can be specified using any of the options documented in [G] **graph twoway line**.

histopts(*hist_options*) are any of the options other than discrete, fraction, frequency, percent, horizontal, and all **Density plots** options documented in [R] **histogram**.

kdenopts(*kden_options*) are any of the options documented in [R] **kdensity**.

addplot(*plot*) provides a way to add other plots to the generated graph; see [G] ***addplot_option***.

scatter_options are any of the options documented in [G] **graph twoway scatter**.

pp_options are any of the options of quantile documented in [R] **diagnostic plots**.

qq_options are any of the options of quantile documented in [R] **diagnostic plots**.

twoway_options are any of the options other than by() documented in [G] ***twoway_options***.

graph_options specify the overall look of a graph. If `normal` is used without `overlay`, *graph_options* are any of the options documented in [G] **graph combine**. Otherwise, *graph_options* are any of the *twoway_options* above.

5 Numerical examples

5.1 Univariate analysis of Australian Institute of Sport data

Our motivating example demonstrated the use of `skewnreg` and `skewtreg` for modeling the distribution of plasma ferritin concentration from the Australian Institute of Sport data. We can also use these commands within the regression framework to accommodate departures from normality of the conditional distribution of the outcome of interest controlling for other covariates.

For the purpose of illustration, consider the conditional distribution of lean body mass, `lbm`, given the weight and height of an athlete. Linearity of lean body mass with respect to weight and height was established by previous analysis of these data (for example, Cook and Weisberg [1994]), so we consider a simple linear regression for modeling the conditional distribution of `lbm`. To obtain more meaningful estimates of main effects, we use recentered versions of covariates, `weight_c` and `height_c`, in our regression analysis. Also, to adjust for likely differences in the relationship due to gender, we interact `weight_c` and `height_c` with `female`. (Alternatively, we could have fit separate regressions for males and females to also allow the variability in the measurements to differ across gender.)

```
. use ais, clear
(Biological measures from athletes at the Australian Institute of Sport)
. summarize weight, meanonly
. generate weight_c = weight - r(mean)
. summarize height, meanonly
. generate height_c = height - r(mean)
```

We first fit a normal linear regression and examine the distribution of the residuals from its fit. It is worth noting that weight and height measurements are highly correlated.

(*Continued on next page*)

```
. regress lbm i.female##c.(weight_c height_c)
      Source |       SS       df       MS              Number of obs =     202
-------------+------------------------------           F(  5,   196) = 1087.52
       Model |  33142.2236     5  6628.44472           Prob > F      =  0.0000
    Residual |  1194.61754   196  6.09498744           R-squared     =  0.9652
-------------+------------------------------           Adj R-squared =  0.9643
       Total |  34336.8411   201  170.830055           Root MSE      =  2.4688

------------------------------------------------------------------------------
         lbm |      Coef.   Std. Err.      t    P>|t|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
    1.female |  -9.014547   .4304858   -20.94   0.000    -9.863526   -8.165568
    weight_c |   .7101775   .0265595    26.74   0.000     .6577985    .7625566
    height_c |   14.83978   4.169091     3.56   0.000     6.617744    23.06182
             |
     female# |
  c.weight_c |
           1 |  -.1765309    .041757    -4.23   0.000    -.2588816   -.0941802
             |
     female# |
  c.height_c |
           1 |  -5.442548   5.965791    -0.91   0.363    -17.20793    6.322834
             |
       _cons |   68.51799   .3006605   227.89   0.000     67.92504    69.11093
------------------------------------------------------------------------------

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

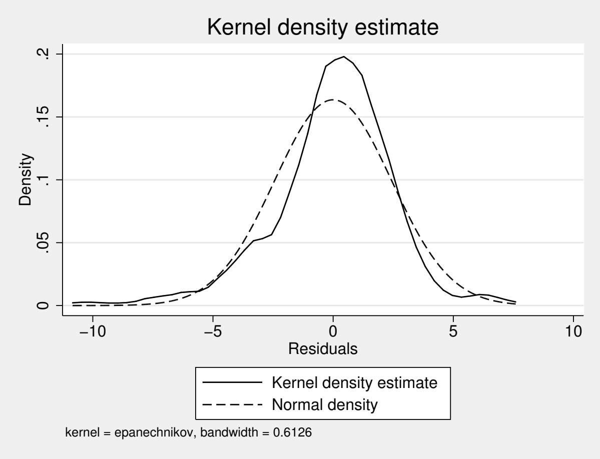

Figure 4. Normal residuals density estimate

Figure 4 demonstrates a slight (longer left tail) skewness in the distribution of residuals compared with the assumed underlying normal distribution. More directly, we can use a Q–Q plot to compare the distribution of residuals with the normal distribution, as shown in figure 5:

```
. qnorm resid
```

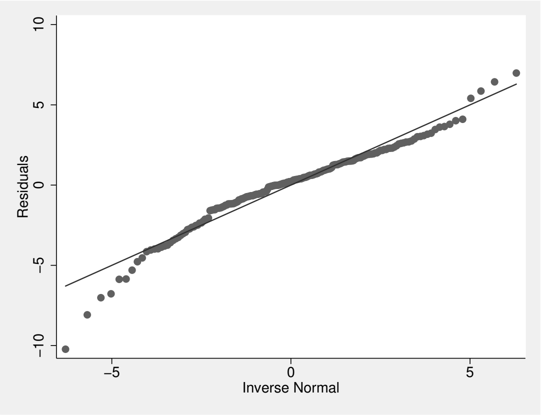

Figure 5. Normal Q–Q plot of residuals

The Q–Q plot confirms the existence of negative skewness in the distribution of residuals from the linear regression fit.

We store estimation results from `regress` for later comparison with skewed models:

```
. estimates store reg
```

To capture asymmetry in the data, we now fit the skew-normal regression:

```
. skewnreg lbm i.female##c.(weight_c height_c), nolog
Skew-normal regression                            Number of obs   =        202
                                                  Wald chi2(5)    =    6773.01
Log likelihood = -457.54665                       Prob > chi2     =     0.0000

------------------------------------------------------------------------------
         lbm |      Coef.   Std. Err.      z    P>|z|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
    1.female |  -8.225366   .4431113   -18.56   0.000    -9.093848   -7.356883
    weight_c |   .7737271   .0306772    25.22   0.000     .7136009    .8338533
    height_c |    9.91473   4.072276     2.43   0.015     1.933216    17.89624
             |
     female#|
  c.weight_c |
          1  |  -.1959762   .0382144    -5.13   0.000    -.2708751   -.1210774
             |
     female#|
  c.height_c |
          1  |  -3.118911   5.621625    -0.55   0.579    -14.13709    7.899271
             |
       _cons |   68.05071   .3032845   224.38   0.000     67.45629    68.64514
-------------+----------------------------------------------------------------
       gamma |  -.6191484   .1192347    -5.19   0.000    -.8528442   -.3854526
-------------+----------------------------------------------------------------
       sigma |   2.416606   .1314719                      2.172189    2.688526
------------------------------------------------------------------------------
LR test vs normal regression:        chi2(1) =    17.18   Prob > chi2 = 0.0000
```

By default, skewnreg estimates and displays model parameters other than the standard deviation sigma in the CP metric, as discussed in section 3.3. The standard deviation is estimated in the log metric. From the output, we can see that both weight and height are strong predictors of lean body mass measurements, and their relationship differs between males and females. The estimated skewness index, labeled as gamma in the output, is -0.62, which suggests that the conditional distribution of lbm adjusted for weight and height is skewed to the left. According to the reported test of H_0: $\gamma = 0$ with the test statistic of -5.19, we have strong evidence of 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, which is 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:

```
. skewnreg, dpmetric
Skew-normal regression                          Number of obs   =        202
                                                Wald chi2(5)    =    6773.01
Log likelihood = -457.54665                     Prob > chi2     =     0.0000

------------------------------------------------------------------------------
         lbm |      Coef.   Std. Err.      z    P>|z|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
    1.female |  -8.225366   .4431113   -18.56   0.000    -9.093848   -7.356883
    weight_c |   .7737271   .0306772    25.22   0.000     .7136009    .8338533
    height_c |    9.91473   4.072276     2.43   0.015     1.933216    17.89624
             |
     female#
  c.weight_c |
          1  |  -.1959762   .0382144    -5.13   0.000    -.2708751   -.1210774
             |
     female#
  c.height_c |
          1  |  -3.118911   5.621625    -0.55   0.579    -14.13709    7.899271
             |
       _cons |   70.78126   .2882586   245.55   0.000     70.21628    71.34624
-------------+----------------------------------------------------------------
       alpha |  -2.718978   .6434226                     -3.980063   -1.457893
-------------+----------------------------------------------------------------
       omega |   3.646351    .273574                      3.147716    4.223975
------------------------------------------------------------------------------
LR test vs normal regression:       chi2(1) =    17.18   Prob > chi2 = 0.0000
```

Notice that all regression coefficients remain the same: the transformation from the CP to the DP metric changes only the intercept. The estimate of the shape parameter alpha is -2.72 with a 95% confidence interval of $[-3.98, -1.46]$. The confidence interval does not include 0, corresponding to the normal regression, which agrees with our earlier findings. Also note that the scale parameter omega is now reported instead of the standard deviation sigma.

Similarly to figure 4, we can use the skewrplot command to plot the residual density estimate obtained nonparametrically against that from the skew-normal distribution evaluated at the MLEs of the model parameters, as shown in figure 6:

```
. skewrplot, kdensity
```

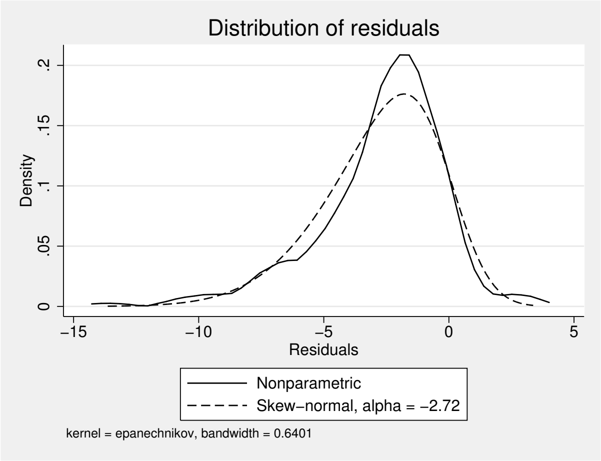

Figure 6. Skew-normal residuals density estimate

Figure 6 demonstrates an improved fit to the distribution of residuals.

Alternatively, we can obtain a Q–Q or P–P plot by using the respective options. For example,

```
. skewrplot, qq
```

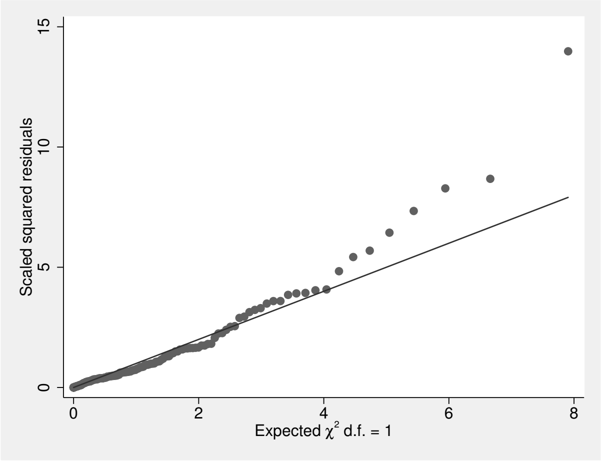

Figure 7. Q–Q plot for the skew-normal model

produces the Q–Q plot of quantiles of the scaled squared residuals from the fitted skew-normal model against the quantiles of the chi-squared distribution with 1 degree of freedom, as shown in figure 7.

According to the Q–Q plot, the skew-normal model fits the data reasonably well, with the exception of several outlying observations in the right tail. See Dalla Valle (2007) for a formal test of the skew-normality in a population.

Next we store estimation results from the skew-normal regression for later comparison with other models. We store results in the DP metric by using the `postdp` option on replay:

```
. skewnreg, postdp
. estimates store skewn_dp
```

To accommodate heavier tails in addition to skewness, we fit the skew-t model:

```
. skewtreg lbm i.female##c.(weight_c height_c), nolog
Skew-t regression                                 Number of obs   =        202
                                                  Wald chi2(5)    =    7955.92
Log likelihood = -450.12502                       Prob > chi2     =     0.0000

------------------------------------------------------------------------------
         lbm |      Coef.   Std. Err.      z    P>|z|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
    1.female |  -8.184854   .3913878   -20.91   0.000     -8.95196   -7.417748
    weight_c |   .7583558   .0300931    25.20   0.000     .6993743    .8173372
    height_c |   12.03037    3.83344     3.14   0.002     4.516964    19.54377
             |
     female#|
  c.weight_c |
          1  |  -.1677404   .0375606    -4.47   0.000    -.2413579   -.0941229
             |
     female#|
  c.height_c |
          1  |  -6.352142   5.232898    -1.21   0.225    -16.60843    3.904149
             |
       _cons |   70.14246   .3307072   212.10   0.000     69.49429    70.79063
-------------+----------------------------------------------------------------
       alpha |  -1.760172   .6463594    -2.72   0.006    -3.027013    -.493331
-------------+----------------------------------------------------------------
       omega |   2.318537   .3619959                       1.70732    3.148569
-------------+----------------------------------------------------------------
          df |   3.658399   1.128259                      1.998842    6.695817
------------------------------------------------------------------------------
LR test vs normal regression:  chibar2(1_2) =   32.02 Prob >= chibar2 = 0.0000
```

As mentioned in section 3.3, the centered parameterization for the skew-t model is still under development and has not yet appeared in the literature. Thus the `skewtreg` command reports results only in the DP metric. Compared with the output of DPs from `skewnreg`, the `skewtreg` command reports an additional estimate of the degrees of freedom. The estimate of the degrees of freedom is 3.66 with a 95% confidence interval of $[2.00, 6.70]$, which implies heavier-than-normal tails for the conditional distribution of `lbm`. The estimate of the shape parameter `alpha` is -1.76 with a 95% confidence interval of $[-3.03, -0.49]$. Again the reported likelihood-ratio test rejects the hypothesis of normality. The reported test of $H_0\colon \alpha = 0, \nu = \infty$ requires a boundary correction because the degrees-of-freedom parameter is tested at its boundary value. As such, the distribution of the likelihood-ratio test statistic is a 50:50 percent mixture of chi-squared distributions with 1 and 2 degrees of freedom, labeled as `chibar2(1_2)` in the output; see, for example, Gutierrez, Carter, and Drukker (2001) and DiCiccio and Monti (2009) for more details.

We can also perform the likelihood-ratio test of the skew-*t* model versus the skew-normal model ($H_0\colon \nu = \infty$) by using the `lrtest` command. 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.

```
. lrtest skewn ., force
Likelihood-ratio test                                 LR chi2(1)  =     14.84
(Assumption: skewn nested in .)                       Prob > chi2 =    0.0001
```

The likelihood-ratio test favors the skew-*t* model over the skew-normal model. The results from this test should be interpreted with caution because it does not automatically account for the fact that the degrees of freedom ν are tested at the boundary value $\nu = \infty$. The distribution of the likelihood-ratio test statistic in this case is a 50:50 percent mixture of a degenerate distribution at 0 and a chi-squared distribution with 1 degree of freedom. As such, the corrected *p*-value is half the uncorrected *p*-value and is 0.000058 in this example:

```
. display r(p)/2
.00005841
```

We can also compare the two fits visually using, for example, a Q–Q plot. We use `skewrplot, qq` to obtain the Q–Q plot of residuals after `skewtreg`:

(*Continued on next page*)

```
. skewrplot, qq
```

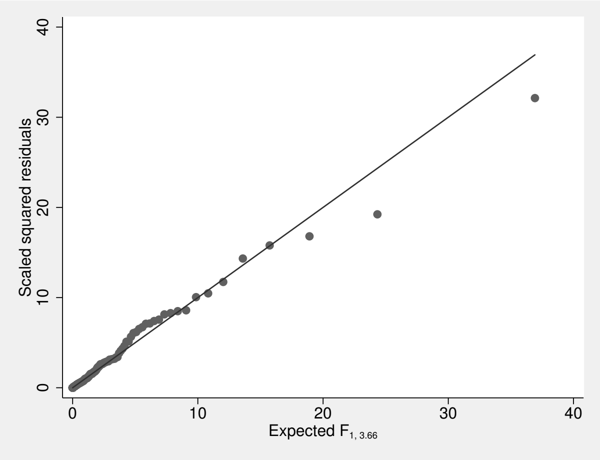

Figure 8. Q–Q plot for the skew-t model

According to figures 7 and 8, the skew-t model fits the `lbm` regression better than the skew-normal model.

Alternatively, we can use information criteria to compare the two models:

```
. estimates stats skewn .

-----------------------------------------------------------------------------
       Model |    Obs    ll(null)   ll(model)     df          AIC         BIC
-------------+---------------------------------------------------------------
       skewn |    202           .   -457.5467      8     931.0933    957.5595
           . |    202           .    -450.125      9       918.25    948.0244
-----------------------------------------------------------------------------
               Note:  N=Obs used in calculating BIC; see [R] BIC note
```

Both Akaike's information criterion and Schwarz's Bayesian information criterion are smaller for the skew-t model, which suggests that it is preferable to the skew-normal model.

We can also compare results from all three regressions, including the normal regression, side-by-side by using `estimates table`.

Because there is no CP parameterization 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
```

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

```
. estimates table reg skewn_dp skewt_dp, equation(1) star(0.05 0.01 0.005) b(%9.3f)

-------------+----------------------------------------------
    Variable |     reg           skewn_dp       skewt_dp
-------------+----------------------------------------------
#1           |
      female |
          1  |    -9.015***      -8.225***      -8.185***
             |
    weight_c |     0.710***       0.774***       0.758***
    height_c |    14.840***       9.915*        12.030***
             |
      female#|
  c.weight_c |
          1  |    -0.177***      -0.196***      -0.168***
             |
      female#|
  c.height_c |
          1  |    -5.443         -3.119         -6.352
             |
       _cons |    68.518***      70.781***      70.142***
-------------+----------------------------------------------
alpha        |
       _cons |                   -2.719***      -1.760**
-------------+----------------------------------------------
omega        |
       _cons |                    3.646***       2.319***
-------------+----------------------------------------------
df           |
       _cons |                                   3.658***
-------------+----------------------------------------------
                      legend: * p<.05; ** p<.01; *** p<.005
```

According to the three regression models, both weight and height are strong predictors of lean body mass measurements. Despite the differences in coefficient estimates, all models lead to similar inferential conclusions. The estimates of the shape parameter `alpha` suggest the presence of negative skewness in the conditional distribution of `lbm` given weight and height. Because tests against zero are not appropriate for the scale and degrees-of-freedom parameters, the significance levels, reported automatically by `estimates table` for these parameters, should be ignored.

5.2 Multivariate analysis of Australian Institute of Sport data

Suppose we are interested in the distribution of `lbm` and `bmi`, the body mass index. In figure 9, the scatterplot 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
```

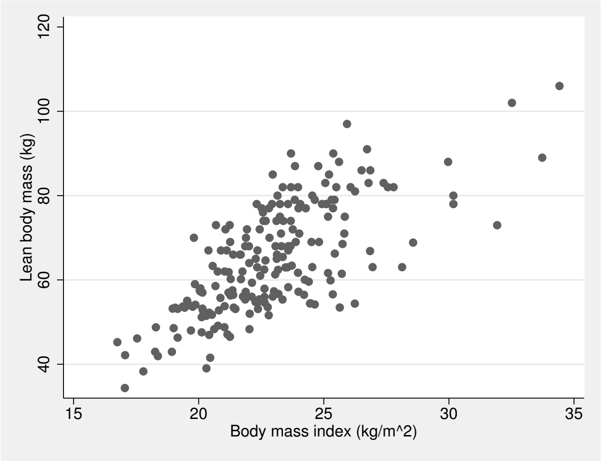

Figure 9. Scatter plot of `lbm` and `bmi`

The scatterplot 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`:

```
. mskewnreg lbm bmi, nolog
Multivariate 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 |   64.92238   .9165846    70.83   0.000     63.12591    66.71886
-------------+----------------------------------------------------------------
bmi          |
       _cons |   22.99999   .1964848   117.06   0.000     22.61489     23.3851
-------------+----------------------------------------------------------------
gamma        |
           1 |   .0061345   .0095526     0.64   0.521    -.0125882    .0248572
           2 |   .4534053   .0936021     4.84   0.000     .2699486     .636862
-------------+----------------------------------------------------------------
Sigma        |
         1 1 |    169.679   16.86076                      139.6514     206.163
         1 2 |   26.31228   3.150039                      20.13832    32.48624
         2 2 |   7.910783   .8210286                      6.454709    9.695323
------------------------------------------------------------------------------
LR test vs MVN regression:           chi2(2) =    37.55   Prob > chi2 = 0.0000
```

By default, mskewnreg reports results in the CP metric. The estimate of the skewness parameter for lbm is close to zero, and according to the z-test ($p = 0.521$), the hypothesis of H_0: $\gamma_1 = 0$ cannot be rejected. For bmi, however, there is strong evidence that the skewness parameter is different from zero. The joint test of H_0: $\gamma_1 = 0, \gamma_2 = 0$ (see below) and the reported likelihood-ratio test strongly reject the hypothesis of bivariate normality for lbm and bmi.

```
. mskewnreg, postcp
. test [gamma1]_cons [gamma2]_cons
 ( 1)  [gamma1]_cons = 0
 ( 2)  [gamma2]_cons = 0
           chi2(  2) =   52.18
         Prob > chi2 =    0.0000
```

To test CPs with mskewnreg, we first need to post CP estimates and their VCE to e(b) and e(V) using the postcp option. By default, mskewnreg saves parameters and their VCE in the estimation metric, which is described in Azzalini and Capitanio (2003) for the multivariate skew-t distribution.

We can also obtain the results in the DP metric by using the dpmetric option:

```
. mskewnreg lbm bmi, dpmetric nolog
Multivariate 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 |   61.76118    1.86054    33.20   0.000     58.11459    65.40777
-------------+----------------------------------------------------------------
bmi          |
       _cons |   20.13548   .2921862    68.91   0.000     19.56281    20.70816
-------------+----------------------------------------------------------------
alpha        |
           1 |   -2.30218   .5772141    -3.99   0.000    -3.433499   -1.170861
           2 |   5.515335   1.301097     4.24   0.000     2.965232    8.065439
-------------+----------------------------------------------------------------
Omega        |
         1 1 |   179.6722   21.30181                      142.4174    226.6725
         1 2 |   35.36759    7.52879                      20.61143    50.12375
         2 2 |   16.11622   2.299581                      12.18449    21.31664
------------------------------------------------------------------------------
LR test vs MVN regression:           chi2(2) =    37.55   Prob > chi2 = 0.0000
```

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 mentioned in section 3.3, 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.

We can compare the fit against the normal model by using, for example, a Q–Q plot:

```
. skewrplot, qq normal
```

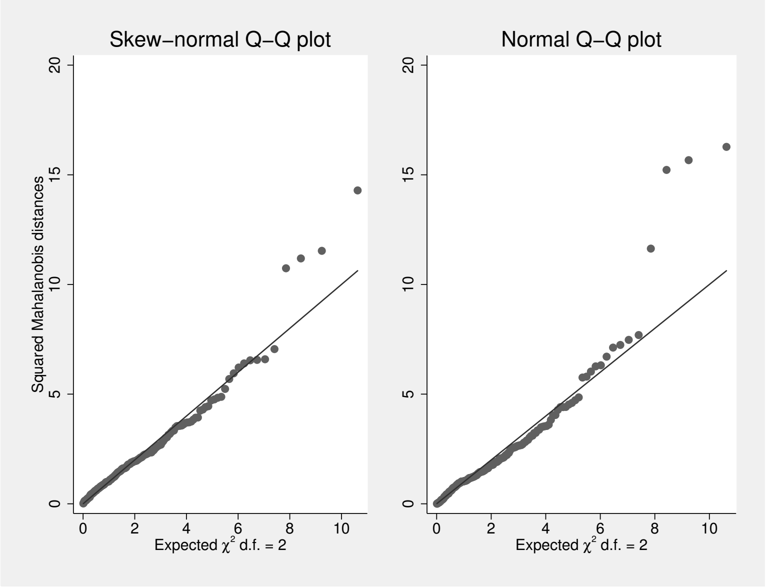

Figure 10. Q–Q plot for bivariate skew-normal and normal model

Figure 10 shows that the bivariate skew-normal model fits the data better than the bivariate normal model.

We can also fit the bivariate skew-*t* model:

```
. mskewtreg lbm bmi, nolog

Multivariate 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   .3165282    63.81   0.000     19.57748    20.81825
-------------+----------------------------------------------------------------
alpha        |
           1 |  -2.234864   .5836011    -3.83   0.000    -3.378702   -1.091027
           2 |   5.242386   1.355911     3.87   0.000      2.58485    7.899922
-------------+----------------------------------------------------------------
Omega        |
         1 1 |   171.7734   24.33629                          130.1249    226.7521
         1 2 |   32.63323     8.5462                          15.88298    49.38347
         2 2 |    14.8864   3.046903                          9.967092    22.23366
-------------+----------------------------------------------------------------
          df |   51.00171   95.45806                          1.301432    1998.702
------------------------------------------------------------------------------
LR test vs MVN regression:     chibar2(2_3) =    37.86 Prob >= chibar2 = 0.0000
```

The estimated degrees of freedom are large, which suggests that the skew-normal model is sufficient for modeling `lbm` and `bmi`.

We can also adjust the location for gender by including `female` as a regressor:

```
. mskewnreg lbm bmi = female, nolog

Multivariate skew-normal regression             Number of obs   =        202
                                                Wald chi2(1)    =     314.13
Log likelihood = -1105.0246                     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   .8219865    91.30   0.000     73.43443    76.65656
-------------+----------------------------------------------------------------
bmi          |
      female |  -2.267239   .3202246    -7.08   0.000    -2.894868   -1.639611
       _cons |   24.13093   .2413993    99.96   0.000     23.65779    24.60406
-------------+----------------------------------------------------------------
gamma        |
           1 |   .1037418   .0543517     1.91   0.056    -.0027856    .2102692
           2 |   .6843178   .0915305     7.48   0.000     .5049213    .8637143
-------------+----------------------------------------------------------------
Sigma        |
         1 1 |   71.51098   7.115973                          58.83973    86.91101
         1 2 |   16.63504   2.002173                          12.71085    20.55923
         2 2 |   6.954864   .7538472                          5.623747    8.601051
------------------------------------------------------------------------------
LR test vs MVN regression:          chi2(2) =    35.63   Prob > chi2 = 0.0000
```

We could also fit separate regressions for males and females to allow all parameters of the joint distribution to vary across gender.

5.3 Log-skew-normal and log-skew-t distributions for modeling positive data

The lognormal and log-*t* distributions are often used to model data such as precipitation data or income data that have a positive support. 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, following Marchenko and Genton (2010). The data are publicly available from the National Climatic Data Center, the largest archive of weather data, and include monthly precipitation measured in inches for the period of 1895–2007 (113 observations per month). The national values could be viewed as weighted averages of station data. More specifically, national values are obtained from the regional values weighted by area. The regional values for each of the nine U.S. climatic regions are computed from the statewide values (which are obtained from the divisional values weighted by area) weighted by area. The divisional monthly precipitation data are obtained as monthly equally weighted averages of values reported by all stations within a climatic division.

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:

```
. use precip07_national
(Precipitation (inches), national U.S. data)

. generate lnprecip = ln(precip)

. skewnreg lnprecip if month==1, nolog

Skew-normal regression                            Number of obs   =        113
                                                  Wald chi2(0)    =          .
Log likelihood =  .71065091                       Prob > chi2     =          .

------------------------------------------------------------------------------
    lnprecip |      Coef.   Std. Err.      z    P>|z|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
       _cons |   .7651154   .0228328    33.51   0.000     .7203639    .8098669
-------------+----------------------------------------------------------------
       gamma |  -.3321967   .1894122    -1.75   0.079    -.7034378    .0390445
-------------+----------------------------------------------------------------
       sigma |   .2428148   .0168615                      .2119171    .2782174
------------------------------------------------------------------------------
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. Below we use the `statsby` command to collect the estimates of skewness indexes and their respective standard errors from `skewnreg` over months and plot them along with their associated 95% confidence intervals (also see Cox [2010] for more examples of `statsby`):

```
. statsby gamma=_b[gamma:_cons] se_gamma=_se[gamma:_cons], by(month) clear:
> skewnreg lnprecip
(running skewnreg on estimation sample)

      command:  skewnreg lnprecip
        gamma:  _b[gamma:_cons]
     se_gamma:  _se[gamma:_cons]
           by:  month

Statsby groups
────┼─── 1 ───┼─── 2 ───┼─── 3 ───┼─── 4 ───┼─── 5
............

. generate lb = gamma-1.96*se_gamma

. generate ub = gamma+1.96*se_gamma

. twoway (line gamma month, sort) (rcap ub lb month, sort), yline(0) xtitle("")
> ytitle("Skewness index") legend(off)  xlabel(1(1)12, valuelabel angle(45))
```

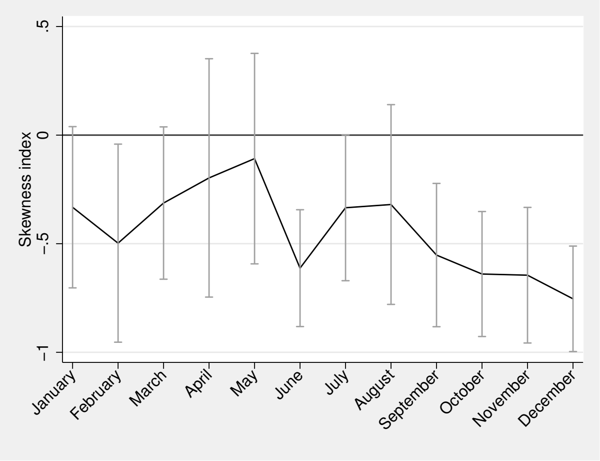

Figure 11. Skewness indexes over months with 95% confidence intervals

From figure 11, we can see that the assumption of the symmetry of the distribution of the log-precipitation is questionable for some months (for example, September, and October). We can see that the distribution of the log precipitation is negatively skewed for summer and fall months and becomes more symmetric in early spring. Similarly, we can investigate the trend in the tails of the distribution over months by plotting the estimated degrees of freedom from `skewtreg`.

6 Conclusion

In this article, we described two flexible parametric models, the skew-normal and skew-t models, which can be used for the analysis of nonnormal data. We presented a suite of commands for fitting these models in Stata to univariate and multivariate data. We also provided postestimation features for obtaining linear predictions and for graphically evaluating the goodness-of-fit of the skewed distributions to the data. We demonstrated how to use the commands for 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 positive support on the example of U.S. precipitation data.

7 Acknowledgments

The authors thank the referee for valuable comments. M. G. Genton's research was partially supported by NSF grants CMG ATM-0620624 and DMS-1007504. We are grateful to R. B. Arellano-Valle, A. Azzalini, T. DiCiccio, and A. C. Monti for access to preliminary versions of their publications.

8 References

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

———. 2009. Parameters and other summary quantities of the skew-t distribution. Manuscript in preparation.

Arnold, B. C., and R. J. Beaver. 2000. The skew-Cauchy distribution. *Statistics & Probability Letters* 49: 285–290.

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

———. 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., and A. Capitanio. 1999. Statistical applications of the multivariate skew normal distribution. *Journal of the Royal Statistical Society, Series B* 61: 579–602.

———. 2003. Distributions generated by perturbation of symmetry with emphasis on a multivariate skew t-distribution. *Journal of the Royal Statistical Society, Series B* 65: 367–389.

Azzalini, A., T. dal Cappello, and S. Kotz. 2002. Log-skew-normal and log-skew-t distributions as models for family income data. *Journal of Income Distribution* 11: 12–20.

Azzalini, A., and A. Dalla Valle. 1996. The multivariate skew-normal distribution. *Biometrika* 83: 715–726.

Azzalini, A., and M. G. Genton. 2008. Robust likelihood methods based on the skew-t and related distributions. *International Statistical Review* 76: 106–129.

Bayes, C. L., and M. D. Branco. 2007. Bayesian inference for the skewness parameter of the scalar skew-normal distribution. *Brazilian Journal of Probability and Statistics* 21: 141–163.

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

Cook, R. D., and S. Weisberg. 1994. *An Introduction to Regression Graphics*. New York: Wiley.

Cox, N. J. 2010. Speaking Stata: The statsby strategy. *Stata Journal* 10: 143–151.

Dalla Valle, A. 2007. A test for the hypothesis of skew-normality in a population. *Journal of Statistical Computation and Simulation* 77: 63–77.

DiCiccio, T. J., and A. C. Monti. 2009. Inferential aspects of the skew-t distribution. Manuscript in preparation.

Genton, M. G., ed. 2004. *Skew-Elliptical Distributions and Their Applications: A Journey Beyond Normality*. Boca Raton, FL: Chapman & Hall/CRC.

Gutierrez, R. G., S. Carter, and D. M. Drukker. 2001. sg160: On boundary-value likelihood-ratio tests. *Stata Technical Bulletin* 60: 15–18. Reprinted in *Stata Technical Bulletin Reprints*, vol. 10, pp. 269–273. College Station, TX: Stata Press.

Hallin, M., and C. Ley. Forthcoming. Skew-symmetric distributions and Fisher information—A tale of two densities. *Bernoulli*.

Ley, C., and D. Paindaveine. 2010. On the singularity of multivariate skew-symmetric models. *Journal of Multivariate Analysis* 101: 1434–1444.

Liseo, B., and N. Loperfido. 2006. A note on reference priors for the scalar skew-normal distribution. *Journal of Statistical Planning and Inference* 136: 373–389.

Marchenko, Y. V., and M. G. Genton. 2010. Multivariate log-skew-elliptical distributions with applications to precipitation data. *Environmetrics* 21: 318–340.

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.

About the authors

Yulia V. Marchenko is a senior statistician at StataCorp. Her research interests include multiple imputation, survival analysis, skewed multivariate non-Gaussian distributions, and statistical software development.

Marc G. Genton is a professor at the Department of Statistics, Texas A&M University, College Station. His research interests include skewed multivariate non-Gaussian distributions, spatial and spatio-temporal statistics, and robustness.