Communications of the Korean Statistical Society Vol. 15, No. 2, 2008, pp. 161–171

Further Applications of Johnson's S_U-normal Distribution to Various Regression Models

Pilsun Choi[1], Insik Min[2]

Abstract

This study discusses Johnson's S_U-normal distribution capturing a wide range of non-normality in various regression models. We provide the likelihood inference using Johnson's S_U-normal distribution, and propose a likelihood ratio (LR) test for normality. We also apply the S_U-normal distribution to the binary and censored regression models. Monte Carlo simulations are used to show that the LR test using the S_U-normal distribution can be served as a model specification test for normal error distribution, and that the S_U-normal maximum likelihood (ML) estimators tend to yield more reliable marginal effect estimates in the binary and censored model when the error distributions are non-normal.

Keywords: S_U-normal distribution; skewness and kurtosis; normality test.

1. Introduction

The normality assumption for error distribution has played a dominant role in both theoretical and applied economics. It is, however, obvious that many real data encountered in economics and finance show significant deviations from the normal distribution. In the field of time series analysis, it is now well-documented that the conditional normal distribution is unable to provide an appropriate description of the statistical properties of asset returns in terms of skewness and excess kurtosis. The studies motivated by inadequateness of the normal assumption proposed a variety of parametric distributions to capture the asymmetry and fat tails of asset returns (Bollerslev, 1987; Theodossiou, 1998, 2000; Rockinger and Jondeau, 2001 to name a few).

In the field of limited dependent variable models as well, several methods of maximum likelihood (ML) estimates with more flexible distributions have been introduced. Among others, McDonald and Yexiao (1996) and McDonald (1996) proposed several skewed and leptokurtic parametric distributions, and compared the performance of some flexible parametric and semi-parametric methods in estimating binary and censored regression models. The main benefit of using a more flexible parametric distribution is that it

1) Assistant Professor, Department of International Trade, Konkuk University, Seoul 143-701, Korea.
2) Assistant Professor, Department of Economics, Kyung Hee University, Seoul 130-701, Korea.
Correspondence: imin@khu.ac.kr

can provide likelihood inferences and thus consistent estimators, and can account for skewness and excess kurtosis.

In this paper, we discuss Johnson's (1949) S_U-normal distribution capturing a wide range of non-normality in various regression models. The S_U-normal distribution defined as an inverse hyperbolic transformation of a normal random variable is one of the most flexible distribution functions in accommodating asymmetry and excess kurtosis of error distribution. In addition, this distribution shares some convenient properties of the normal distribution such as the easy derivations of quantiles, cumulative probabilities and joint density function.

In spite of the flexibility and simplicity, the S_U-normal distribution had not been introduced in the field of econometrics until Choi and Nam (2007) and Choi (2002) applied the density to modeling multivariate generalized autoregressive conditional heteroskedasticity (GARCH) models. Recently the S_U-normal distribution has also been applied to the estimation of switching regression models by Choi and Min (2007). In this paper, we attempt further applications of the S_U-normal distribution to various regression models including limited dependent variable models. We provide likelihood inferences using Johnson's S_U-normal distribution, and propose a likelihood ratio (LR) test for normality. We also apply the S_U-normal distribution to the binary and censored regression models. Monte Carlo simulations are used to show that the LR test using the S_U-normal distribution can be served as a model specification test for normal error distribution, and that the S_U-normal ML estimators tend to yield more reliable marginal effect estimates in the binary and censored model when the error distributions are non-normal.

The remainder of this paper is organized as follows. In Section 2, we provide the likelihood inference using Johnson's S_U-normal distribution. Section 3 proposes a likelihood ratio test for normality and ML estimates for the binary and censored regression models based on the S_U-normal distribution. Section 4 concludes the paper.

2. Johnson's S_U-normal distribution

Johnson (1949) introduced a flexible parametric distribution by simply transforming a normal random variable. The distribution of a random variable Y is considered to be S_U-normal, if

$$\sinh^{-1}(Y) = \lambda + \theta Z,$$

where Z is a standard normal random variable, $\lambda \in \Re, \theta \in \Re^+$ and $-\infty < Y < \infty$. $\sinh^{-1}(Y)$ denotes the inverse hyperbolic sine of Y. The S_U-normal distribution has the density

$$f(Y;\lambda,\theta) = \frac{1}{\sqrt{2\pi\theta^2(Y^2+1)}}\exp\left[-\frac{\left(\sinh^{-1}(Y)-\lambda\right)^2}{2\theta^2}\right]. \quad (2.1)$$

The mean and variance of Y are given as, respectively, $m = w^{1/2}\sinh(\lambda)$ and $s^2 = (1/2)(w-1)\{w\cosh(2\lambda)+1\}$, where $w = \exp(\theta^2)$ (See Appendix for detailed derivations).

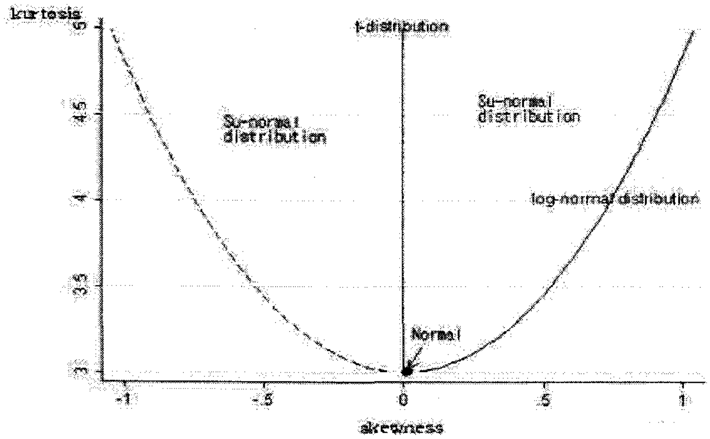

Figure 2.1: Flexibility of the S_U-normal distribution

Figure 2.1 shows the flexibility of the S_U-normal distribution. In the skewness and kurtosis plane, one can depict the area that each distribution may cover as its skewness and kurtosis coefficients. The normal distribution can cover only one point in this plane ($skewness = 0$ and $kurtosis = 3$) and the log-normal distribution covers the upward-sloping line in the figure. The Student's t distribution is symmetric and may have a kurtosis coefficient that lies between three and infinity. The S_U-normal distribution can cover all kurtosis larger than that covered by a log-normal distribution. It implies that for any thicker tail distribution that is a log-normal distribution, there exists an appropriate S_U-normal distribution that covers a wide range of negative skewness with excess kurtosis.

3. Further Applications

3.1. Likelihood inferences for S_U-normal distribution model

Suppose that data consist of $\{y_i, x_i\}$ and an *i.i.d.* sample drawn from the distribution of (y, x) with the joint density function. Consider a regression model with S_U-normal ($S_U N$) error term,

$$y_i = x_i\beta + u_i, \quad u_i \sim S_U N(0, \sigma^2; \lambda, \theta).$$

Therefore, the conditional distribution of $(y_i|x_i) \sim S_U N(x_i\beta, \sigma^2; \lambda, \theta)$ and the corresponding conditional density function is given by

$$
\begin{aligned}
f(y_i;x_i,\sigma) &= \frac{s}{\sigma}f_y\left\{\frac{s}{\sigma}(y_i - x_i\beta) + m\right\} \\
&= \frac{s/\sigma}{\sqrt{2\pi\theta^2\left[\left\{\frac{s}{\sigma}(y_i - x_i\beta)+m\right\}^2+1\right]}} \\
&\quad \times \exp\left[-\frac{\left(\sinh^{-1}\left\{\frac{s}{\sigma}(y_i - x_i\beta)+m\right\}-\lambda\right)^2}{2\theta^2}\right],
\end{aligned}
$$

where $\delta = (\beta, \sigma, \lambda, \theta)$ is the parameter set and $f_Y(\cdot)$ is the density function provided in (2.1). The log-likelihood function is constructed as follows

$$
\begin{aligned}
\ln L &= \sum_{i=1}^{n} \ln f(y_i;x_i,\delta) \\
&= \sum_{i=1}^{n}\left[\ln s - \ln\sigma - \frac{1}{2}\ln(2\pi) - \ln\theta - \frac{1}{2}\ln(U_i^2+1) - \frac{\{\sinh^{-1}(U_i)-\lambda\}^2}{2\theta^2}\right],
\end{aligned}
$$

where $U_i = s/\sigma(y_i - x_i\beta) + m$. The score function is $\sum_{i=1}^{n} S_\delta(y_i;x_i,\delta)$, where $S_\delta = \partial \ln f(y_i;x_i,\delta)/\partial\delta = (S_\beta, S_\sigma, S_\lambda, S_\theta)$. Thus $\hat{\delta} = (\hat{\beta}, \hat{\sigma}, \hat{\lambda}, \hat{\theta})$ is the ML estimates satisfying $\sum_{i=1}^{n} S_\delta(y_i;x_i,\delta) = 0$. Considering the space restriction, we provide the score function only with respect to β in this section.

$$
S_\beta = \frac{\partial \ln f(y_i;x_i,\delta)}{\partial\beta} = -\frac{U_i}{U_i^2+1}\frac{\partial U_i}{\partial\beta} - \frac{\{\sinh^{-1}(U_i)-\lambda\}}{\theta^2}\frac{1}{\sqrt{1+U_i^2}}\frac{\partial U_i}{\partial\beta},
$$

where $\partial U_i/\partial\beta = -(s/\sigma)x_i$.[3)]

3.2. Model specification test for normality

In many economic applications, the distribution of one variable conditional on some other variables is of particular interest. Using Student's t distribution, Bollerslev (1987) suggested an LR test against null hypothesis of conditional normal errors in the GARCH model. Semiparametric and nonparametric tests of conditional distributions are also suggested by Zheng (2000), Andrews (1997), Klaauw and Konig (2003) and Fan *et al.* (2006).

In this subsection we propose a parametric test of error distributions using the SU-normal distribution. In fact, Bollerslev's LR test can not be a consistent one in case where the alternative error distribution is sharply skewed. The LR test statistic in this study is based on the SU-normal distribution that can capture the skewness and excess kurtosis.

3) The other score functions are available upon request.

The data generating process (DGP) for null model is a linear regression model with normal homoskedastic errors:

$$DGP_0 : y_i = \beta_0 + \beta_1 x_i + u_i,$$

where $\{x_i\}_{i=1}^n$ is a random sample from $U(-2,2)$ and the error term $\{u_i\}$ follows *i.i.d.* $N(0,\sigma^2)$. Moreover, x_i and u_i are independent of each other. The true parameters are $\delta \equiv \{\beta_0, \beta_1, \sigma\} = \{1,1,1\}$. The null and alternative hypotheses are, respectively

$$H_0 : f(y|x,\delta) = \phi\left[(y - \beta_0 - \beta_1 x)/\sigma\right]/\sigma \quad \text{and}$$
$$H_1 : f(y|x,\delta) \text{ belongs to a non-normal distribution.}$$

The three DGPs for H1 are considered as follows:[4]

$$\begin{aligned} DGP_1^a &: y_i = \beta_0 + \beta_1 x_i + u_i, \quad u_i \sim t(5), \\ DGP_2^a &: y_i = \beta_0 + \beta_1 x_i + u_i, \quad u_i \sim S_U N(0,1;\lambda = 1, \theta = 0.3), \\ DGP_3^a &: y_i = \beta_0 + \beta_1 x_i + u_i, \quad u_i \sim skewed-t(0,1;\eta = 5, \lambda = 0.2), \end{aligned}$$ [5]

The proposed LR test statistic is written as

$$LR = -2(\log L_{Normal} - \log L_{S_U N}), \tag{3.1}$$

where $\log L_{Normal}$ and $\log L_{S_U N}$ are the log-likelihood function values estimated under the assumption of normal and S_U-normal error distributions, respectively. Because the S_U-normal converges to the normal distribution as the shape parameter $\theta \to 0$, the distribution of LR test statistic is expected to deviate from a $\chi^2(1)$ distribution; thus, the test statistic is more concentrated towards the origin than in case of a $\chi^2(1)$ distribution (see Bollerslev, 1987). For empirical critical values, we employ the parametric bootstrap procedure that mimics the sample distribution of the test statistic in (3.1).[6]

Monte Carlo evidence reported in Table 3.1 indicates that the simulated rejection probabilities for the LR test based on (3.1) closely replicates the nominal sizes of 10%, 5% and 1% under the null hypothesis. Moreover, the simulated power of the test increases as the sample size increases when the data generating process moves away from the null model (DGP_0). With regard to the alternative DGP_1^a with a symmetric and leptokurtic t distribution, we observe that our test becomes powerful and thus consistent as the sample size is $n = 400$. For the alternative models with an asymmetric and leptokurtic distribution, namely, DGP_2^a and DGP_3^a, our test is still consistent in showing that the simulated rejection rates are asymptotically one.

4) Although we provide the simulation results only for one special set of parameter values in the paper, we obtained basically similar results even when we used different set of parameter values for each alternative DGP.
5) We adopt Hansen (1994) type skewed-t distribution.
6) The number of simulations is 1,000 and the numbers of bootstrapping are 1,000 for size estimation and 500 for power estimation.

Table 3.1: Simulated sizes and powers of the normality test

	$n = 100$			$n = 200$			$n = 300$		
	1%	5%	10%	1%	5%	10%	1%	5%	10%
DGP_0 (size)	0.9	5.5	11.7	1.3	5.5	11.3	0.9	5.3	10.1
DGP_1^a (power)	48.3	67.1	74.9	74.5	86.7	91.0	96.5	98.3	99.1
DGP_2^a (power)	52.5	73.0	81.4	86.1	94.5	97.8	99.4	99.9	100
DGP_3^a (power)	55.0	74.1	79.7	88.7	93.3	97.0	99.3	99.8	99.8

3.3. Binary regression model

The binary dependent variable model with the S_U-normal error distribution is given by

$$\begin{aligned} y_i^* &= x_i\beta + \varepsilon_i, \quad \varepsilon_i \sim S_U N(0, \sigma^2; \lambda, \theta), \\ y_i &= \begin{cases} 1, \text{ if } y_i^* > 0, \\ 0, \text{ if } y_i^* \leq 0, \end{cases} \end{aligned}$$

where y_i is observed instead of the latent variable y_i^*. It is noted that $z_i = \{\sinh^{-1}(\varepsilon_i + m) - \lambda\}/\theta$ or $\varepsilon_i = \{\sinh(\lambda + \theta z_i) - m\}$ to make the zero-mean distribution, where z_i is a standard normal random variable. Further, we can derive the following:

$$\begin{aligned} \Pr[\varepsilon_i \leq -x_i\beta] &= \Pr[\{\sinh(\lambda + \theta z_i) - m_i\} \leq -x_i\beta] \\ &= \Pr[z_i \leq \{\sinh^{-1}(m - x_i\beta) - \lambda\}/\theta] = \Phi(a_i^*), \end{aligned}$$

where $\Phi(\cdot)$ and is the standard normal *cdf*. As a result, the likelihood function of the model is constructed as

$$\log L = \sum_{y_i=0} \log\{\Phi(a_i^*)\} + \sum_{y_i=1} \log\{1 - \Phi(a_i^*)\}.$$

The marginal effect of the probability that $y_i = 1$ is

$$\frac{\partial \Pr[y_i = 1|x_i, \beta, \lambda, \theta]}{\partial x_k} = \frac{\partial[1 - \Phi(a^*)]}{\partial x_k} = -\frac{\partial a^*}{\partial x_k}\phi(a^*) = \frac{\beta_k}{\theta}\frac{1}{\sqrt{1 + (m - x_i\beta)^2}}\phi(a^*).$$

We design the following binary dependent model for the simulation:

$$y_i^* = \beta_0 + \beta_1 x_i + \varepsilon_i, \quad i = 1, \ldots, n,$$

where true parameters are $\{\beta_0, \beta_1\} = \{-5, 1\}$. The exogenous variable x_i is uniformly distributed on [0,10]. We choose two sample sizes ($n = 500$ and $n = 1000$). The error terms are drawn from three underlying distributions: a standard normal distribution, a t-distribution with degrees of freedom parameter 4 as a symmetric and thick tail distribution, a standardized S_U-normal distribution with $\lambda = 1$ and $\theta = 1$ as the asymmetric and leptokurtic distributions.[7)]

7) We obtained basically similar results when we used different set of parameter values.

Table 3.2: Monte Carlo simulation results for binary model

(1) $\varepsilon_i \sim N(0,1)$

	True value	Probit		S_U-normal ML	
		$n = 500$	$n = 1000$	$n = 500$	$n = 1000$
$ME(x = 3)$	0.054	0.054(0.020)	0.052(0.013)	0.053(0.020)	0.051(0.012)
$ME(x = 5)$	0.399	0.401(0.043)	0.402(0.026)	0.407(0.047)	0.407(0.027)
$ME(x = 7)$	0.054	0.053(0.018)	0.053(0.012)	0.050(0.020)	0.050(0.012)
$\log L$		−89.52	−178.48	−89.29	−176.22

(2) $\varepsilon_i \sim t(4)$

	True value	Probit		S_U-normal ML	
		$n = 500$	$n = 1000$	$n = 500$	$n = 1000$
$ME(x = 3)$	0.034	0.056(0.031)	0.058(0.029)	0.042(0.021)	0.041(0.016)
$ME(x = 5)$	0.530	0.398(0.141)	0.392(0.142)	0.485(0.129)	0.497(0.099)
$ME(x = 7)$	0.034	0.055(0.031)	0.056(0.027)	0.039(0.020)	0.038(0.015)
$\log L$		−88.37	−179.07	−86.29	−172.95

(3) $\varepsilon_i \sim S_U N(0, 1; \lambda = 1, \theta = 1))$

	True value	Probit		S_U-normal ML	
		$n = 500$	$n = 1000$	$n = 500$	$n = 1000$
$ME(x = 3)$	0.030	0.055(0.038)	0.056(0.034)	0.029(0.010)	0.030(0.006)
$ME(x = 5)$	0.521	0.410(0.130)	0.410(0.120)	0.515(0.122)	0.5140(0.086)
$ME(x = 7)$	0.002	0.043(0.046)	0.045(0.046)	0.002(0.005)	0.002(0.004)
$\log L$		−82.34	−166.78	−71.41	−141.48

Note: $ME(x = a)$ indicates the marginal effect at $x = a$. The value in parenthesis is the root mean squared error.

Table 3.2 provides the results for simulation with the number of simulation 500. For normally distributed errors, the Probit estimators slightly dominate the S_U-normal MLE. As the sample size is $n = 1000$, the marginal effects and root mean squared error (RMSE) from the S_U-normal ML model are similar to those from the Probit model.

For a symmetric and leptokurtic distribution, however, we observe that the S_U-normal ML model obviously dominates the Probit model in terms of the precision and RMSE of estimates at the given points. Particularly, at the tail distribution, the reliability of marginal effects from the S_U-normal model outperforms the Probit model. When the true distribution is the S_U-normal distribution, we expect that the non-normality with asymmetry and excess kurtosis leads to imprecise marginal effect estimates and a serious loss of efficiency in the Probit model. Meanwhile, the S_U-normal ML model seems to sufficiently accommodate the deviations from the normality.

3.4. Censored regression model

The censored model with the S_U-normal error distribution is given by

$$y_i^* = x_i\beta + \varepsilon_i, \quad \varepsilon_i \sim S_U N(0, \sigma^2; \lambda, \theta),$$

where we observe $y_i = \begin{cases} y_i^*, \text{ if } y_i^* > 0, \\ 0, \text{ if } y_i^* \leq 0 \end{cases}$. Note that $\varepsilon_i = \sigma\{\sinh(\lambda + \theta z_i) - m\}/s$, where z_i is a standard normal random variable and by applying the change of variable, $z_i = [\sinh^{-1}\{s(y_i - x_i\beta)/\sigma + m\} - \lambda]/\theta$. The log-likelihood function is given by

$$\begin{aligned} \log L &= \sum_{y_i=0} \log(\Pr[y_i = 0]) + \sum_{y_i>0} \log(f(y_i|y_i > 0)\Pr[y_i > 0]) \\ &= \sum_{y_i=0} \log(\Pr[y_i^* \leq 0]) + \sum_{y_i=1} \log(f(y_i)). \end{aligned}$$

Under the S_U-normal error assumption,

$$\begin{aligned} \Pr[y_i^* \leq 0] &= \Pr[\varepsilon_i \leq -x_i\beta] = \Pr[\sigma\{\sinh(\lambda + \theta z_i) - m\}/s \leq x_i\beta] \\ &= \Pr[z_i \leq \{\sinh^{-1}(m - sx_i\beta/\sigma) - \lambda\}/\theta] = \Phi(a_i^*), \end{aligned}$$

where $a_i^* = \{\sinh^{-1}(m - sx_i\beta/\sigma) - \lambda\}/\theta$. Hence, we have

$$\log L = \sum_{y_i=0} \log(\Phi(a_i^*)) + \sum_{y_i>0} \log(J_i\phi(z_i)).$$

Whatever the distribution with mean zero and variance σ^2 is used, the marginal effect for $E(y|x)$ is derived as

$$\frac{\partial E(y|x)}{\partial x_k} = \beta_k \Pr[y^* > 0] = \beta_k[1 - F(x_i\beta)],$$

where $F(\cdot)$ is the *cdf* of ε_i. Thus the marginal effect in the censored model with the S_U-normal error is derived as

$$\frac{\partial E(y|x)}{\partial x_k} = \beta_k \Pr[y^* > 0] = \beta_k[1 - \Phi(b_i^*)],$$

where $b_i^* = \{\sinh^{-1}(m + sx_i\beta/\sigma) - \lambda\}/\theta$.

The following simulation experiment is designed as

$$y_i^* = \beta_0 + x_{1i}\beta_1 + \varepsilon_i, \quad i = 1, \ldots, n,$$

where we observe $y_i = y_i^*$ when $y_i^* > 0$ and $y_i = 0$, otherwise. Most of simulation designs are same as in the binary model except for setting $\sigma = 2$.[8)] The simulation results are provided in Table 3.3. For normal error distribution, both Tobit and S_U-normal models yield the coefficient and marginal effect estimation results. When $t(4)$ distribution is assumed, we do not find any difference in the coefficient estimation, but the estimated marginal effects from the Tobit model are much deviated from the true value. Meanwhile, S_U-normal ML model produces the accurate and reliable estimated under the symmetric and leptokurtic distribution. As predicted, when the error distribution is a skewed and thick tail distribution, the Tobit model shows a seriously bad estimation results. On the contrary, S_U-normal ML model tends to yield consistent estimates and marginal effects as the sample size is larger.

8) We obtained basically similar results when we used different set of parameter values.

Table 3.3: Monte Carlo simulation results for binary model

(1) $\varepsilon_i \sim N(0,2)$

	True value	Tobit		S_U-normal ML	
		$n = 500$	$n = 1000$	$n = 500$	$n = 1000$
β_0	−5.000	−5.013(0.337)	−4.990(0.0.237)	4.998(0.348)	−4.982(0.239)
β_1	1.000	1.000(0.046)	0.998(0.033)	0.999(0.048)	0.998(0.033)
$ME(x=3)$	0.158	0.157(0.017)	0.158(0.012)	0.156(0.0.017)	0.157(0.012)
$ME(x=5)$	0.500	0.497(0.024)	0.499(0.016)	0.495(0.025)	0.498(0.017)
$ME(x=7)$	0.841	0.840(0.042)	0.840(0.029)	0.842(0.044)	0.842(0.029)
$\log L$		−616.25	−1233.00	−615.95	1232.71

(2) $\varepsilon_i \sim t(4)$

	True value	Tobit		S_U-normal ML	
		$n = 500$	$n = 1000$	$n = 500$	$n = 1000$
β_0	−5.000	−5.048(0.325)	−5.031(0.265)	−5.005(0.270)	−4.991(0.198)
β_1	1.000	1.005(0.045)	1.003(0.036)	0.999(0.037)	0.998(0.027)
$ME(x=3)$	0.058	0.073(0.025)	0.075(0.022)	0.059(0.012)	0.059(0.008)
$ME(x=5)$	0.500	0.495(0.022)	0.497(0.016)	0.496(0.028)	0.500(0.021)
$ME(x=7)$	0.941	0.927(0.044)	0.924(0.034)	0.941(0.036)	0.940(0.027)
$\log L$		−496.91	−1000.84	−477.63	−959.34

(3) $\varepsilon_i \sim S_U N(0,2;\lambda=1,\theta=1))$

	True value	Tobit		S_U-normal ML	
		$n = 500$	$n = 1000$	$n = 500$	$n = 1000$
β_0	−5.000	−6.069(1.392)	−6.117(1.286)	−4.988(0.309)	−5.014(0.235)
β_1	1.000	1.115(0.165)	1.121(0.148)	0.998(0.041)	1.001(0.030)
$ME(x=3)$	0.093	0.151(0.065)	0.151(0.061)	0.092(0.011)	0.092(0.008)
$ME(x=5)$	0.338	0.470(0.134)	0.470(0.133)	0.337(0.023)	0.0.336(0.017)
$ME(x=7)$	0.977	0.977(0.142)	0.849(0.134)	0.975(0.039)	0.978(0.028)
$\log L$		−644.14	1292.47	−515.40	−1032.82

Note: $ME(x=a)$ indicates the marginal effect at $x=a$. The value in parenthesis is the root mean squared error.

4. Conclusions

In this paper, we attempt further applications of the S_U-normal distribution to various regression models including limited dependent variable models. The S_U-normal distribution is one of the most flexible distributions in capturing skewness and excess kurtosis. It can represent all kurtosis values bigger than that of a lognormal distribution for all coefficients of skewness. The S_U-normal distribution also has simplicity due to the fact that it is derived from the method of transformation to normality. In this paper, we show that the S_U-normal distribution can be an alternative to various distributions in capturing asymmetry and excess kurtosis. The simulation results suggest that the LR test statistic based on the S_U-normal distribution can be served as a model specification test. It is also shown that the advantages of assuming error distributions with a flexible parametric form are considerable for estimating the marginal effects in binary and censored regression models.

Appendix: Derivation of Moments of S_U-normal Distribution

The S_U-normal random variable y is defined as $y = \sinh(x)$ where $x \sim N(\lambda, \theta^2)$. Note that the moment generating function of a normal random variable is

$$E(e^{my}) = \exp\left(m\lambda + \frac{m^2\theta^2}{2}\right).$$

We use the following relationships of hyperbolic functions;

$$\begin{aligned}
\sinh(x) &= \frac{1}{2}(e^x - e^{-x}), \\
\sinh^2(x) &= \frac{1}{2}[\cosh(2x) - 1] \quad \text{and} \\
\cosh &= \frac{1}{2}(e^x + e^{-x}).
\end{aligned}$$

The first moment of y is

$$\begin{aligned}
E(y) &= E(\sinh(x)) = \frac{1}{2}[E(e^x) - E(e^{-x})] = \frac{1}{2}[e^{\lambda+\theta^2/2} - e^{\lambda+\theta^2/2}] \\
&= \frac{1}{2}e^{\theta^2/2}(e^{\lambda} - e^{-\lambda}) = e^{\theta^2/2}\sinh(\lambda) = w^{1/2}\sinh(\lambda),
\end{aligned}$$

where $w = \exp(\theta^2)$. The second moment of y is

$$\begin{aligned}
E(y^2) &= E(\sinh^2(x)) = \frac{1}{2}\left[\frac{1}{2}\left\{E(e^{2x}) + E(e^{-2x})\right\} - 1\right] \\
&= \frac{1}{2}\left[\frac{1}{2}\left(e^{2\lambda+2\theta^2} + e^{-2\lambda+2\theta^2}\right) - 1\right] \\
&= \frac{1}{2}\left[\frac{1}{2}e^{2\theta^2}\left(e^{2\lambda} + e^{-2\lambda}\right) - 1\right] = \frac{1}{2}\left[w^2\cosh(2\lambda) - 1\right].
\end{aligned}$$

Hence the variance of y is given by

$$\begin{aligned}
\text{var}(y) &= E(y^2) - [E(y)]^2 = \frac{1}{2}[w\cosh(2\lambda) - 1] - w\sinh^2(\lambda) \\
&= \frac{1}{2}[w^2\cosh(\lambda) - 1] - \frac{1}{2}w[\cosh(2\lambda) - 1] = \frac{1}{2}(w-1)[w\cosh(2\lambda) + 1].
\end{aligned}$$

References

Andrews, D. W. K. (1997). A conditional Kolmogorov test. *Econometrica*, **65**, 1097–1128.

Bollerslev, T. (1987). A conditionally heteroskedastic time series model for speculative prices and rates of return. *The Review of Economics and Statistics*, **69**, 542–547.

Choi, P. (2002). Essays on modeling asymmetric and leptokurtic distributions of asset returns, Ph.D. Thesis, Texas A&M University.

Choi, P. and Min, I. (2007). Estimating endogenous switching regression model with a flexible parametric distribution function: application to Korean housing demand. Forthcoming in Applied Economics.

Choi, P. and Nam, K. (2007). Asymmetric and leptokurtic distribution for heteroscedastic asset returns: the S_U-normal distribution. *Forthcoming in Journal of Empirical Finance.*

Fan, Y., Li, Q. and Min, I. (2006). A nonparametric bootstrap test of conditional distributions. *Econometric Theory*, **22,** 587–613.

Hansen, B. E. (1994). Autoregressive conditional density estimation. *International Economic Review*, **35,** 705–730.

Johnson, N. L. (1949). Systems of frequency curves generated by methods of translation. *Biometrika*, **36,** 149–176.

Klaauw, B. and Konig, R. H. (2003). Testing the normality assumption in the sample selection model with an application to travel demand. *Journal of Business and Economic Statistics*, **21,** 31–42.

van der Klaauw, B. and Konig, R. H. (2003). Testing the normality assumption in the sample selection model with an application to travel demand. *Journal of Business and Economic Statistics*, **21,** 31–42.

McDonald, J. B. (1996). An application and comparison of some flexible parametric and semi-parametric qualitative response models. *Economics Letters*, **53**, 145–152.

McDonald, J. B. and Yexiao, J. X. (1996). A comparison of semi-parametric and partially adaptive estimators of the censored regression model with possibly skewed and leptokurtic error distributions. *Economics Letters*, **53,** 153–159.

Rockinger, M. and Jondeau, E. (2001). Entropy Densities with an application to autoregressive conditional skewness and kurtosis. *Journal of Econometrics*, **106,** 119–142.

Theodossiou, P. (1998). Financial data and the skewed generalized T distribution. *Management Science*, **44,** 1650–1661.

Theodossiou, P. (2000). Distribution of financial asset prices, the skewed generalized error distribution, and the pricing of options. Working Paper, Rutgers University.

Zheng, J. X. (2000). A consistent test of conditional parametric distributions. *Econometric Theory*, **16,** 667–691.

[Received August 2007, Accepted Septmber 2007]