ACTA UNIVERSITATIS LODZIENSIS
FOLIA OECONOMICA 196, 2006

*Maria Jadamus-Hacura**

SKEW-NORMAL DISTRIBUTION – BASIC PROPERTIES AND AREAS OF APPLICATIONS

Abstract. The skew-normal is a class of distribution that includes the normal distribution as a special case. A systematic treatment of the skew-normal distribution has been given in Azzalini (1985, 1986); generalizations to the multivariate case are given in Azzalini and Capitanio (1999), while Azzalini and Capitanio (2003) propose a further extension with a skew-*t* distribution. In this paper we study the properties of this class of density functions and we investigate the applicability of this distributions for modeling some financial and income data.

Key words: skewness, skew normal distribution, skew-*t* distribution.

1. THE SKEW-NORMAL AND THE SKEW-*T* PROBABILITY DISTRIBUTIONS

A random variable X will be said to have a standard skew-normal distribution $SN(0, 1, \lambda)$ with parameter λ if its density function is of the form:

$$f(x; \lambda) = 2\phi(x)\Phi(\lambda x) \tag{1}$$

where ϕ and Φ denote the standard normal $N(0, 1)$ density and distribution functions respectively. The density is symmetric if $\lambda = 0$ (in this case it coincides with the standard normal density). As a limit case, when λ tends to $+\infty$ $(-\infty)$, X tends to a positive (negative) half normal random variable. It can be shown, that the variable X^2 is distributed as χ_1^2, irrespective of the value of λ.

In applications, it is common to work with transformed variable $Y = \mu + \sigma X$, with $\mu \in R$ and $\sigma > 0$. Hence, the density for the random variable Y, denoted $SN(\mu, \sigma, \lambda)$ is:

$$f(y; \mu, \sigma, \lambda) = \frac{2}{\sigma}\varphi\left(\frac{y-\mu}{\sigma}\right)\Phi\left(\lambda\frac{y-\mu}{\sigma}\right). \tag{2}$$

* Ph.D., Department of Econometrics, The Karol Adamiecki University of Economics, Katowice.

Here λ is the shape parameter, which determines the skewness, and μ and σ represent the usual location and scale parameters. The first four moments of Y are:

$$E(Y) = \mu + \sigma\sqrt{2/\pi\delta},$$

$$\mathrm{Var}(Y) = \sigma^2(1 - 2\delta^2/\pi\pi,$$

$$Y_1 = \frac{\sqrt{2}(4-\pi)}{\pi^{3/2}} \frac{\delta^3}{(1 - 2\delta^2/\pi\pi^{3/2}},$$

$$Y_2 = \frac{8(\pi - 3)}{\pi^2} \frac{\delta^4}{(1 - 2\delta^2/\pi\pi^2},$$

where $\delta = \lambda/\sqrt{1+\lambda^2} \in (-1, 1)$. The skew normal distribution shares particular theoretical properties with the normal distribution and has the advantage of being suitable for the analysis of data with unimodal and skewed empirical distributions.

The graphs in Figure 1 show the effect of different values of λ on the shape of the probability density function. The family of densities (2) was apparently first discussed in detail in Azzalini (1985), generalizations to the multivariate case are given in Azzalini and Dalla Valle (1996), Azzalini and Capitanio (1999), while Azzalini and Capitanio (2003) propose a further extension with a skew-*t* distribution. The skew-*t* distribution is a natural generalization of the skew normal distribution, which allows modeling the tails of the distribution as well as its skewness. Below, we give the definition of Azzalini and Capitanio (2003), although this coincides with the proposal of Branco and Dey (2001). Both papers deal with the general multivariate case. Here, we restrict our attention to the scalar case. Let us consider a random variable $X \sim SN(\lambda)$, with density given by (1), and another random variable V, independent of X, with distribution $V \sim \chi^2_v/v$. Then

$$Y = \mu + \sigma V^{-1/2} X$$

has the skew-*t* distribution, written $Y \sim ST(\mu, \sigma, \lambda, v)$. The density of Y is given by:

$$f(y; \mu, \sigma, \lambda, v) = 2t(y; v)T\left[\lambda \frac{y-\mu}{\sigma}\left(\frac{v+1}{Q_y + v}\right)^{1/2}; v+1\right], \tag{3}$$

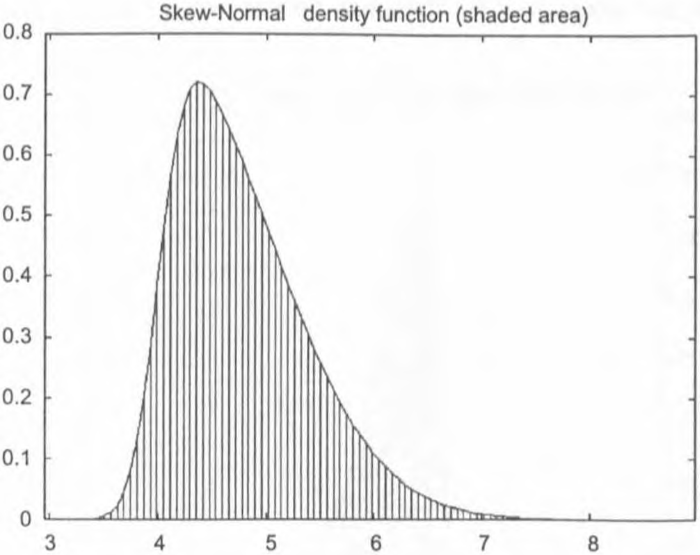
where

$$Q_y = \left(\frac{y-\mu}{\sigma}\right)^2$$

$$t(y; v) = \frac{1}{\sigma} \frac{1}{\sqrt{\pi v}} \frac{\Gamma\{(v+1)/2}{\Gamma(v/2)} \left(1 + \frac{Q_y}{v}\right)^{-(v+1)/2}$$

and $T(x; y_0)$ is the *t*-distribution function with v_0 degrees of freedom.

(a)



(b)

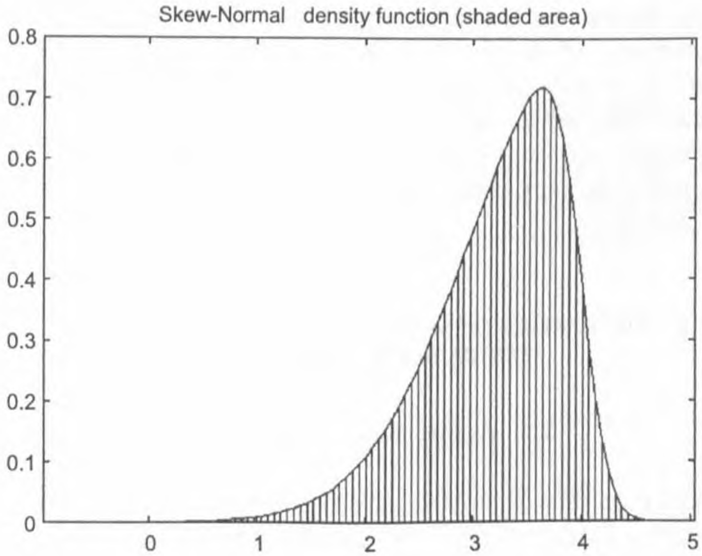


Fig. 1. Skew-normal densities with parameters: (a) $\mu = 0$, $\sigma = 1$, $\lambda = 5$; (b) $\mu = 0$, $\sigma = 1$, $\lambda = -5$

Note that, that as ν goes to infinity, the skew-t distribution tends to the skew-normal distribution. In Figure 2 we observe, that the small value of ν indicates tail much heavier than the skew-normal distribution.

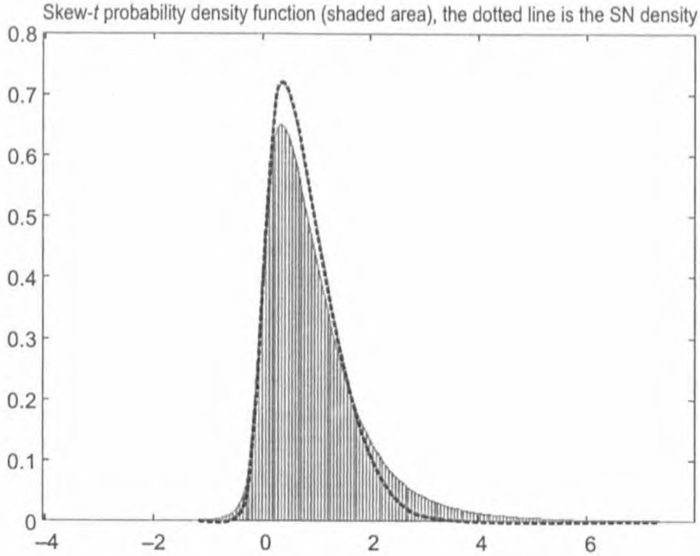


Fig. 2. Skew t-density when: $\mu = 0$, $\sigma = 1$, $\lambda = 5$, $\nu = 4$

For data fitting we can apply software "library sn" witch is available on the homepage of Adelhi Azzalini: http://azzalini.stat.unipd.it/SN/. The "library sn" is a suite of functions for handling skew-normal and skew-t distributions, both in the univariate and the multivariate cases. The available facilities include various standard operations (density function, random number generation, etc), data fitting via MLE, plotting log-likelihood surfaces and others.

2. EXAMPLES OF APPLICATIONS OF THE SKEW-NORMAL AND THE SKEW-*T* PROBABILITY DISTRIBUTIONS

2.1. Family Income Data Models

The family income data are usually fitted using log-normal, Gamma, Singh-Maddala, Dagum type I and generalized Beta of second kind (GB2) distributions. Azzalini et al. (2003) supplement these fittings by adding the log-skew-normal (log-SN) and log-skew-t distributions (log-ST). The value of the SSE (sum of squared errors) multiplied by 10^3, between the fitted and observed values of distributions of the income for several European countries in 1997 are presented in Table 1.

Table 1. Values of SSE, multiplied by 10^3, between the fitted and observed values of income for the chosen European countries in 1997

Country	n	Log N	Gamma	Dagum Type 1	Sinh-Maddala	Log-SN	GB2	Log-ST
Demark	2506	6.249	3.502	6.407	4.136	2.758	3.544	2.709
The Netherlands	4922	7.139	2.126	0.862	1.549	2.159	0.891	0.778
Belgium	2863	1.552	2.592	1.596	1.358	1.263	1.222	1.199
France	5851	1.22	0.867	0.703	0.335	0.307	0.2	0.216
Ireland	2723	4.966	2.343	4.236	3.003	2.001	2.569	2.003
Italy	6478	3.561	0.541	1.053	0.422	0.713	0.425	0.472
Greece	4171	3.721	0.334	0.668	0.23	0.281	0.201	0.280
Spain	5439	0.574	2.478	2.143	1.648	0.410	0.395	0.401
Portugal	4666	4.641	1.118	0.126	0.631	1.037	0.215	1.036
Austria	2952	2.911	0.377	1.233	0.568	0.286	0.459	0.291
Sweden	5788	17.35	6.999	0.793	4.217	4.363	0.551	0.668
Germany	5956	4.415	1.196	0.385	0.553	1.235	0.376	0.397

Source: Azzalini et al. (2003).

The best fitting we have for log-ST and GB2 distributions; which are essentially equivalent.

Both the log-SN and log-ST distribution allow for a multivariate version, whose properties are described in detail in the above cited papers. These multivariate versions would allow consideration of joint distribution of family income.

2.2. Portfolio Selection Models

The returns on most financial assets exhibit kurtosis and many also have probability distributions that exhibit skewness as well. Addock (2002) presents a general multivariate model for the probability distribution of assets returns which incorporates both skewness and kurtosis. The model is based on multivariate skew-Student distribution, whose main properties are applied to the tasks of assets pricing and portfolio selection.

Harvey et al. (2003) propose the use of the skew normal distribution as a characterization of the asset returns. They show that this distribution has many attractive features as far as modeling multivariate returns is concerned. They propose portfolio selection based on the maximization of utility function using Bayesian methods. The authors show that this approach leads to higher expected utility than the resampling methods common in the practice of finance.

3. CONCLUSIONS

It is reasonable to apply the family of skew normal and skew-*t* distribution for modeling data which exhibits skewness and kurtosis. Besides the examples mentioned in the paper the latest findings of Dominguez-Molina et al. (2003) also report the use of skew normality in Stochastic Frontier Analysis where skewness is applied as a measure of technical inefficiency.

REFERENCES

Adock C. J. (2002), *Asset Pricing and Portfolio Selection Based on the Multivariate Skew-Student Distribution.*

Azzalini A. (1985), "A Class of Distributions which Includes the Normal Ones", *Scandinawian Journal of Statistics,* **12**, 171–178.

Azzalini A. (1986), "Further Results on a Class of Distributions which Includes the Normal Ones", *Statistica,* **46**, 199–208.

Azzalini A., Dalla Valle A. (1996), "The Multivariate Skew-normal Distribution", *Biometrika,* **83**, 715–726.

Azzalini A., Capitanio A. (1999), "Statistical Applications of the Multivariate Skew-normal Distribution", *Journal of Royal Statistical Society B*, **61**, 579–602.

Azzalini A., Capitanio A. (2003), "Distributions Generated by Perturbation of Symmetry with Emphasis on a Multivariate Skew-*t* Distribution", *Journal of Royal Statistical Society B,* **65**, 367–389.

Azzalini A., Capello T., Kotz S. (2003), "Log-skew-normal and Log-skew-*t* Distributions as Models for Family Income Data", *Journal of Income Distribution,* **11**, 3–4.

Domínguez-Molina J. A., González-Farías G., Ramos-Quiroga R. (2003), "Skew-Normality in Stochastic Frontier Analysis", *Comunicación Técnica*, I-03-18/06-10.

Harvey C. R., Liechty J. C., Liechty M. W. (2003), "Portfolio Selection with Higher Moments", *Working Paper*, Duke University.

Maria Jadamus-Hacura

SKOŚNY ROZKŁAD NORMALNY – PODSTAWOWE WŁASNOŚCI I OBSZARY ZASTOSOWAŃ

(Streszczenie)

Klasa skośnych rozkładów normalnych zawiera jako szczególny przypadek rozkład normalny. Szczegółowemu omówieniu własności rozkładu skośnego normalnego poświęcona jest praca Azzalini (1985, 1986); przypadek wielowymiarowy przedstawili Azzalini i Capitanio (1999), natomiast w pracy tych autorów z roku 2003 można znaleźć dalsze rozszerzenie tej klasy rozkładów o rozkłady skośne *t*-Studenta. W niniejszym artykule przedstawiono podstawowe własności funkcji gęstości omawianych rozkładów i pokazano możliwości ich wykorzystania w modelowaniu dochodów i danych finansowych.