

Research Article

A New Algorithm of Parameter Estimation for the Logistic Equation in Modeling CO₂ Emissions from Fossil Fuel Combustion

Ming Meng,^{1,2} Wei Shang,³ Dongxiao Niu,¹ and Qian Gao¹

¹ School of Economics and Management, North China Electric Power University, No. 619 Yonghua Street, Baoding, Hebei 071003, China

² Soft Science Research Base of Hebei Province, North China Electric Power University, Baoding, Hebei 071003, China

³ School of Economics, Hebei University, Baoding, Hebei 071002, China

Correspondence should be addressed to Ming Meng; ncepumm@126.com

Received 29 March 2014; Accepted 10 July 2014; Published 20 July 2014

Academic Editor: Xian Liu

Copyright © 2014 Ming Meng et al. This is an open access article distributed under the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

CO₂ emissions from fossil fuel combustion have been considered as the most important driving factor of global climate change. A complete understanding of the rules of CO₂ emissions is warranted in modifying the climate change mitigation policy. The current paper advanced a new algorithm of parameter estimation for the logistic equation, which was used to simulate the trend of CO₂ emissions from fossil fuel combustion. The differential equation of the transformed logistic equation was used as the beginning of the parameter estimation. A discretization method was then designed to input the observed samples. After minimizing the residual sum of squares and letting the summation of the residual be equal to 0, the estimated parameters were obtained. Finally, this parameter estimation algorithm was applied to the carbon emissions in China to examine the simulation precision. The error analysis indicators mean absolute percentage error (MAPE), median absolute percentage error (MdAPE), maximal absolute percentage error (MaxAPE), and geometric mean relative absolute error (GMRAE) all showed that the new algorithm was better than the previous ones.

1. Introduction

Global climate change is presently one of the most important issues in scientific and political agenda [1–4]. The Intergovernmental Panel on Climate Change (IPCC) has consistently documented a scientific consensus on the link between anthropogenic greenhouse gas (GHG) emissions and climate change [5–8]. As a direct result of fossil fuel combustion, energy-related CO₂ emissions contribute well to over 80% of the world's total and account for over two-thirds of all anthropogenic GHG emissions [9]. A complete understanding of the rules of CO₂ emissions from fossil fuel combustion is important in modifying the climate change mitigation policy.

Similar to other energy-related indicators, the development of CO₂ emissions from fossil fuel combustion has its

own law. Köne and Büke [10] have first researched the long-term characteristics of energy-related CO₂ emissions. They drew the emission figures of the top 25 emitters and attempted to simulate them with linear models. On this basis, Meng and Niu [11] have analyzed the reasons for the long-term change in energy-related CO₂ emissions and proposed that the S-shaped model is more appropriate than the linear one in simulating the long-term emission curve. They have also offered three similar algorithms to estimate the parameters of the logistic equation, the representative of the S-shaped model. An empirical analysis in China shows that their logistic model is better than the linear one in terms of both the maximum empirical risk and the quality of fit.

In essence, all three parameter estimation algorithms advanced by Meng and Niu [11] have two key logical steps. First, the logistic equation must be transformed into a

linear structure because the unknown parameters of this equation exist as a whole in the denominator, rendering direct parameter estimation impossible. Second, new parameters and variables are used to replace each term of the linear structure, and their values are estimated by the ordinary least square (OLS) algorithm. These processes can help estimate the parameters of the logistic equation but, at the same time, affect the estimate precision. One of the important reasons is that the replacement for the linear structure may change the relative importance of different samples.

For example, as demonstrated by Meng and Niu [11], the logistic equation is given by

$$x_t = \frac{1}{c + ae^{bt}}. \quad (1)$$

It can be transformed into the linear structure as

$$\frac{x_{t+1} - x_t}{x_{t+1}} = (1 - e^b) - c(1 - e^b)x_t. \quad (2)$$

By letting $z_t = (x_{t+1} - x_t)/x_{t+1}$, $\gamma = 1 - e^b$, and $\beta = -c(1 - e^b)$, (2) can be transformed into the linear model $z_t = \gamma + \beta x_t$. Using the OLS method, the linear parameters γ and β are obtained. Consequently, the parameters a , b , and c are easily obtained.

When estimating the parameters, the influence of a sample is positively correlated with its residual; that is, a larger residual indicates greater influence. In fact, according to (1) and (2), a sample with a large residual in (1) does not necessarily have a larger residual in (2) because the latter residual is not only determined on its own but also significantly affected by the next sample. As a result, the relative importance of different samples may be changed in the process of replacement for the linear structure. In other words, the estimated optimal parameters for (2) may not be the optimal parameters for (1). Thus, the present paper proposes a new algorithm (NA) of parameter estimation for the logistic equation, which does not require the process of replacement for the linear structure.

The current paper is organized as follows. Section 2 introduces the complete NA of parameter estimation for the logistic equation. Section 3 presents a case study on CO₂ emissions from fossil fuel combustion in China to test the NA. A detailed comparison between the NA and the three previous algorithms (PAs) is offered. Finally, Section 4 provides the conclusions.

2. Parameter Estimation Algorithms

By letting $y_t = 1/x_t$, the logistic equation (1) is written as

$$y_t = c + ae^{bt}. \quad (3)$$

Equation (3) is an exponential curve and parameters a , b , and c jointly determine the exact shape. Parameter b is at the exponential location; thus, the estimated values of the three parameters cannot be obtained by OLS. The Grey theory [12, 13] has advanced an idea to treat this kind of problem. Following this idea, we developed the algorithm of parameter estimation.

Given that the three parameters of (3) cannot be directly estimated, the form of differential equation is selected as the starting point. Equation (3) is written as

$$\frac{dy_t}{dt} - by_t = -cb. \quad (4)$$

According to the definition of the derivative and considering the statistical results, the first term on the left of (4) can be approximated as

$$\frac{dy}{dt} = \lim_{\Delta t \rightarrow 0} \frac{y_{t+\Delta t} - y_t}{\Delta t} \approx \frac{y_{k+1} - y_k}{k+1-1} = y_{k+1} - y_k. \quad (5)$$

Considering that y_k and y_{k+1} are used in discretizing the first term on the left of (4), the second term can reasonably adopt the mean value of y_k and y_{k+1} :

$$by_t \approx \frac{b}{2}(y_k + y_{k+1}). \quad (6)$$

As a result, (4) can be written as

$$y_{k+1} - y_k - \frac{b}{2}(y_k + y_{k+1}) = -cb. \quad (7)$$

Inputting the samples into (4) yields the following results:

$$\begin{aligned} y_2 - y_1 &= \frac{\hat{b}}{2}(y_1 + y_2) - \hat{c}\hat{b} + \varepsilon_1, \\ y_3 - y_2 &= \frac{\hat{b}}{2}(y_2 + y_3) - \hat{c}\hat{b} + \varepsilon_1, \\ &\vdots \\ y_n - y_{n-1} &= \frac{\hat{b}}{2}(y_{n-1} + y_n) - \hat{c}\hat{b} + \varepsilon_{n-1}, \end{aligned} \quad (8)$$

where n is the number of samples and ε is the residual.

Equation (8) can be written as

$$\mathbf{D} = \mathbf{G}\hat{\mathbf{Z}} + \mathbf{E}, \quad (9)$$

where

$$\mathbf{D} = \begin{bmatrix} y_2 - y_1 \\ y_3 - y_2 \\ \dots \\ y_n - y_{n-1} \end{bmatrix}, \quad \mathbf{G} = \begin{bmatrix} \frac{1}{2}(y_1 + y_2) & -1 \\ \frac{1}{2}(y_2 + y_3) & -1 \\ \dots & \dots \\ \frac{1}{2}(y_{n-1} + y_n) & -1 \end{bmatrix}, \quad (10)$$

$$\hat{\mathbf{Z}} = \begin{bmatrix} \hat{b} \\ \hat{c}\hat{b} \end{bmatrix}, \quad \mathbf{E} = \begin{bmatrix} \varepsilon_1 \\ \varepsilon_2 \\ \dots \\ \varepsilon_{n-1} \end{bmatrix}.$$

To obtain the optimum values of parameters b and c , the residual sum of squares must be minimized.

TABLE 1: Simulation results of each algorithm for each industrial sector.

	1998	1999	2000	2001	2002	2003	2004	2005	2006	2007	MAPE (%)
AC											
REs	0.224	0.224	0.226	0.229	0.239	0.243	0.297	0.304	0.317	0.311	—
A1	0.222	0.231	0.241	0.25	0.26	0.268	0.277	0.285	0.294	0.301	6.27
A2	0.221	0.229	0.238	0.247	0.256	0.265	0.274	0.284	0.293	0.302	5.80
A3*	0.222	0.228	0.234	0.241	0.249	0.258	0.268	0.279	0.292	0.306	4.97
NA	0.222	0.228	0.235	0.242	0.25	0.259	0.269	0.281	0.293	0.308	4.91
IC											
REs	8.892	8.862	9.006	9.182	9.945	11.766	13.916	15.572	17.395	18.807	—
A1	8.478	8.986	9.56	10.214	10.966	11.839	12.865	14.088	15.573	17.411	6.93
A2*	8.626	9.038	9.517	10.08	10.75	11.559	12.555	13.807	15.429	17.607	6.91
A3	8.684	9.056	9.493	10.016	10.649	11.428	12.408	13.674	15.367	17.736	6.93
NA	8.735	9.114	9.563	10.099	10.749	11.553	12.566	13.88	15.646	18.136	6.49
CC											
REs	0.058	0.056	0.059	0.062	0.065	0.069	0.078	0.083	0.087	0.088	—
A1	0.057	0.059	0.062	0.065	0.068	0.071	0.074	0.078	0.081	0.086	4.43
A2	0.057	0.059	0.062	0.064	0.067	0.07	0.073	0.077	0.081	0.086	4.16
A3*	0.057	0.059	0.061	0.064	0.066	0.069	0.073	0.077	0.081	0.087	3.91
NA	0.058	0.059	0.062	0.064	0.067	0.07	0.073	0.077	0.082	0.087	3.87
TSC											
REs	0.437	0.497	0.531	0.542	0.581	0.661	0.78	0.873	0.976	1.086	—
A1	0.435	0.474	0.518	0.568	0.624	0.689	0.763	0.85	0.951	1.072	3.26
A2	0.435	0.474	0.518	0.568	0.624	0.689	0.763	0.849	0.95	1.07	3.3
A3*	0.437	0.474	0.517	0.565	0.621	0.685	0.76	0.849	0.955	1.085	2.95
NA	0.437	0.474	0.517	0.565	0.621	0.685	0.76	0.849	0.955	1.086	2.95
WC											
REs	0.088	0.094	0.09	0.092	0.095	0.106	0.118	0.127	0.135	0.145	—
A1	0.086	0.089	0.093	0.097	0.102	0.108	0.114	0.121	0.13	0.14	4.07
A2	0.086	0.09	0.093	0.097	0.102	0.107	0.114	0.121	0.13	0.141	3.97
A3*	0.087	0.09	0.093	0.097	0.101	0.106	0.112	0.12	0.13	0.142	3.59
NA	0.087	0.09	0.093	0.097	0.101	0.107	0.113	0.121	0.131	0.143	3.47
OC											
REs	0.693	0.689	0.674	0.684	0.679	0.711	0.752	0.784	0.783	0.782	—
A1	0.681	0.69	0.7	0.709	0.719	0.728	0.738	0.747	0.756	0.766	2.99
A2	0.68	0.69	0.699	0.708	0.718	0.727	0.737	0.746	0.756	0.766	2.95
A3*	0.681	0.687	0.693	0.7	0.708	0.717	0.727	0.738	0.75	0.764	2.82
NA	0.684	0.69	0.697	0.704	0.713	0.722	0.732	0.744	0.757	0.772	2.67

* is the best algorithm of A1, A2, and A3.

AC: carbon emitted in agriculture, forestry, animal husbandry, fishery, and water conservancy; IC: carbon emitted in industry (including the electricity generation industry); CC: carbon emitted in construction; TSC: carbon emitted in transport, storage, and post; WC: carbon emitted in wholesale and retail trades; OC: carbon emitted in others.

By letting

$$\mathbf{Q} = \|\mathbf{D} - \mathbf{G}\widehat{\mathbf{Z}}\|^2, \tag{11}$$

the derivative of \mathbf{Q} to $\widehat{\mathbf{Z}}$ must equal zero. Consider

$$\frac{\partial \mathbf{Q}}{\partial \widehat{\mathbf{Z}}} = \mathbf{G}^T \mathbf{G} \widehat{\mathbf{Z}} - \mathbf{G}^T \mathbf{D} = 0. \tag{12}$$

In other words,

$$\widehat{\mathbf{Z}} = (\mathbf{G}^T \mathbf{G})^{-1} \mathbf{G}^T \mathbf{D}. \tag{13}$$

Accordingly, the estimations of b and c are as follows:

$$\widehat{b} = \widehat{\mathbf{Z}}(1), \quad \widehat{c} = \frac{\widehat{\mathbf{Z}}(2)}{\widehat{\mathbf{Z}}(1)}. \tag{14}$$

By introducing \widehat{b} , \widehat{c} , and all samples into (3), the following equation must be valid to ensure the estimated curve across the center of the samples (the summation of the residual equals zero):

$$\sum_{t=1}^n y_t = n\widehat{c} + \widehat{a} \sum_{t=1}^n e^{\widehat{b}t}. \tag{15}$$

The estimation of a is then obtained as follows:

$$\hat{a} = \frac{\sum_{t=1}^n y_t - n\hat{c}}{\sum_{t=1}^n e^{\hat{b}t}} = \frac{(\sum_{t=1}^n y_t - n\hat{c})(1 - e^{\hat{b}})}{e^{\hat{b}}(1 - e^{n\hat{b}})}. \quad (16)$$

Finally, the estimations of a , b , and c are all obtained.

3. Experimental Setup and Results

In [11], the three PAs (A1, A2, and A3) have been adopted to simulate separately the carbon emissions of each main industrial sector in China from 1998 to 2007. The best algorithm for each time series is then selected, and the simulated results of each best algorithm are summed to obtain the simulation results of the total carbon emissions.

To compare the NA offered in this paper with the three PAs [11], the same data were used and the same process was followed to simulate the total carbon emissions.

First, A1, A2, A3, and NA were adopted to simulate the carbon emissions of each main industrial sector. The simulation results of each algorithm and their mean absolute percentage error (MAPE) (17) to the real emissions (REs) are listed in Table 1. Consider

$$\text{MAPE} = \text{mean}_i \left(\left| \frac{\hat{x}(i) - x(i)}{x(i)} \right| \right) \cdot 100, \quad i = 1, 2, \dots, N, \quad (17)$$

where $x(i)$ is the RE of the i th sample, $\hat{x}(i)$ is the simulation result, and N is the number of data used in the MAPE calculation.

Using MAPE as an indicator, the best algorithms of A1, A2, and A3 for each time series of carbon emissions were selected.

As shown in Table 1, the NA always has a better MAPE than A1, A2, and A3. This preliminarily result proved the advantage of the NA.

Second, following the same process in [11], the selected best results of A1, A2, and A3 (see Table 1) were summed and the simulation results of the PAs for the total emissions were obtained. We also summed the simulation results of the NA for each time series of carbon emission. The real total emissions (RTEs), simulation results of the PAs and NA, and relative errors of each algorithm in each year are listed in Table 2.

As RTEs are usually influenced by stochastic factors, forecasting results of NA are not always more precise than PAs. That is, for some forecasting points in Table 2, the relative errors of NA are less than PAs and, for other forecasting points, PAs perform better than NA. As a result, the performance of each method should be evaluated comprehensively by comparison. To compare the two methods, the median absolute percentage error (MdAPE), maximal absolute percentage error (MaxAPE), and geometric mean

TABLE 2: Comparison of the simulation results of the total emissions.

	RTEs	PAs	Relative error (%)	NA	Relative error (%)
1998	10.392	10.11	-2.71	10.223	-1.63
1999	10.422	10.576	1.48	10.655	2.24
2000	10.586	11.116	5.01	11.167	5.49
2001	10.791	11.748	8.87	11.771	9.08
2002	11.604	12.496	7.69	12.501	7.73
2003	13.556	13.395	-1.19	13.396	-1.18
2004	15.941	14.494	-9.08	14.513	-8.96
2005	17.743	15.870	-10.56	15.952	-10.09
2006	19.693	17.637	-10.44	17.864	-9.29
2007	21.219	19.991	-5.79	20.532	-3.24

relative absolute error (GMRAE) were also used as evaluation indicators in addition to the MAPE [14–16]:

$$\text{MdAPE} = \text{median}_i \left(\left| \frac{\hat{x}(i) - x(i)}{x(i)} \right| \right) \cdot 100, \quad i = 1, 2, \dots, N$$

$$\text{MaxAPE} = \max_i \left(\left| \frac{\hat{x}(i) - x(i)}{x(i)} \right| \right) \cdot 100, \quad i = 1, 2, \dots, N$$

$$\text{GMRAE} = \left[\prod_{i=1}^N \left| \frac{\hat{x}(i) - x(i)}{\hat{x}^*(i) - x(i)} \right| \right]^{1/N} \cdot 100, \quad i = 1, 2, \dots, N, \quad (18)$$

where $\hat{x}^*(i)$ is the simulation result obtained from the benchmark method.

The values of the different error analysis indicators were obtained using (17)–(18) and are listed in Table 3.

In essence, the above error analysis indicators have similar functions in distinguishing the better algorithm. However, they still have fine distinctions. The MAPE is an indicator of accuracy reflecting the general closeness of the simulation results to the real data. The MdAPE is the middle value of all absolute percentage errors ordered by size. Similar to the MAPE, the MdAPE has the ability to reflect the general closeness of the simulation results to the real data, but it also has the ability to overcome the influence of a few outliers. The MaxAPE is the worst simulation result; it reflects the maximal simulation risk. The GMRAE calculates the extent of improvement of the new model compared with the previous model. It adopts the geometric mean algorithm and is especially suitable for comparison between different models. Table 3 indicates that the NA is always better than the PAs for each indicator.

4. Conclusions

The intrinsic trend of CO₂ emissions from fossil fuel combustion is an S-shaped curve and is feasibly simulated by the logistic equation. To use the OLS method for parameter estimation, PAs replaced each term of the linearized logistic equation with new parameters and variables. Given that the replacement may change the relative importance of different

TABLE 3: Values of error analysis indicators.

	MAPE	MdAPE	MaxAPE	GMRAE
PAs	6.28	6.74	10.56	0.95
NA	5.89	6.61	10.09	0.88

samples, the precision of the estimated parameters may also be affected. In the current paper, a new parameter estimation algorithm that does not require replacement was advanced. It adopted the differential equation form of the transformed logistic equation as the starting point. After discretization to the differential equation, the observed samples were inputted into the equation. By minimizing the residual sum of squares, the estimated values of parameters b and c were obtained. Inputting the estimated values of b and c into the logistic equation and letting the summation of the residual equal zero, the estimated value of a was also obtained. Finally, the carbon emissions in China were chosen to test the precision of the NA. Error analysis indicators (MAPE, MdAPE, MaxAPE, and GMRAE) all showed that the NA was better than the PAs.

Conflict of Interests

The authors declare that there is no conflict of interests regarding the publication of this paper.

Acknowledgments

This work was supported by “the Fundamental Research Funds for the Central Universities (2014MS148)” and “the National Natural Science Foundation of China (NSFC) (71201057 and 71071052).”

References

- [1] A. Sözen and I. Alp, “Comparison of Turkey's performance of greenhouse gas emissions and local/regional pollutants with EU countries,” *Energy Policy*, vol. 37, no. 12, pp. 5007–5018, 2009.
- [2] J. Kysely, J. Picek, and R. Beranová, “Estimating extremes in climate change simulations using the peaks-over-threshold method with a non-stationary threshold,” *Global and Planetary Change*, vol. 72, no. 1-2, pp. 55–68, 2010.
- [3] S. Ludig, M. Haller, E. Schmid, and N. Bauer, “Fluctuating renewables in a long-term climate change mitigation strategy,” *Energy*, vol. 36, no. 11, pp. 6674–6685, 2011.
- [4] J. P. Smol, “Climate change: a planet in flux,” *Nature*, vol. 483, no. 7387, pp. S12–S15, 2012.
- [5] IPCC(Intergovernmental Panel on Climate Change, *Climate Change 1995: The Science of Climate Change*, Cambridge University Press, New York, NY, USA, 1995.
- [6] IPCC (Intergovernmental Panel on Climate Change), *Climate Change 2001: The Scientific Basis*, Cambridge University Press, New York, NY, USA, 2001.
- [7] IPCC (Intergovernmental Panel on Climate Change), *Climate Change 2007: The Physical Science Basis*, Cambridge University Press, New York, NY, USA, 2007.
- [8] J. Byrne, K. Hughes, W. Rickerson, and L. Kurdgelashvili, “American policy conflict in the greenhouse: divergent trends in federal, regional, state, and local green energy and climate change policy,” *Energy Policy*, vol. 35, no. 9, pp. 4555–4573, 2007.
- [9] V. Taseska, N. Markovska, A. Causevski, T. Bosevski, and J. Pop-Jordanov, “Greenhouse gases (GHG) emissions reduction in a power system predominantly based on lignite,” *Energy*, vol. 36, no. 4, pp. 2266–2270, 2011.
- [10] A. I. Köne and T. Büke, “Forecasting of CO₂ emissions from fuel combustion using trend analysis,” *Renewable and Sustainable Energy Reviews*, vol. 14, no. 9, pp. 2906–2915, 2010.
- [11] M. Meng and D. Niu, “Modeling CO₂ emissions from fossil fuel combustion using the logistic equation,” *Energy*, vol. 36, no. 5, pp. 3355–3359, 2011.
- [12] J. Deng, *Introduction to Grey Mathematical Resource Science*, Huazhong University of Science & Technology Press, Wuhan, China, 2010.
- [13] P.-Y. Chen and H.-M. Yu, “Foundation settlement prediction based on a novel NGM model,” *Mathematical Problems in Engineering*, vol. 2014, Article ID 242809, 8 pages, 2014.
- [14] R. J. Hyndman and A. B. Koehler, “Another look at measures of forecast accuracy,” *International Journal of Forecasting*, vol. 22, no. 4, pp. 679–688, 2006.
- [15] M. Meng, D. Niu, and W. Sun, “Forecasting monthly electric energy consumption using feature extraction,” *Energies*, vol. 4, no. 10, pp. 1495–1507, 2011.
- [16] N. Kourentzes, F. Petropoulos, and J. R. Trapero, “Improving forecasting by estimating time series structural components across multiple frequencies,” *International Journal of Forecasting*, vol. 30, no. 2, pp. 291–302, 2014.



Hindawi

Submit your manuscripts at
<http://www.hindawi.com>

