Study on the Influencing Factors of Carbon Intensity Using Skew-normal Mixed Model

Ye Ren-dao*, Xu Li-jun1, Luo Kun2, Zhang Yong1

1College of Economics, Hangzhou Dianzi University, Hangzhou, China
2Alibaba Business College, Hangzhou Normal University, Hangzhou, China
*Corresponding author: yerendao2003@163.com

Abstract This paper firstly applies the EM algorithm and gives the maximum likelihood estimates of unknown parameters in skew-normal mixed model. For empirical analysis, we verify the skew-normal distribution characteristics of provincial carbon intensity data in China from 2000 to 2014. A skew-normal mixed model is then constructed to study the main influencing factors of carbon intensity of China. It is found that energy intensity would have the most significant influence on carbon intensity, among a group of factors including GDP per capita, proportion of secondary industry, and dependence on foreign trade. Finally, the results are compared with those based on normal mixed model, so as to confirm the statistical excellent properties of skew-normal mixed model.

Keywords: skew-normal distribution, carbon intensity, EM algorithm, mixed model


1. Introduction

Global warming caused by excessive carbon emissions has become a serious environmental problem, and every country around the world needs to be involved in emission control. In 2015, The Chinese government submitted to the UN Framework Convention on Climate Change the document titled Enhanced Actions on Climate Change: China’s Intended Nationally Determined Contribution. According to the document, China's carbon emissions would peak around 2030, and the Chinese government would endeavor to bring down carbon intensity by 60% to 65% compared to 2005, which is no doubt a real challenge for energy conservation, emissions reduction and economic development. It is urgent to analyze the influencing factors of carbon intensity, with a view of improving emission reduction policies and achieving the development of low carbon economy in China.

Main methods applied in research on influencing factors of carbon intensity include decomposition analysis and econometric analysis. For example, based on structural decomposition analysis (SDA) and index decomposition analysis (IDA), it is found that factors including economic growth, energy efficiency, population, and urbanization level would have great influences on carbon emissions [1,2,3]. Whereas, econometric models are used to study the influencing factors of carbon intensity in different regions of China [4,5,6]. The findings show that adjusting the industrial structure and optimizing energy consumption structure would help reduce carbon intensity.

Note that most of the existing research tend to assume normal distribution when constructing models for the sake of simplicity in mathematical processing. However, the distribution of actual data is often much more complex. In fact, actual longitudinal data show a variety of skew-distribution, including skew-normal distribution, skew-t distribution and skew-elliptical distribution, among which skew-normal distribution is most common in actual data distribution. The loose use of normal distribution hypothesis obviously lacks robustness, and has aroused skepticism [7,8,9]. If skewed data is carelessly assumed to be normal, estimation results will be biased, making statistical inference unreliable. For example, [10] showed that the use of normal distribution failed to reveal the important characteristics of inter-object variables, since Framingham heart data does not follow normal distribution. Therefore, statistical modeling under skewed-normal assumption has become a focus of recent research. Among them, the estimation of unknown parameters for skew-normal mixed model is studied [9,11,12].

In this paper, we firstly apply the EM algorithm and give the maximum likelihood estimates of unknown parameters in skew-normal mixed model. Then, for empirical analysis, we verify the skew-normal distribution characteristics of provincial carbon intensity data in China from 2000 to 2014. A skew-normal mixed model is then constructed to study the main influencing factors of carbon intensity of China. Finally, the estimation results are compared with those based on normal mixed model, so as to confirm the statistical excellent properties of skew-normal mixed model.

2. Skew-normal Distribution

According to [14], if the n-dimensional random variable Y follows skew-normal distribution with the location
parameter $\mu \in \mathbb{R}^n$, scale parameter $\Sigma \in \mathbb{M}_{n \times n}$ and skewness parameter $\alpha \in \mathbb{R}^n$, denoted as $Y \sim SN_n(\mu, \Sigma, \alpha)$, then the density function can be given by

$$f_Y(x; \mu, \Sigma, \alpha) = 2\phi_\Sigma(x; \mu, \Sigma)\Phi(\alpha\Sigma^{-1/2}(x - \mu)), x \in \mathbb{R}^n, (1)$$

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

In this paper, we apply the following properties of skew-normal distribution during parameter estimation of skew-normal mixed model (See [15] for more about the properties of skew-normal distribution).

**Proposition 1** If $W \sim SN_n(\alpha)$, then

$$W = \delta[1][X_0] + (I_n - \delta\delta')^{1/2}X_1, (2)$$

where $\delta = \frac{\alpha}{\sqrt{1 + \alpha^2}}$, $X_0 \sim N(0,1)$, $X_1 \sim N_n(0, I_n)$.

**Lemma 1** If $Y = \mu + \Sigma^{1/2}W$, where $W \sim SN_n(\alpha)$, then

$$Y \sim SN_n(\mu, \Sigma, \alpha)$$

and

$$E(Y) = \mu + \frac{\Sigma^{1/2}\delta}{\pi},$$

$$\text{Cov}(Y) = \Sigma - \frac{2}{\pi} \Sigma^{1/2}\delta\Sigma^{1/2}. $$

**Proposition 2** If $Y \sim N_p(\mu, \Sigma, \alpha)$, $X \sim N_q(\eta, \Omega)$, then

$$\phi_p(y|\mu + Ax, \Sigma, \phi_q(x|\eta, \Omega)$$

$$\phi_q(x|\eta + A\Sigma^{-1}(\eta - \mu - Ap), \Lambda),$$

where $\Lambda = (\Omega^{-1} + A\Sigma^{-1}A)^{-1}$.

### 3. Parameter Estimation

Due to the existence of redundant parameters in skew-normal mixed model, the maximum likelihood estimate gives no explicit solutions. Therefore, we apply the EM algorithm to give the maximum likelihood estimate of unknown parameters. As an iterative algorithm, the EM algorithm can be used to give the maximum likelihood estimate of parameters in probability models with implicit variables (See [16] for more about the EM algorithm).

Then, the skew-normal mixed model is given as in

$$Y_j = X_j(\beta + Z_j\epsilon_{1j} + \epsilon_{0j}), j = 1, 2, \cdots, m$$

(3)

where $Y_j$ is $n \times 1$ random vector denoting the carbon intensity, $\beta$ and $\epsilon_{1j}$ respectively denote fixed and random effects. $\epsilon_{1j} \sim SN(0, \sigma_1^2, \alpha)$ and $\epsilon_{0j} \sim N_n(0, \sigma_0^2 I_n)$ are independent. According to the above-mentioned skew-normal distribution properties, we get

$$\epsilon_{1j} = \sigma_1\delta_t + \sigma_1(1 - \delta^2)^{1/2}X_j,$$

where $\delta = \alpha\sqrt{1 + \alpha^2}$, $t_j$ is the absolute value of the standard normal distribution random variable, $X_j$ is the random variable following standard normal distribution, $t_j$ and $X_j$ are independent. Therefore, (3) can be expressed as

$$Y_j = X_j\beta + Z_j\epsilon_{1j} + \epsilon_{0j} = X_j\beta + Z_j\sigma_1\delta_t + r_j,$$

where $r_j = Z_j\sigma_1(1 - \delta^2)^{1/2}X_j + \epsilon_{0j}$. In the application of EM algorithm, $t_j (j = 1, 2, \cdots, m)$ are implicit variables. Let

$$\Psi_j = \sigma_1^2 I_n + Z_jZ_j^\prime(1 - \delta^2),$$

and

$$Y_j | f_j \sim N_n(X_j\beta + Z_j\sigma_1\delta_t, \Psi_j).$$

(4)

Let $\theta = (\beta', \sigma_1, \sigma_0)'$, then $\theta$ and $\delta$ are the parameters to be estimated in (3). Under Proposition 2 and (4), it follows that the joint distribution of $Y_j$ and $t_j$ can be given by

$$f_{Y_j, t_j}(Y_j, t_j | \theta, \delta)$$

$$= 2\phi_\Psi(Y_j|X_j\beta + Z_j\sigma_1\delta_t, \Psi_j)\phi(t_j | \{t_j > 0\}$$

$$= 2\phi_\Psi(Y_j|X_j\beta, \sigma_0^2 I_n + \sigma_1^2 Z_jZ_j^\prime)\phi(t_j | \{t_j > 0\},$$

where $\eta_j = (Z_j\sigma_1\delta)'\Psi_j^{-1}(Y_j - X_j\beta)$ and $\tau_j^2 = \frac{1}{1 + (Z_j\sigma_1\delta)'\Psi_j^{-1}(Z_j\sigma_1\delta)}$.

And the log-likelihood function of (3) is given by

$$ll(\theta, \delta) = -\frac{1}{2} \sum_{j=1}^m \ln |\Psi_j| - \frac{1}{2} \sum_{j=1}^m (Y_j - X_j\beta)'\Psi_j^{-1}(Y_j - X_j\beta)$$

$$- \frac{1}{2} \sum_{j=1}^m (t_j - \eta_j)^2$$

(5)

where $\Sigma_j = \sigma_0^2 I_n + Z_jZ_j^\prime\sigma_1^2$.

The conditional distribution, first order and second order original moments of $t_j$ with regard to $Y_j$ can be obtained as in

$$f_{t_j|Y_j} (t_j | Y_j) = 2\phi(t_j | \eta_j, \tau_j^2)\mathbb{I}\{t_j > 0\},$$

$$E[t_j|Y_j] = \eta_j + \frac{\phi(\eta_j | \tau_j^2)}{\Phi(\eta_j | \tau_j)^2} \tau_j,$$

$$E[t_j^2|Y_j] = \eta_j^2 + \tau_j^2 + \frac{\phi(\eta_j | \tau_j^2)}{\Phi(\eta_j | \tau_j)^2} \tau_j \eta_j.$$ 

Then, the following steps can be used to estimate the parameters of (3). 

**E-step**: Given $Y_j$, $X_j$ and the parameters of the last iteration ($\hat{\theta}, \hat{\delta}$), using (5) we can get
\[
\hat{t}_j = \hat{n}_j + \frac{\phi(\hat{n}_j / \hat{z}_j)}{\Phi(\hat{n}_j / \hat{z}_j)} \tilde{t}_j,
\]

\[
\hat{t}_j^2 = \hat{n}_j^2 + \tilde{t}_j^2 + \frac{\phi(\hat{n}_j / \hat{z}_j)}{\Phi(\hat{n}_j / \hat{z}_j)} \tilde{t}_j \hat{n}_j.
\]

**M-step:** Based on the estimated parameters of E-step, update \((\hat{\theta}, \hat{\delta})\) by \(\frac{\partial \ln L(\theta, \delta)}{\partial \beta} = 0\), and get

\[
\hat{\beta} = \left( \sum_{j=1}^{m} X_j \left[ \tilde{\xi}_j^{-1} + \tilde{\sigma}_j^{-2} \psi_j^{-1} \left( Z_j \tilde{\sigma}_j \delta \right)^T \psi_j^{-1} X_j \right] \right)^{-1} \times \left[ \sum_{j=1}^{m} X_j \tilde{\xi}_j^{-1} + \tilde{\sigma}_j^{-2} \psi_j^{-1} \left( Z_j \tilde{\sigma}_j \delta \right)^T \psi_j^{-1} Y_j \right],
\]

where \(\tilde{\sigma}_1, \tilde{\sigma}_0\) and \(\tilde{\delta}\) are based on the previous iterations.

Given the original values of parameters to be estimated, which are often chosen to be the corresponding estimates under a normal assumption, we repeat E-step and M-step till convergence, and then obtain maximum likelihood estimate.

4. Empirical Analysis

4.1. Specification of Variables and Data

Our preliminary research show that carbon intensity is mainly influenced by factors including income elasticity of environmental quality demand, scale effects, technical effects and structure effects. Herein, we use the following independent variables.

- GDP per capita, denoted by CG, represents the scale of economic activities.
- Energy intensity, denoted by EI, refers to energy consumption per unit of GDP.
- The proportion of secondary industry, denoted by IS, refers to the proportion of the second industry value added in GDP.
- The dependence on foreign trade, denoted by FT, refers to the total value of commodity imports and exports with regard to GDP.
- As for the dependent variable, carbon intensity denoted by CI refers to carbon emission per unit of GDP. And carbon emission is calculated using a carbon emission conversion method (See [17]) and coefficients as in Table 1.

**Table 1. Coefficients of Energy Conversion and Carbon Emission**

<table>
<thead>
<tr>
<th>Conversion Coefficients of Energy to Standard Coal</th>
<th>Carbon Emission Coefficients</th>
</tr>
</thead>
<tbody>
<tr>
<td>Coal</td>
<td>0.7143</td>
</tr>
<tr>
<td>Coke</td>
<td>0.9714</td>
</tr>
<tr>
<td>Crude Oil</td>
<td>1.4286</td>
</tr>
<tr>
<td>gasoline</td>
<td>1.4714</td>
</tr>
<tr>
<td>kerosene</td>
<td>1.4714</td>
</tr>
<tr>
<td>diesel</td>
<td>1.4571</td>
</tr>
<tr>
<td>Fuel Oil</td>
<td>1.4286</td>
</tr>
<tr>
<td>Natural Gas</td>
<td>1.33</td>
</tr>
<tr>
<td>Electric power</td>
<td>0.1229</td>
</tr>
</tbody>
</table>

We use the longitudinal data of the above variables in 30 provinces of China from 2000 to 2014. Data are derived from China Statistical Yearbook and China Energy Statistics Yearbook.

4.2. Test of Skew-normal Distribution

Descriptive data analysis show that the mean, variance and skewness of carbon intensity is 0.223, 0.293 and 0.435 respectively, all in logarithm forms. Figure 1 gives the frequency distribution histogram. Further, results of Shapiro-Wilk test, Kolmogorov-Smirnov test and Cramer-von Mises test show that the p-values are 9.447e-05, 4.646e-05 and 3.642e-04 respectively, all significant at 5% significance level. These normality testing indicates that the carbon intensity data are not normally distributed.

![Figure 1. Histogram of carbon intensity data](image)

4.3. Parameter Estimation of Skew-normal Mixed Model

The above-mentioned maximum likelihood estimation method is adopted to construct the skew-normal mixed model as in

\[
\ln CI_{ij} = \beta_0 + \beta_1 \ln CG_{ij} + \beta_2 \ln EI_{ij} + \beta_3 \ln IS_{ij} + \beta_4 \ln FT_{ij} + \epsilon_{1j} + \epsilon_{0ij} \tag{6}
\]

where \(i = 2000, 2001, \ldots, 2014, \quad j = 1, 2, \ldots, 30, \) random effects \(\epsilon_{1j} \sim SN(0, \sigma^2_1, \alpha)\), error term \(\epsilon_{0ij} \sim N(0, \sigma^2_0)\).

Table 2 gives the estimation results of two mixed models under different assumptions that random effects follow skew-normal distribution and normal distribution respectively. Compared with the normal mixed model, the AIC and BIC values of the skew-normal mixed model
dropped by 15.4 and 11.3, respectively. It is obvious that the skew-normal mixed model fits the longitudinal data better than the normal one.

Table 2. Estimation Results of Skew-normal and Normal Mixed Models

<table>
<thead>
<tr>
<th></th>
<th>Skew-normal mixed model</th>
<th>Normal mixed model</th>
</tr>
</thead>
<tbody>
<tr>
<td>Constant</td>
<td>-3.2680</td>
<td>-3.1622</td>
</tr>
<tr>
<td>CG</td>
<td>0.3194</td>
<td>0.3234</td>
</tr>
<tr>
<td>EI</td>
<td>0.9643</td>
<td>0.9914</td>
</tr>
<tr>
<td>IS</td>
<td>0.1531</td>
<td>0.1423</td>
</tr>
<tr>
<td>FT</td>
<td>-0.0752</td>
<td>-0.0744</td>
</tr>
<tr>
<td>Skewness</td>
<td>27.6</td>
<td></td>
</tr>
<tr>
<td>Log-likelihood</td>
<td>716.1</td>
<td>707.4</td>
</tr>
<tr>
<td>AIC</td>
<td>-1416.2</td>
<td>-1400.8</td>
</tr>
<tr>
<td>BIC</td>
<td>-1383.4</td>
<td>-1372.1</td>
</tr>
</tbody>
</table>

We then further analyze the estimation results of the skew-normal mixed model, which can be summarized as in

\[ \ln CI = -3.268 + 0.3194 \ln CG + 0.9843 \ln EI + 0.1531 \ln IS - 0.0752 \ln FT \]  
(7)

GDP per capita, energy intensity and the proportion of secondary industry would have positive impacts on carbon intensity, while the dependence on foreign trade would negatively affect carbon intensity. With other conditions given, carbon intensity would rise by 0.32% as GDP per capita increases by 1%. An increase in energy intensity and secondary industry proportion by 1% would lead to an increase in carbon intensity by 0.98% and 0.15%, respectively. On the contrary, as dependence on foreign trade rises by 1%, carbon intensity would drop by nearly 0.08%. It is found that China's carbon intensity could be significantly cut down by reducing energy intensity and the proportion of second industry, as well as increasing the dependence on foreign trade. In addition, due to the positive impact of GDP per capita on carbon intensity, it is crucial that China transforms its current extensive mode of economic growth by which significant increases in carbon emissions are fundamentally determined.

These estimation results could be biased and unreliable, if we tried to fit our skewed data under the assumption of normal distribution. As shown in Table 2, the impacts of GDP per capita and energy intensity on carbon intensity would be underestimated, while the impacts of secondary industry proportion and foreign trade dependence would be overestimated.

5. Conclusions

This paper firstly discusses the parameter estimation method for skew-normal mixed model based on the EM algorithm. Then, for empirical analysis, the skew-normal distribution characteristics of longitudinal carbon intensity data are verified. A skew-normal mixed model is then constructed to study the main influencing factors of carbon intensity of China, the estimation results of which are compared with those based on normal mixed model. It is confirmed that the skew-normal mixed model fits the longitudinal carbon intensity data of China better than the normal model, and therefore have statistical excellent properties.

It is found that factors including GDP per capita, energy intensity, proportion of secondary industry, and dependence on foreign trade would have significant impacts on China’s carbon intensity. In particular, carbon intensity could be reduced by lowering energy intensity and proportion of secondary industry, as well as opening more to the world economy. Still, there exists harsh challenges as China accelerates the process of economic growth mode transformation, and fundamentally control carbon emission with a more intensive economic development. Accordingly, the following suggestions are put forward.

Firstly, our empirical analysis show that energy intensity is the most significant influencing factor of carbon intensity. And therefore, it would be the most direct and effective measure to reduce the carbon intensity by lowering energy intensity. To this end, the efficiency of energy utilization needs to be improved by strengthening R&D of new energy, energy conservation and emission reduction.

Secondly, since carbon intensity drops with the proportion of secondary industry, it makes sense that the secondary industry be reduced in a more carbon-emission-friendly way, and the industrial structure be optimized. The emission-intensive industries needs to be limited or upgraded. In fact, not only the secondary industry but also the economic growth mode of the whole country needs transformation urgently.

Thirdly, it would help to reduce carbon intensity by more opening-up in international trade. However, special attention needs to be paid to avoid the expansion of emission-intensive production and exportation. Therefore, the export of goods highly intensive in carbon emission and energy consumption should be under control, while the export of low-carbon goods could be supported.

Acknowledgements

This research was supported by Zhejiang Provincial Philosophy and Social Science Planning Zhijiang Youth Project of China (16ZJQN017YB), National Natural Science Foundation of China (11401148), National Social Science Foundation of China (12CJY012), National Bureau of Statistics Key Project of China (2015LZ14), Outstanding Post-graduate Dissertation Cultivation Fund of Hangzhou Dianzi University (yxlw2016014), and Outstanding Post-graduate Dissertation Cultivation Foundation of Zhejiang Provincial University Humanities and Social Science Key Research Base (201627).

References


