International Journal of Smart Home
Vol. 9, No. 8 (2015), pp. 73-82
http://dx.doi.org/10.14257/ijsh.2015.9.8.09

An Empirical Study on the Energy Intensity in China Based on the Skew-normal Distribution

Ye Rendao[1], Zhou Longquan[1] and Luo Kun[2]

[1]*College of Economics, Hangzhou Dianzi University, Hangzhou 310018, China*
[2]*Alibaba Business College, Hangzhou Normal University, Hangzhou 310018, China*
yerendao2003@163.com, zlq880711@163.com, olivelk@163.com

Abstract

Firstly, simulation studies on the type I error probability and power of non-central skew-F testing statistic are given by using Monte Carlo method in this paper. Then, the characteristics of the skew-normal distribution are showed on the data of energy intensity in China. Also, skew-normal multivariate regression models are established to study the main influence factors of energy intensity for the whole country, eastern China and mid-western China respectively. On this basis, a GEE model is constructed to verify the robustness of the above skew-normal multivariate regression results. It shows that the technology progress, industry structure and energy consumption structure have significant influences on the energy intensity for the whole country, eastern and mid-western region, with the R&D input and electricity consumption proportion influencing negatively while the secondary industry proportion influencing positively. Moreover, the impacts of technology progress and energy consumption structure turn out to be quite different among regions. Finally, some countermeasures and suggestions are recommended in this paper.

Keywords: *Skew-normal Distribution, Energy Intensity, Monte Carlo Simulation, Skew-normal Multivariate Regression Model, GEE Model*

1. Introduction

In recent years, the energy consumption has grown sharply in China and its main source is coal, which results in increasingly severe environmental problems. The severe nationwide haze recently makes energy consumption and environmental problems become the focus of people. It is not the expansion of total energy consumption but the low level of energy utilization efficiency that aggravate environmental pollution. How to improve the energy efficiency has become an important topic that our economy and sustainable development of environment must deal with.

As an important indicator of measuring energy efficiency, the energy intensity turns into one of research focuses. Existing studies on China's energy intensity usually use factor decomposition and econometric modeling to analyze the variation of energy intensity as well as its influence factors. Using structural decomposition and index decomposition methods to analyze energy intensity, Wang [1] and Li and Wang [2] found that industrial structure and energy consumption structure have significant influence on China's energy intensity. Besides, an econometric model was established by Wu *et al.*, [3], Li and Zhou [4], and Liu *et al.*, [5] to study the influence factors of the energy intensity in various regions of China. Generally, adjusting industrial structure and optimizing energy consumption structure can reduce energy intensity. However, because of rebound effects, there remains controversy over whether the technology development can reduce energy intensity.

ISSN: 1975-4094 IJSH

For studies on energy intensity, it is usually assumed that random error term follows normal distributions for mathematical convenience. But the distribution of actual data is often much more complicated. The actual data of energy intensity usually follows skew distribution, such as skew- normal distribution, skew-t distribution, or skew-elliptical distribution. Further, the skew-normal distribution is the most common one for fitting actual data. Obviously, the normal distribution assumptions may result in a lack of robustness and invalid statistical inference when data show multimodality and skewness. Therefore, it is important to develop statistical models with flexible distribution assumptions to analyze the energy intensity of China. The research of statistical modeling with skew-normal distribution is popular in recent years, which has been discussed in detail by Gupta and Huang [6], Wang *et al.*, [7], and Lachos *et al.*, [8]. They consider that the random error term follows non-normal and asymmetric distribution. Recently, Alejandro *et al.*, [9] and Lin et al. [10] established the skew-normal mixed effects model to estimate unknown parameters. However, it is interested in applying the skew-normal model on real data analysis and obtaining some important results.

In this paper, based on the data of energy intensity of China during 2000-2011, we first prove the skew-normal distribution of the actual data. Then the skew-normal multivariate regression models are established to study the main influence factors of energy intensity. Furthermore, GEE models are used to show the robustness of the results of the skew-normal multivariate regression model.

2. Skew-normal Multivariate Regression Model

Based on the results given by Wang *et al.*, [7] and Ye and Wang [11], the definitions of skew- normal distribution, non-central skew chi-square distribution and non-central skew-F distribution are present as follows. Then, the skew-normal multivariate regression models are established.

2.1. Skew-normal Distribution, Non-central Skew Chi-square Distribution and Non-central Skew-F Distribution

Let $M_{n\times k}$ be the set of all n×k matrices over the real field $\Re$ and $\Re^n = M_{n\times 1}$. For any $B \in M_{n\times k}$, use B', B^-, and $r(B)$ to denote the transpose , the Moore-Penrose inverse, and the rank of B, respectively.

Definition 2.1 The random vector Y follows a multivariate skew-normal distribution, denoted by $Y \sim SN_n(\mu, \Sigma, \alpha)$, if its density function is

$$f_Y(x;\mu,\Sigma,\alpha) = 2\phi_n(x;\mu,\Sigma)\Phi(\alpha'\Sigma^{-1/2}(x-\mu)), \qquad x \in \Re^n, \tag{1}$$

Where $\phi_n(x;\mu,\Sigma)$ the n-dimensional normal density is function with mean vector μ and covariance matrix Σ, and $\Phi(\cdot)$ is the standard normal distribution function.

Definition 2.2 Let $Y \sim SN_n(\mu, I_m, \alpha)$. The distribution of $Y'Y$ is defined as the non-central skew chi-square distribution with degrees of freedom m, the non-centrality parameter $\lambda = \mu'\mu$, and the skewness parameters $\delta_1 = \alpha'\mu$ and $\delta_2 = \alpha'\alpha$, denoted by $Y'Y \sim S\chi^2_m(\lambda, \delta_1, \delta_2)$.

Definition 2.3 Assume that $Z_1 \sim S\chi^2_{n_1}(\lambda, \delta_1, \delta_2)$, $Z_2 \sim \chi^2_{n_2}$, and Z_1 and Z_2 are independent. The distribution of $F = (Z_1/n_1)/(Z_2/n_2)$ is called the non-central skew-F distribution with degrees of freedom n_1 and n_2, the non-central parameter λ, and the skewness parameters δ_1 and δ_2, denoted by $F \sim SF_{n_1,n_2}(\lambda, \delta_1, \delta_2)$.

2.2. Skew-normal Multivariate Regression Model

In this paper, the skew-normal multivariate regression model is given by

$$Y = X\beta + \varepsilon , \tag{2}$$

Where Y is a n×1 random vector, X is a n×p design matrices, $\beta = (\beta_0, \beta_1, \cdots, \beta_{p-1})'$ is a p×1 vector of regression coefficient, and ε is a n×1 vector of random errors with $\varepsilon \sim SN_n(0, \sigma_1^2 I_n, \alpha)$. Based on the test statistic defined by Ye and Wang [11], we consider the following hypothesis testing problem.

$$H_0: H\beta = 0 \text{ vs } H_1: H\beta \neq 0 , \tag{3}$$

Where $H \in M_{m\times p}$ is of rank m.

Let $Z_1 = (H\hat{\beta})'(H(X'X)^- H')^{-1}(H\hat{\beta})$, $Z_2 = Y'(I_n - P_X)Y$, and $\hat{\beta} = (X'X)^- X'Y$. Then $Z_1 \sim S\chi_m^2(\lambda, \delta_1, \delta_2)$, $Z_2 \sim \chi_{n_1}^2$, and Z_1 and Z_2 are independent. For the model Y given in (2), the test statistic is defined by

$$F(Y) = \frac{(H\hat{\beta})'(H(X'X)^- H')^{-1}(H\hat{\beta})/m}{Y'(I_n - P_X)Y/n_1}, \tag{4}$$

Where $n_1 = n - r(X)$

Based on the Definition 2.3, it is easy to obtain $F(Y) \sim SF_{m,n_1}(\lambda, \delta_1, \delta_2)$. Thus, $H_0: H\beta = 0$ is rejected if $F(Y) \geq F_{m,n_1}(\gamma)$, where $F_{m,n_1}(\gamma)$ is the γth quantile of F_{m,n_1}.

2.3. The Simulation Study of Non-central Skew-F Test Statistic

In this section, based on the Monte Carlo methods, we use R software to simulate the type I error probabilities and power function of non-central skew-F test statistic. Consider the following hypothesis

$$H_0: \beta_1 = 0 \text{ vs. } H_1: \beta_1 \neq 0 . \tag{5}$$

For the hypothesis given in (5), simulation studies on the type I error probabilities and power of the non-central skew-F test statistic have been carried out at various values of nominal level γ. In the simulation, it is assigned that sample size n=15, 30 and 50, regression coefficient β_1=0.3, 0.8 and 1.5, scale parameter σ=0.5, 1 and 2, and skew parameter $\boldsymbol{\alpha} = \alpha_1 \mathbf{1}_n$ with α_1=0.5, 1.5 and 3. Table 1 presents the simulated type I error probabilities of the test statistic F(Y) for various combinations of n, α_1 and σ. Table 2 reports the simulated power of the test statistic F(Y) for various combinations of n, α_1, σ and β_1.

Table 1. The Simulation Results Of Type I Error Probabilities of F(Y)

n	α_1	σ	γ			
			0.025	0.05	0.075	0.1
15	0.5	0.5	0.0226	0.0476	0.0692	0.0978
		1	0.0214	0.0468	0.0722	0.0982
		2	0.0244	0.0498	0.0724	0.0986
30	1.5	0.5	0.0224	0.0492	0.0750	0.1034
		1	0.0240	0.0512	0.0758	0.1001
		2	0.0246	0.0498	0.0742	0.1014
50	3	0.5	0.0240	0.0484	0.0716	0.0944
		1	0.0250	0.0486	0.0726	0.0964
		2	0.0204	0.0468	0.0692	0.0970

Table 2. The Simulation Results of Power Functions of F(Y)

n	α_1	σ	β_1	γ			
				0.025	0.05	0.075	0.1
15	0.5	0.5	0.3	0.0398	0.0762	0.1096	0.1428
			0.8	0.1908	0.2986	0.3718	0.4298
			1.5	0.6264	0.7530	0.8164	0.8594
30	1.5	1	0.3	0.0712	0.1234	0.1674	0.2048
			0.8	0.4310	0.5516	0.6244	0.6782
			1.5	0.9496	0.9766	0.9860	0.9910
50	3	2	0.3	0.1012	0.1702	0.2142	0.2566
			0.8	0.6034	0.7142	0.7796	0.8178
			1.5	0.9934	0.9984	0.9986	0.9990

From Table 1 and Table 2, we have the following conclusions.

The type I error probabilities of F(Y) are very close to the various nominal levels across the wide array of scenarios, which have nothing to do with σ for the fixed values of n and α_1. In cases where β_1 departs from the null hypothesis H_0 in (5), the power increases significantly with the sample size enlarging. In short, the overall picture that emerges from the simulation results is that the proposed test is extremely satisfactory, and it is applicable regardless of sample size.

3. The Empirical Analysis

3.1. Variable Selection and Data Description

It is known that the energy intensity, along with its variations, is mainly influenced by technical progress, industrial structure, energy consumption structure and some other factors. Thus, we consider the energy intensity (EI) as the dependent variable, and the R&D input (RI), the industrial structure (IS) and the energy consumption structure (ES) as the independent variables. The energy intensity is the energy consumption per unit of GDP (Unit: tce/10 000 *yuan*); the R&D investment is defined as the research and experiment expenditure of each region (Unit: 100 million *yuan*); the industrial structure is denoted by the second industrial proportion (Unit: %), while the energy consumption structure is denoted by the percentage of electricity consumption in total energy consumption (Unit: %).

The data consists of energy intensity of 30 provinces and autonomous regions in China from 2000 to 2011. The data of the energy consumption and the electricity consumption of each region are gathered from "China Energy Statistical Yearbook"; the data of GDP and the second industrial value of each region are obtained from "China Statistical Yearbook"; the data of R&D investment are collected from "China Scientific Statistical Yearbook". Besides, we calculated the regional real GDP data with the year 2000 as the base period, so as to eliminate the interference of price fluctuations.

3.2. The Verification of Skew-normal Distribution

By the moment method, we get the mean 0.484, the variance 0.182, and the skewness 0.485 of the energy intensity data. The histogram of energy intensity is present in Figure 1. To test the normality of the data, we obtain the values of the statistics of Shapiro-Wilk, Kolmogorov-Smirinov and Cramer-Von, along with their p-values 3.86e-07, 0.0001 and 1.36e-05. All p-values are less than the significance level 0.05, which shows that the distribution of China's energy intensity data is not normal.

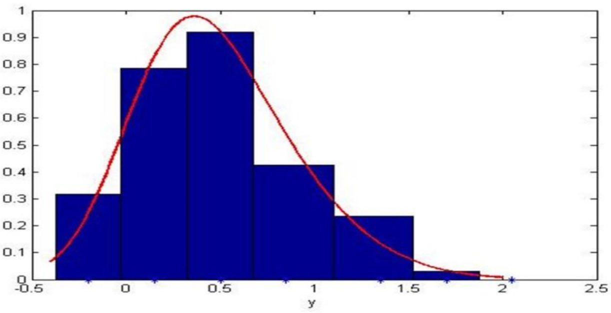

Figure 1. The Histogram of Energy Intensity Data

Based on the Definition 2.1 and moment method, the energy intensity data follow a skew-normal distribution SN (0.04, 0.379, 2.116), and the density fitting curve is showed in Figure 1.

In addition, the test statistic value of the chi-square goodness-of-fit test is 5.31, which is less than $\chi^2_{0.05,2}$ =5.9915. Therefore, the data of energy intensity of 30 provinces and autonomous regions follow a skew-normal distribution under the significance level 0.05.

Considering that there exists territorial difference on energy intensity, especially between eastern region and mid-western region in China, we also analyze the influence factors in the two regions[1]. We can easily show that the data of eastern region and mid-western region follow the skew- normal distribution, because they are the subsets of the skew-normal distribution data.

Based on the above discussion, the skew-normal distribution of the energy intensity data is confirmed. Then, a skew-normal multivariate regression model is established with further statistical analyses[2].

3.3. The Estimation of Skew-normal Multivariate Regression Model

In order to eliminate the side effects in the statistic models caused by outlier, we analyze the outliers of the dataset before parameter estimation. The results of the outlier analysis are displayed in Figure 2 and Figure 3. In Figure 2, it shows that the energy intensity of Ningxia, Guizhou, Shanxi and Qinghai have been far away from the cut-off line, which means these data are outliers. In Figure 3, there are significant differences in energy intensity trend between these four provinces (Ningxia, Guizhou, Shanxi, and Qinghai) and other provinces. As a result, the energy intensity data of Ningxia, Guizhou, Shanxi and Qinghai will be eliminated in the following modeling process.

[1] According to the criteria of region dividing in China's Statistical Yearbook, Beijing, Shanghai, Jiangsu, Zhejiang, Guangdong, Fujian, Tianjin, Shandong, Hebei, Liaoning, and Hainan are divided into the eastern region; while Jilin, Jiangxi, Heilongjiang, Henan, Hunan, Hubei, Chongqing, Sichuan, Yunnan, Shaanxi, Anhui, Guangxi, Shanxi, Xinjiang, Inner Mongolia, Gansu, Qinghai, Guizhou, and Ningxia are divided into mid-western region.

[2] The current empirical study on the intensity of energy consumption mostly assumed that the actual data followed the normal distribution and thus built statistical models. In contrast, this paper firstly verifies the distribution characteristics of the actual data, and then builds a statistical model based on the corresponding energy intensity to a more reliable result.

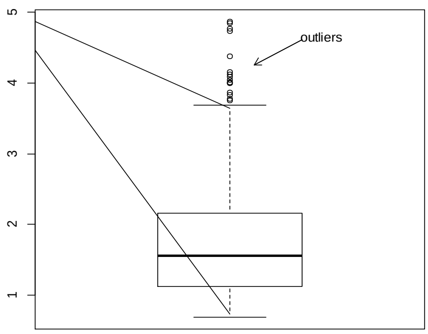

Figure 2. The Boxplot of Energy Intensity Data of China

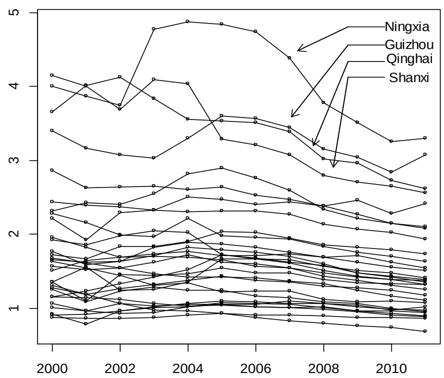

Figure 3. The Trend Chart of Energy Intensity Data of China

Then, we establish the skew-normal multivariate regression models for the whole country, eastern and mid-western regions respectively, and OLS methods are used to estimate the parameters. The skew-normal multivariate regression model is defined as follows.

$$\ln EI_{it} = \beta_0 + \beta_1 \ln RI_{it} + \beta_2 \ln IS_{it} + \beta_3 \ln ES_{it} + \varepsilon_{it} \quad (6)$$

Where $i = 1, 2, \cdots, 26$, $t = 2000, 2001, \cdots, 2011$, $\varepsilon_{it} \sim SN(0, \sigma^2, \alpha)$

The results of skew-normal multivariate regression are shown in Table 3 and Table 4.

Table 3. The Results of Least Squared Estimates of Model (6)

	Intercept(β_0)	**RI(β_1)**	**IS(β_2)**	**ES(β_3)**
Whole country	-0.2557	-0.0871	0.6721	-0.7041
F(Y)	2.4568	71.0390	96.0662	164.2945
p-value	0.1180	0.000*	0.000*	0.000*
Eastern	-1.4210	-0.0201	0.5739	-1.0575
F(Y)	92.2251	4.7089	125.3678	360.4851
p-value	0.000*	0.0326*	0.000*	0.000*
Mid-western	0.8262	-0.1218	0.5815	-0.2615
F(Y)	10.1119	50.4118	13.7584	11.7825
p-value	0.000*	0.000*	0.000*	0.000*

Note: The asterisk (*) indicates that the null hypothesis (H_0) is rejected under the significance level 0.05.

Table 4. The Results of Significance Tests of Model (6)

	Whole country	Eastern	Mid-western
F(Y)	148.0636	257.0811	13.6314
p-value	0.000*	0.000*	0.000*

Note: The asterisk (*) indicates that the null hypothesis (H_0) is rejected under the significance level 0.05.

According to the results of Table 3, the coefficients of the whole country, eastern region and mid-western region are all significant under the significance level 0.05. Table 4 also shows that the regression models of the three regions are significant. The results are given as follows.

Whole country model: $lnEI_1 = -0.2557 - 0.0871 lnRI + 0.6721 lnIS - 0.7041 lnES$, (7)

Eastern model: $lnEI_2 = -1.4210 - 0.0201 lnRI + 0.5739 lnIS - 1.0575 lnES$, (8)

Mid-western model: $lnEI_3 = 0.8262 - 0.1218 lnRI + 0.5815 lnIS - 0.2615 lnES$. (9)

In model (7), it shows that the R&D investment and the proportion of electricity consumption have a negative impact on energy intensity, while second industry proportion has a positive impact. With 1% increase on R&D input, energy intensity decreases by 0.09%; when the proportion of electricity consumption increases by 1%, energy intensity will decrease by 0.7%. For the second industry proportion, when it increases by 1%, energy intensity will increase by 0.67%. Therefore, raising R&D input and electricity consumption proportion, and reducing the second industry proportion, can significantly reduce the energy intensity of China.

From the models (7)-(9), the estimation results of eastern and mid-western models are similar to those of the whole country model, especially for the effects of the secondary industry proportion influence on energy intensity. However, the impacts of technology progress and energy consumption structure turn out to be quite different among regions. Increasing electricity consumption plays an important role in reducing energy intensity in the eastern, while the technological progress is the most important factor in mid-western region. From the models (8)-(9), 1% increase on R&D input leads to a 0.12% reduce of energy intensity in mid-western region. With 1% increase on electricity consumption proportion, energy intensity of eastern region reduces 1.06%. It is concluded that promoting technological progress and improving energy consumption structure are the two main way to reduce energy intensity in China. But the above two ways play a different role in the eastern and mid-western region. Therefore, the choice of reducing energy intensity in different regions should have a more clear orientation.

3.4. Robustness Test: A Comparison between Skew-normal Multivariate Regression Model and GEE

In order to verify the robustness of the results obtained by the above skew-normal multivariate regression models, we also establish GEE models for the whole country, eastern region and mid-western region. The results of GEE model are displayed in Table 5.

Table 5. The Estimation Results of GEE Models

	Intercept(β_0)	RI(β_1)	IS(β_2)	ES(β_3)
Whole country	-0.0647	-0.0408	0.2800	-0.3829
Z-value	-0.4429	-5.4165	4.6211	-7.5823
p-value	0.6578	0.000*	0.000*	0.000*
Eastern	-0.0190	-0.0244	0.5571	-0.4500
Z-value	-0.8803	-2.4174	7.5734	-5.5758
p-value	0.3787	0.0156*	0.000*	0.000*
Mid-western	-1.1297	-0.0571	0.5805	-1.0276
Z-value	1.5782	-3.0936	5.8244	-6.3950
p-value	0.1145	0.002*	0.000*	0.000*

Note: The asterisk (*) indicates that the null hypothesis (H_0) is rejected under the significance level 0.05.

Comparing the results in Table 3 and Table 5, it shows that the results of GEE model are similar to those of the skew-normal multivariate regression model on the direction of the coefficients and values. Thus, the results of the skew-normal multivariate regression model seem to be reasonable; that is, increasing R&D investment and electricity consumption proportion will reduce the energy intensity in China, while increasing the second industry proportion will cause the opposite consequence. Therefore, the results by GEE model indicate that the skew-normal multivariate regression model is robust for the data from nationwide and each region.

4. Conclusions

Our study shows that several factors such as technology progress, industry structure and energy consumption structure have significant influence on the energy intensity. Increasing R&D input and electricity consumption proportion, improving technology, and decreasing the secondary industry proportion are useful to reduce the energy intensity. In addition, the impacts of technology progress and energy consumption structure turn out to be quite different among regions. Therefore, some suggestions are given as follows.

(1) The government is suggested to increase the R&D expenditure, especially the research expenditure on new energy technology. It will not only promote the development of energy technology, but also increase energy efficiency and lower energy intensity in China. More attention should be paid to the introduction of relevant technology research and development to the mid-western region. Increasing support on energy technology in this region will decrease energy intensity dramatically.

(2) Accelerate the optimizing and upgrading of industrial structure, and transform economic growth mode. Since the manufacturing industry is relatively centralized in eastern region, the government is suggested to eliminate high energy consumption and high pollution industry to form a new industrial structure with high-tech industries as the lead. Meanwhile, it is also important to develop the third industry and the high-tech industries to improve the current situation with economic development stimulated by high energy consumption.

(3) Our country can optimize the energy consumption structure, reduce the consumption ratio of coal, oil and other conventional fossil energy, and promote the mode of energy consumption. In addition, by optimizing the structure of energy production, especially accelerating transformation of the electricity production mode, we can get out of the plight that the coal is the main source of current electricity production. Specifically, the government is suggested to pay more attention to improve the energy consumption structure and promote the use of clean energy in eastern region, which will markedly reduce the energy intensity in this region.

Acknowledgements

This material is based upon work funded by National Natural Science Foundation of China (Grant No. 11401148), National Social Science Foundation of China (Grant No. 12CJY012), Ministry of Education of China, Humanities and Social Science Projects (Grant Nos. 14YJC910005, 10YJC790184), Zhejiang Provincial Natural Science Foundation of China (Grant Nos. LY14A010030, Y6110017), Zhejiang Provincial Philosophy and Social Science Planning Project of China (Grant No. 13NDJC089YB), Statistics Research Center of Zhejiang Province (Grant No (2014)78), Scientific Research Innovation Fund of Hangzhou Dianzi University (Grant No KYCX2013JJ026), and Outstanding Post-graduate Dissertation Cultivation Fund of Hangzhou Dianzi University (Grant No. yxlw2014007)

References

[1] Y. Q. Wang, “The methods of analysis factors of energy intensity and its application”, The Journal of Quantitative & Technical Economics, vol. 8, **(2003)**, pp. 151-154.
[2] L. Li, F. Wang, “China’s manufacturing energy intensity decomposition research”, The Journal of Quantitative & Technical Economics, vol. 10, **(2008)**, pp. 66-74.
[3] Q. S. Wu, J. H. Cheng and H. Wang, “Change of energy consumption with the process of industrialization in China”, China Industrial Economics, vol. 4, **(2005)**, pp. 30-37.
[4] L. S. Li and Y. Zhou, “Technology progress can improve energy efficiency, an empirical test based on chines industrial sector”, Management World, vol. 10, **(2006)**, pp. 82-89.
[5] C. Liu, Y. H and Cui, “Comparison research of dynamic relationships on China’s regional energy consumption intensity an empirical analysis based on panel data models of different provinces”, China Industrial Economics, vol. 4, **(2008)**, pp. 34-43.
[6] A. K. Gupta and W. J. Huang, “Quadratic forms in skew normal varieties”, Journal of Mathematical Analysis and Applications, vol. 2, no. 273, **(2002)**.
[7] T. Wang, B. Li and A. K. Gupta, “Distribution of quadratic forms under skew normal settings”, Journal of Multivariate Analysis, vol. 3, no. 100, **(2009)**.
[8] V. H. Lachos, P. Ghosh and R. B. Arellano-Valle, “Likelihood based inference for skew-normal independent linear mixed models”, Statistica Sinica, vol. 1, no. 20, **(2010)**.
[9] R. B. Arellano-Valle, H. Bolfarine and V. H. Lachos, “Skew-normal linear mixed models”, Journal of Data Science, vol. 4, no. 3, **(2005)**.
[10] T. I. Lin and J. C. Lee, “Estimation and prediction in linear mixed models with skew-normal random effects for longitudinal data”, Statistics in medicine, vol. 9, no. 27, **(2008)**.
[11] R. D. Ye and T. Wang, “Inferences in linear mixed models with skew-normal random effects”, Accepted by Acta Mathematica Sinica (English Series), **(2014)**.

Author

Rendao Ye, Associate Professor, College of Economics, Hangzhou Dianzi University, His research interests include linear models, regression analysis, multivariate statistical analysis, mixed–effects models, longitudinal data analysis, generalized inference theory and so on.