Research on Prediction of China's Total Energy Production Based on MCMC Method

Chunjiao Gao
Fuzhou University of International Studies and Trade, Fuzhou 350202, Fujian, China

Abstract. Based on the application of Bayesian method in parameter estimation and combined with posterior distribution, this paper proposes a reasonable prior distribution. This paper applies the total energy production of China for empirical analysis in the past 20 years. Firstly, the data are processed into stable data, and then the model is determined. Finally, the parameters of the model are estimated by MCMC method, and the posterior distribution diagram and 95% confidence interval of the parameters of ARIMA model are obtained.

1. Introduction
China is the largest energy consumer in the world. The continuous promotion of industrialization and urbanization leads to an increased demand for energy consumption. Due to the increasingly growth of energy demand, China has become one of the most important energy production and consumption countries in the world. The contradiction between economic and social development and energy consumption growth constraints is increasingly prominent. [1-3] Because of the important role of the national economy, social development and construction of a harmonious and resource-conserving society, all countries highly focus on energy all the world. Energy is not only an important factor that affects a country's economic development level, but also an important production factor that improves a country's output level. Nowadays, as the new era of Chinese economic development, the continuous improvement of the comprehensive national strength allows all aspects of economic strength to occupy a more important position on the international stage. For a country, sufficient energy supply and efficient energy efficiency are some of the most important indicators to measure its sustainable economic development, and a country's overall national strength and people's quality of life. [4] Therefore, we must follow the law of green and low-carbon sustainable development path and reducing the dependence on energy in the future development of the economic and social. [5] It is important to grasp the trend of energy consumption demand in the future for the formulation and implementation of China's energy strategy and for the sustainable development of economy and society. [6]

2. Literature review
Shang Yuping (2019) found that in China, total energy production would continuously rise in the next three years, but the growth rate would gradually slow down. Many scholars at home and abroad have also researched the statistical inference theory and method, applied it to practice, and obtained some results. Shang Yuping (2019) found that in China, total energy production would continuously rise in the next three years, but the growth rate would gradually slow down. Many scholars at home and abroad have also researched the statistical inference theory and method, applied it to practice, and obtained some results. For example, Griffiths and Bewley proposed the logarithmic diffusion model based on
Bayes; TanemruaandKasuya put forward the small-scale Bayes Japanese economic prediction model by posterior information criterion. DuanZhiying (2019) analyzed the dynamic influence relationship through VAR model and impulse response function. The research showed that there was a long-term equilibrium relationship among energy consumption, carbon emission, and economic growth. For the discussion about the Bayesian time series model, the above literature mainly focuses on the estimation of model parameters. For Bayesian estimation of model parameters, there are both conjugate priors selected for the prior distribution of parameters, and robust priors proposed by James, an American statistician, as the representative of HaoHuibing, which are used to obtain Bayesian prediction model.

Compared with other methods, the Bayesian method considers the prior information of parameters with more accurate estimation under the reasonable prior distribution. However, during the application, Bayesian estimation involves the calculation of high-dimensional integral causing to more complex of the parameter estimation. In China, many scholars estimate the parameters of the model by MCMC, as a result, it can effectively solve high-dimensional problems with high accuracy. Based on the advantage of Bayesian method in parameter estimation, this paper proposes a reasonable prior distribution combined with the characteristics of posterior distribution obtained by many literatures using this method. [7] Based on the advantage of Bayesian method in parameter estimation, this paper proposes a reasonable prior distribution combined with the characteristics of posterior distribution obtained by many literatures using this method. The difficulty of calculation makes this paper estimate the parameters of the model by MCMC. Finally, empirical analyzed the total energy production data of China in the past 20 years. The posterior distribution diagram and 95% confidence interval of the parameters of ARIMA model are obtained.

3. Empirical analysis

3.1. Sample data

In this paper, the annual data of China's total energy production from 2001 to 2018 published in "China Statistical Yearbook" are selected for analysis. The data are shown in Table 1 below:

| t (Year) | x(Total Energy production) | Year | X(Total Energy production) |
|----------|---------------------------|------|---------------------------|
| 2018     | 377000                    | 2009 | 286092.2                  |
| 2017     | 358500                    | 2008 | 277419.4                  |
| 2016     | 346037.3                  | 2007 | 264172.6                  |
| 2015     | 361476                    | 2006 | 244762.9                  |
| 2014     | 361866                    | 2005 | 229036.7                  |
| 2013     | 358783.8                  | 2004 | 206107.7                  |
| 2012     | 351040.8                  | 2003 | 178298.8                  |
| 2011     | 340177.5                  | 2002 | 156277                    |
| 2010     | 312124.8                  | 2001 | 147425                    |

Note: year is represented by time variable t, and total energy production is represented by variable x

Data from National Bureau of Statistics

3.2. Stationarity test

Take the logarithm of the data and record the data after the first difference as X (1), take yi =xi –x, where i=1, 2, ...,57, the unit root test to the sequence {Yi}, and the results are shown in Figure 1: when the significance level is 0.05, the sequence is not stable, but after eliminating the linear trend, the sequence is stable.
### Augmented Dickey-Fuller Unit Root Tests

| Type      | Lags | Rho  | Pr < Rho | Tau  | Pr < Tau | F     | Pr > F |
|-----------|------|------|----------|------|----------|-------|-------|
| Zero Mean | 0    | -0.0739 | 0.6494 | -4.62 | 0.001    |       |       |
|           | 1    | -0.0718 | 0.6483 | -1.23 | 0.1745   |       |       |
|           | 2    | -0.0662 | 0.6487 | -1.45 | 0.1294   |       |       |
| Single Mean | 0  | 1.9424 | 0.9944 | 3.23  | 0.9999   | 23.55 | 0.0010 |
|           | 1  | 1.6386 | 0.9913 | 0.95  | 0.9933   | 1.29  | 0.7528 |
|           | 2  | 1.7962 | 0.9925 | 1.64  | 0.9987   | 2.59  | 0.4537 |
| Trend     | 0   | -0.4606 | 0.9863 | -0.30 | 0.8829   | 7.48  | 0.0468 |
|           | 1   | -4.3814 | 0.8287 | -1.59 | 0.7406   | 4.34  | 0.3747 |
|           | 2   | -1.7908 | 0.9619 | -1.28 | 0.8674   | 7.86  | 0.0389 |

**Figure 1.** Results of stationary test

### 3.3. Model decision

In order to estimate the parameters $p$ and $q$, the smaller AIC and BIC are selected for parameter estimation, but it is found that when $p > 3$ or $Q > 3$, both $\varphi_P$ and $\Theta_Q$ cannot pass the parameter test. Now consider the case of $p \leq 3$, $q \leq 3$, it can be seen from Table 3 that the AIC and BIC of ARMA (3, 2) are the minimum, so it is determined that $p = 3$, $q = 2$, and then through the maximum likelihood estimation of ARMA (3,2) model, it can be obtained that $\varphi_1$ and $\Theta_1$ cannot pass the ordinal test. But other ordinal numbers can pass the test. At this time, AIC and BIC are -9.75894 and -4.41671 respectively, indicating that the model has been improved. Its model to be evaluated is: $\left( y_t - \varphi_2 y_{t-2} - \varphi_3 y_{t-3} \right)(1-B)d = \epsilon_t - \Theta_2 \epsilon_{t-2}$.

### 3.4. MCMC parameter estimation

Continue to estimate the parameters of stationary time series $\{Y_i\}$ by MCMC. The Markov chain formed after sampling must be convergent. As known from the following results from Figure 2 to Figure 6, three lines converging to 1 are obtained after 1000 times of sampling, which shows that the Markov chain is convergent.

**Figure 2.** $d$ convergence discriminant  
**Figure 3.** $\varphi_2$ convergence discriminant  
**Figure 4.** $\varphi_3$ convergence discriminant  
**Figure 5.** $\Theta_2$ convergence discriminant
From figure 7 to figure 11, it can be seen that the posterior distribution patterns of $\theta_2$, $\varphi_2$, $\Theta_2$ and $\varphi_2$ are normal except d, and the 95% confidence interval is $[0.3665,0.5005]$, $[-4.356,4.865]$, $[-4.021,5.104]$, $[0.5736,1.769]$, and $[-0.5936,0.721]$.

4. Conclusion
This paper simulates and analyzes the total energy production data by MCMC method in China, and obtains the numerical solution of the posterior distribution of model parameters. The ARIMA (p,d,q) model established in this paper only considers the correlation and long memory of time series. ARIMA (p, d, q) model estimates parameters on the premise of the same variance. The existence of the
heteroscedasticity may cause the parameter estimation and test distortion of the model, and failure of the prediction. Therefore, it can be considered a combination model form the established model and conditional heteroscedasticity model, which makes the model include the heteroscedasticity of time series. This combination model more accurately analyzes financial time series with better prediction effect. The parameter points of the model can be estimated by the mean or mode of the model, but this may be quite different from the estimation obtained by other methods. For example, from table 1, it can be seen that \( d = 0.399 \) is within 95% confidence interval, but it deviates from the mean and mode. The possible reason is that the sample data is relatively small, but it is also noted that the ARIMA model may lead to model failure. There are further studies on whether the prior distribution hypothesis is reasonable and whether the prediction effect of the model obtained by the Bayesian method is significant.

References

[1] Bao Sen, Tian Lixin, Wang Junshuai. Research on energy production and consumption trends and carbon emissions in China [J]. Journal of Natural Resources, 2010, 25 (08): 1248-1254.

[2] Fu Hao, Cheng Yuanlin, Xu Jian. Study on the correlation between energy development and industrial structure in Hunan province [J]. Journal of Shanxi Energy College, 2019, 32 (04): 75-79.

[3] Li Liang. Research on energy demand and structure prediction based on combined Model [D]. Xi’an: master thesis of Xi'an University of technology, 2015.

[4] Di Chang, Xia Zhang, Qiong Liu, Ge Gao, and Yue Wu. "Location based robust audio watermarking algorithm for social TV system." In Pacific-Rim Conference on Multimedia, pp. 726-738. Springer, Berlin, Heidelberg, 2012.

[5] Di Chang, Xia Zhang, and Yue Wu. "A Multi-Source Steganography for Stereo Audio." Journal of Wuhan University (Natural Science Edition), 2013 (3): 277-284.

[6] Xia Zhang, Di Chang, et al. "An Audio Steganography Algorithm Based on Air-Channel Transmitting." Journal of Wuhan University (Natural Science Edition) 57, no. 6 (2011): 499-505.

[7] Xia Zhang, Di Chang, et al. "Tree-like Dimensionality Reduction for Cancer-informatics." In IOP Conference Series: Materials Science and Engineering, vol. 490, no. 4, pp. 042028. IOP Publishing, 2019.