Mixing data modelling method for response of water environment quality to economic development

To cite this article: Jiefang Wang et al 2019 IOP Conf. Ser.: Earth Environ. Sci. 227 052024

View the article online for updates and enhancements.
Mixing data modelling method for response of water environment quality to economic development

Jiefang Wang 1,3, Shuqi Feng 1, Dongyang Pang 1, Yung Zhou 1 and Muyuan Liu 2

1 Management and Economic department, North China University of Water Resource and Electric Power, Zhengzhou;
2 Business School, Nanjing University, Nanjing, China.
3 Email: wangjiefang_@163.com

Abstract. Environmental Kuznets Curve (EKC) is one of the main tools to analyze the response relationship between water environmental quality and economic development. In classical EKC curve modeling process, the environmental index time series and economic index time series must be the same frequency. Because basin water-quality in China has been gradually monitored and released with high frequency, and most macroeconomic indicators are on annual basis, therefore, by the construction of three classes mix data polynomial equation of EKC curve, the mixing data modeling method of the response relationship between water quality and economic development is proposed to adapt to the situation. The mix data model not only avoids destroying original information due to simple low-frequency measurement of high frequency water quality data using classical model, but also through direct mining of high frequency water quality data, the lag time of water environment quality response to economic development can be obtained. To verify the validity of the model, according to the monitoring and statistical data of 2009-2017, a prefecture level city in Henan province of Haihe River Basin is taken as the research object, and EKC three classes mix data polynomial equation of chemical oxygen demand (COD) concentration and ammonia nitrogen concentration is established.

1. Introduction

In order to simulate the response of environmental quality to economic development, based on the Kuznets curve (EKC curve) (1955) proposed by famous economist Simon Kuznets, Grossman and Krueger (1991) extended Environmental Kuznets Curve[1]. Combing the application results of EKC curve model in water environment field, it is not difficult to find the selection of environmental indicators can be divided into two types: the total quantity and the quality index. The former mainly includes "wastewater discharge" and "the total amount of pollution factors in wastewater"[2-6], and the latter mainly refers to "pollution factor concentration"[6-10]. Among them, quality indicators come from the monitoring results of water bodies, and thus can more directly and objectively reflect the quality of water environment.

Since classical EKC curve model requires environmental indicators and economic indicators be of the same frequency. Thus, in order to keep consistence with the characteristics of annual statistics of macroeconomic indicators, annual averages are often used in pollution factor concentration. However, with the refinement of water quality sampling management, the concentration of water pollutants can already be monitored and released at high frequency and even in real time. It is necessary to expand the classical EKC curve model to achieve direct modeling of high-frequency water quality monitoring.
data and low-frequency economic statistics, and to tap deep-seated response relationship between water environment quality and economic development.

This paper inherits the idea of mixing data processing from Ghysels et al. (2004)[11] and improves it. Based on the connotation features of the response relationship between water environment and economic development, the three classes mixing data polynomial equation of EKC curve is constructed.

2. Classical cubic polynomial equation of EKC curve
At the beginning of EKC curve theory, Grossman and Krueger established a quadratic polynomial equation to verify the relationship between SO2 emissions and economic growth[1]. Later, they modified quadratic polynomial equation to a more practical cubic polynomial equation[12] and analyzed the variation of 14 kinds of air and water pollutants in 66 countries. Since then, the cubic polynomial equation of EKC curve has become the mainstream model for studying the relationship between environmental quality and economic development. Using the cubic polynomial model of EKC curve, the related literature has studied the response relationship of water environment quality to social and economic development of cities[2][3][9][10], provincial areas[7-8], and watershed[4-6]. The literature [13-15] used panel data to analyze the relationship between water environment quality and economic development by horizontal comparison in different regions of China. In above studies, while validating classical “inverted U” EKC curve, it also demonstrated there are other response relationship types between water quality and economic development such as “positive U”, “N”, and “inverted N”. The driving factors and causes were also discussed which has important reference value for the formulating of water environment management and control policies.

3. Construction of three classes mixing data polynomial equation of EKC curve
Definition 1 Let $X = (x_1, x_2, \ldots, x_n)$ be time series of economic indicators, $Y^{(m)} = (y_1^{(m)}, y_2^{(m)}, \ldots, y_u^{(m)})$ be time series of environmental indicators, and $Y^{(m)}$ denotes there are m observations from time t-1 to t, which are $y_{j/m}^{(m)}$ ($j = 0, 1, \ldots, m-1$). Then $B(L^{(m)}/0)Y^{(m)} = \beta_0 x^1 + \beta_1 x^2 + \beta_2 x^3 + \beta_3 + \epsilon$ is the three classes mixing polynomial equation of EKC curve. Among them, $B(L^{(m)}) = \sum_{k=0}^{m-1} \beta(k, 0)L^{(m-k+1)}$ is the action item, $L^{(m)}$ is the lag operator, satisfying the condition of $L^{(m)}y_{j/m}^{(m)} = y_{j/m-k}^{(m)}$ ($j = K+1, \ldots, K+m$).

Obviously, the number of indicators in sequence $X$ and $Y^{(m)}$ is different, and one economic indicator corresponds to $m$ environmental indicators. $x$ is a low frequency sequence and $Y$ is a high frequency sequence, which constitute a mixing sequence. For the mixing data composed of similar low-frequency (annual statistics) GDP per capita representing economic indicators, and high-frequency (acquisition of data at hour, week, month, etc.) pollution factor concentration representing water environment indicators, the three classes mixing polynomial equation of EKC curve can be modeled directly.

In addition to the lag operator $L^{(m)}$, by setting the action item $B(L^{(m)}/0)$, the three classes mixing polynomial equation of EKC curve introduces two important parameters: the lag order $K$ and weight function $\theta(k, 0)$.

(1) The lag order $K$. For example, based on the characteristics of water environment quality response to economic development indicators, it is assumed economic production and pollutant emissions are synchronized, but pollutants migrating and accumulating in water and reflecting water environment quality indicators may have a lag period of $\xi$ ($K = 0, 1, 2, L$). When $K = 0$, it means there is no lag period, and $m$ environmental index observations $y_{j/m}^{(m)}$ from time t-1 to t responding to the economic index $y_i$. When $K > 0$, it indicates there is a lag period, and from time t-1 to t, the later $m-K$ environmental indicator observations $y_{j/m-K+1}^{(m)}, y_{j/m-K+2}^{(m)}, \ldots, y_{j/m}^{(m)}$ and from time t to t+1, the former $K$ environmental indicator observations $y_{j/m-K}^{(m)}, y_{j/m-K-1}^{(m)}, \ldots, y_{j/m-K}^{(m)}$ responding to the economic indicator $y_i$. 

2
(2) Weight function $\omega(k, \theta)$. The main function of weight function $\omega(k, \theta)$ is to gather a large amount of information of high-frequency data, avoid information loss caused by direct frequency reduction for high frequency data, and reduce noise at the same time. In the analysis of the mixing of financial data and macroeconomic data, Ghysels et al. [11][16] proposed Almon-weight functions and Beta-weights etc. Among them, Almon-weight function

$$\omega(k, \theta) = \frac{\exp(0_k + \theta_1 k^1 + \ldots + \theta_{mm} k^{mm})}{\sum \exp(0_k + \theta_1 k^1 + \ldots + \theta_{mm} k^{mm})}$$

has strong adaptability. Normally, when two parameters $\theta_1$ and $\theta_2$ are selected, a rich form of weight functions can be obtained, presenting common forms such as equal weight, descending, increasing, hump, normal, etc. Weights may appear increasing, descending, speeding up or slowing down with the variation of the lag order $K$. Therefore, in this paper, the two-parameter Almon-weight function is used as the general weight form in the three classes mixing polynomial equation.

4. Solving steps of the three classes mixing polynomial equation of EKC curve

Under the criterion of minimizing the average simulated residuals, according to the following steps, the optimization of $K$ and $\omega(k, \theta)$ and the solution of corresponding regression estimation of $\beta_1, \beta_2, \beta_3, \beta_4$ can be achieved.

Step 1: Set initial value $K_i = 0$

Step 2: Let $K = K_i + \Delta$ ($\Delta$ is an integer), and then use the search algorithm to find the optimal solution of $\theta_1, \theta_2, \beta_1, \beta_2, \beta_3, \beta_4$ in Programming equation and the optimal value when the lag order takes different values.

$$\min \left\{ \frac{1}{n} \sum_{t=1}^{n} \left[ B(L^{m\omega}, \beta) y^{m\omega} - \beta_1 x_t^1 + \beta_2 x_t^2 + \beta_3 x_t^3 + \beta_4 x_t^4 \right] \right\}$$

$$\left[ \beta_1, \beta_2, \beta_3, \beta_4 \right] = \text{regress}(B(L^{m\omega}, \beta) y^{m\omega}, X)$$

$$B(L^{m\omega}, \beta) = \sum_{t=1}^{n} \omega(k, \theta) L_{k\omega}$$

$$\omega(k, \theta) = \frac{\exp(0_k + \theta_1 (k - 1) + \theta_2 (k - 2) + \ldots + \theta_m (k - m))}{\sum \exp(0_k + \theta_1 (k - 1) + \theta_2 (k - 2) + \ldots + \theta_m (k - m))}$$

$$L_{k\omega} = \left[ L_{1\omega}, L_{2\omega}, \ldots, L_{j\omega} \right]$$

Among them, the meaning of $(\beta_1, \beta_2, \beta_3, \beta_4) = \text{regress}(B(L^{m\omega}, \theta) y^{m\omega}, X)$ is: when the explained variable is $B(L^{m\omega}, \theta) y^{m\omega}$, and the explanatory variable is $x$, $\beta_1, \beta_2, \beta_3, \beta_4$ is equal to the regression coefficient of the cubic polynomial equation.

Step 3: Find the corresponding lag order $K^*$ when the target value of the planning model (1) takes the minimum $e^*$, and the weight parameter $\theta_1^*, \theta_2^*$, and the regression parameters $\beta_1^*, \beta_2^*, \beta_3^*, \beta_4^*$ corresponding to the optimal lag order $K^*$. The type and the changing trend of the EKC curve can be judged by the size of $\beta_1^*, \beta_2^*$, and $\beta_3^*$.

5. Empirical analysis

5.1. Data acquisition and preprocessing

According to Water Quality Monitoring Report of Henan Province, the weekly monitored of COD and ammonia-nitrogen concentration during the period from January 2009 to July 2017 in the monitoring section of Huojia East Stele Village under the jurisdiction of Henan were obtained (about 52 weeks per year). Due to the strong randomness of water quality fluctuations, the time series of COD and ammonia nitrogen concentration were constructed as the indicators for water environment by taking the four weeks’ mean value of section water quality.
The monitoring section of Huojia East Stele Village is one of the main assessment sections for water environment quality in Jiaozuo. In order to explore the response relationship of water environment quality to economic development, the time series of economic indicators for per capita GDP from 2009 to 2016 are selected.

5.2. Model solving and analysis

The time series of water environment indicators represented by concentration of two pollution factors in Huojia East Stele Village and the time series of economic development represented by per capita GDP of Jiaozuo constitute a typical mixing frequency sequence. Therefore, the three classes mixing polynomial equation of EKC curve was established respectively for COD and ammonia nitrogen concentration. According to step 1 and step 2, the optimal solution of the planning equation (1) under different lagging order can be obtained by using matlab. The parameters $\theta_1$ and $\theta_2$ take the optimal value, when the lag order $K$ is 0 to 7, the optimal weights corresponding to COD and ammonia nitrogen are shown in Figure 1.

![Figure 1. The optimal weights for different lag orders.](image1)

The EKC cubic polynomial equation curve simulation results corresponding to each lag order under optimal weight are shown in Figure 2.

![Figure 2. Fitting results of EKC curves for COD and ammonia nitrogen at different lag orders.](image2)

It can be seen with the change of lag order setting, the trends of simulation curve shows a difference: one cluster of COD EKC curve monotonously decreases, another increases first and then decreases; one cluster of the ammonia-nitrogen environment EKC curve shows a trend of “first decrease, then increase, and then decrease”, and the other shows a “first increase and then decrease” trend.
According to the principle of minimizing the average simulated residual, the optimal lag order corresponding to COD and ammonia nitrogen respectively is $K=4$ and $K=3$ (see Figure 3). The conclusion shows as a product of economic system, pollutants need to undergo a process of migration and accumulation. When analyzing the response relationship between water environment and economic development, the lag period should be considered. The lag time chosen in this empirical study is approximately 12-16 weeks (a lag level represents 4 weeks). The corresponding simulation curves at $K=4$ and $K=3$ can be used as the best simulation curves for the COD and ammonia nitrogen EKC curve (see Figure 3).

![Figure 3](image)

**Figure 3.** Fitting residuals of COD and ammonia nitrogen at different lag orders.

The corresponding parameters are shown in Table 1.

### Table 1 Parameters of EKC curve corresponding to COD and ammonia nitrogen

| Pollution factor | Lag order | $\theta_1$ | $\theta_2$ | $\beta_1$ | $\beta_2$ | $\beta_3$ | Residual (t-value) | $R^2$ | Curve shape | Trend of change |
|------------------|-----------|------------|------------|-----------|-----------|-----------|-------------------|------|-------------|-----------------|
| COD              | K=4       | 4          | -0.7       | -4.1408   | 52.19979  | -222.853  | 1.5237           | 0.9611 | Inverted N  | Decrease         |
| Ammonia nitrogen | K=3       | 4.2        | -0.5       | -0.7267   | 8.896836  | -35.5718  | 0.1189           | 0.9740 | Inverted N  | first increase, then decrease |

---

a * *, ***, and *** shows the significant level are 10%, 5% and 1% respectively.

b The "inverted N" of diagonal tension and the turning point is not obvious.

The results in Table 1 are analyzed as follows:

1. The fitting significance levels of the three classes mixing polynomial curves of EKC curve of COD and ammonia nitrogen reached 0.0028 and 0.0013, all less than 1%; the judgment coefficient reached 0.9611 and 0.9740, both greater than 0.95. The equation simulation works well.

2. The curve shape corresponding to COD and ammonia nitrogen is “Inverted N” (the first turning point is not obvious, especially COD, see Figure 2). The EKC curve of COD showed a declining trend, and the deceleration was slowing down when the per capita GDP was 3.5196. When the per capita GDP was increased to 4.8643, it accelerated again. The EKC curve of ammonia-nitrogen shows a trend of “first decrease, then increase, and then decrease”. When the per capita GDP is 3.5196, it “increases from decreasing”; when the per capita GDP increases to 4.8643, it “decreases from increasing”.

3. The GDP per capita equals 3.5196 and 4.8643 which are two important turning points for EKC curve of COD and ammonia-nitrogen in Jiaozuo of Haihe River Basin under the jurisdiction of Henan. The corresponding year is 2010 and 2013. This period of 4 years was the bottleneck period for pollution control in Jiaozuo (The deceleration of EKC curve of COD is slowing down; the EKC curve of ammonia-nitrogen increases). With the proposition of “Water Quality Assessment for Cross-administrative Regions” in the “Key Watershed Water Pollution Prevention Plan (2011-2015)” promulgated by the Ministry of Environmental Protection in 2012. People's Government of Henan issued “The Comprehensive Treatment of Water Pollution in Haihe River Basin under the jurisdiction of Henan in 2014” and relevant policies. After 2014, as the implementation of related policies, the EKC curve of COD and ammonia nitrogen both showed a rapid decline.
There are differences in the EKC curve between COD and ammonia nitrogen. Part of the reason is the treatment of COD and ammonia nitrogen at different historical stages. Since the "Eleventh Five-Year Plan" period, COD has been the main object of the governance of the water environment of the Haihe River Basin, and the situation has improved year by year. With the intensification of pollution, ammonia nitrogen was used as the main control indicator during the "12th Five-Year Plan" period. In addition, the agricultural non-point source pollution became increasingly prominent and difficult to control, and the ammonia nitrogen control situation was easy to repeat.

6. conclusions
A mixing modeling method for the response relationship of water environment quality to economic development is proposed. (1) This model makes full use of monitoring water quality at high frequency, excavates the lag of water environmental quality indicators to economic indicators and simulates the EKC curve of water environment quality and economic development on the basis of the lag period. (2) Through the presupposition of the weight function with parameters, and optimizing solution through data driven, the information loss or increase virtually caused by the simple mean processing of water environment quality index is avoided. (3) The classical EKC curve theory model can be regarded as the three classes mixing polynomial equation of the EKC curve under the two special conditions: the weight function of high frequency data is equal weight distribution and the lag order is 0. (4) The years of high frequency monitoring of water quality in our country are not long, so the optional sample size is small. As time goes on, the test and application under larger samples need to be carried out.

Acknowledgement
The relevant researches done in this paper are supported by Natural Science Foundation of China (No.71303080); Ministry of Education Research of Social Sciences Youth Funded projects of China (No.14YJC630121); Science and technology innovation talents social science support program in Henan of China (No.2018-cx-026). Science and Technology Major Projects of Colleges and Universities in Henan of China (No.18A630033).

References
[1] GROSSMAN G M, KRUEGER A B 1991 Environmental Impacts of the North American Free Trade Agreement [R] National Bureau of Economic Research working paper 3914
[2] Zheng X, Zhao J, Zhu Y, Wang T J 2013 China Population, Resources and Environment 23(5) 87-91
[3] Zhuang D C, Ye H, Zhang H X J 2013 Economic Geography 33(12) 38-41
[4] Li N, Wei Y N, Wang L CH J 2016 Natural Hazards 81(2) 1193-1207
[5] Feng Y, Li X N, Qu GJ, Xu M 2017 Journal of northwest A&F University(Social Science Edition) 17(5) 67-74
[6] Du X, Xu D, Fu X, Wu G J 2015 Acta Ecologica Sinica 35(06) 1955-1960
[7] PAUDEL K P, ZAPATA H, SUSANTO D J 2005 Environmental and Resource Economics 31(3) 325-334
[8] FARZIN Y H, GROGAN K A J 2013 Environmental Economics and Policy Studies 15(1) 1-37
[9] Sun W G, Xing J, Pan B Y J 2015 Environmental Protection and Circular Economy 35(02) 40-44
[10] Lv P Y, Chen H J 2011 Yangtze River 42(19) 62-65+87
[11] GHYSELS E, SANTA-CLARA P, VALKANOV R 2004 The MIDAS Touch: Mixed Data Sampling Regressions[R] Cirano Working Papers 5(1) 512-517
[12] GROSSMAN G M, KRUEGER A B J 1995 Q J Econ 110(2) 352-377
[13] Guo Q B, Zou J F, Chen J W, Song J 2018 Journal of Economics of Water Resources 36(2) 49-53
[14] Wang Q, Gao J J 2011 Science Research Management 23(7) 157-164
[15] Li P T J 2017 China Population, Resources and Environment 27(5) 22-24
[16] GHYSELS E, SANTA-CLARA P, VALKANOV R J 2006 Journal of Econometrics 131 59-95