Research Journal of Applied Sciences, Engineering and Technology 4(11): 1484-1487, 2012
ISSN: 2040-7467

Submitted: November 28, 2011 Accepted: January 04, 2012 Published: June 01, 2012

Research on Recursive Grouping Data Barycenter Method and its Application

Ji-Lin Zhang
Mathematics and Physics Department, Fujian University of Technology, Fuzhou-350108, China

Abstract: A new and useful parameter estimating method for econometric dynamic model is proposed in this paper. Moreover, a new forecasting method is also proposed in this paper based on it. These methods could deal with the fitting and forecasting of economy dynamic model and could greatly decrease the forecasting errors result from the singularity of the real data. Moreover, the strict hypothetical conditions in least squares method were not necessary in the method presented in this paper, which overcome the shortcomings of least squares method and expanded the application of data barycentre method. The new methods are applied to Chinese steel consumption forecasting based on the historic data. It is shown that the result of fitting and forecasting was satisfactory. From the comparison between new forecasting method and least squares method, we could conclude that the fitting and forecasting results using data barycentre method was more stable than that using least squares regression forecasting method, and the computation of data barycentre forecasting method was simpler than that of least squares method. As a result, the data barycentre method was convenient to use in technical economy.

Key words: Data barycentre, parameter estimation, steel consumption forecasting, triangle recursive

INTRODUCTION

The purpose of this study is to bring to their attention of techno-economist and economy forecasting community a new and useful parameter estimation method and a new forecasting method, which is applied to a case for the understanding and analysis of new methods. As we know, there were many researched results on this fields (Paul, 1999; Labson and Crompton, 1993; Meadows *et al*., 1972), but most of them were based on least squares method. Although the theory of Ls was completed, the fitting and forecasting result based on least squares was unstable, especially to the small sample economy volume. Recently, China has undergone a profound economy change, so the variables affecting economy was changing and random. As a result, it is necessary to set up a new method based on small sample to forecast the changing economy volume. Therefore, a new method that data barycentre economy forecasting method in light of the stability of gravity center of object was presented in this paper. This method could deal with the fitting and forecasting of small sample of economy volume and could greatly decrease the errors of the fitting and forecasting results; Moreover, the strict hypothetical conditions in least squares method were not necessary in the method presented in this paper, which overcome the shortcomings of least squares method and expanded the application of data barycentre method.

CONCEPT AND PROPERTIES OF DATA BARYCENTRE

Definition 1: Let (x_i, y_i) (I = 1,2,$\cdots$,n) denoted one point in the coordinates, Liu and yang (2003) the data barycentre of one point was itself, the data barycentre of two points was the midpoint of them, and the data barycentre of three points was the points which divided into the line between the third point and the data barycentre of other two points in 1:2. Generally speaking ,the data barycentre of n points was the point which divided into the line between the nth point and the data barycentre of other n-1 points in 1: (n-1).

Assumed the coordinates of n points were (x_i, y_i) (i = 1, 2, ..., n). According to the concept of data barycentre, the coordinates of barycentre of n point is:

$$x^{(n)} = \frac{x_n + (n-1)X^{(n-1)}}{n} = \frac{1}{n}\sum_{i=1}^{n} x_i$$

$$y^{(n)} = \frac{y_n + (n-1)y^{(n-1)}}{n} = \frac{1}{n}\sum_{i=1}^{n} y_i$$

where ($x^{(n)}$,$y^{(n)}$) was represented the barycentre of n points. We could achieve the following properties of data barycentre according to above concept:

Corresponding Author: Ji-Lin Zhang, Mathematics and Physics Department, Fujian University of Technology, Fuzhou-350108, China

Theorem 1: The data barycentre of n points is unique.
Theorem 2: The data barycentre of the group of (m+n) points is the point which divides into the line between the barycentre of m points and the barycentre of n points in n:m. According to Theorem 2, the abscissa of the braycentre of x_1, $(x_1+x_2)/2$, ..., $(x_1+x_2+ ... +x_n)/n$ is given by:

$$x = \frac{x_1 + (x_1 + x_2) + ... + (x_1 + x_2 + ... + x_n)}{1 + 2 + ... + n}$$
$$= (\sum_{i=1}^{n} \sum_{t=1}^{i} x_{i'}) / (\sum_{i=1}^{n} i)$$

(Only abscissa was given here. We did the same in the following analysis).
Similarly, the abscissa of the barycentre of:

$$x_1, \frac{x_1 + (x_1 + x_2)}{1+2}, \frac{x_1 + (x_1 + x_2) + (x_1 + x_2 + x_3)}{1 + 2 + ... + 3}$$
$$..., \frac{x_1 + (x_1 + x_2) + ... + (x_1 + x_2 + ... + x_n)}{1 + 2 + ... + n}$$

is given by:

$$x = \frac{x_1 + [x_1 + (x_1 + x_2)] + ... + [x_1 + (x_1 + x_2) + ... + (x_1 + ... + x_n)]}{1 + (1+2) + ... + (1 + 2 + ... + n)}$$
$$= (\sum_{j=1}^{n} \sum_{i=1}^{j} \sum_{i'=1}^{i} x_{i'}) / (\sum_{i=1}^{n} \sum_{i'=1}^{i} x_{i'})$$

Let

$$\sum_{i=1}^{n} x_i = \sum_{i=1}^{n} {}^{(1)}x_i,$$

we called

$$\frac{1}{n} \sum_{i=1}^{n} {}^{(1)}x_i$$

as one-order barycentre operator.
Let

$$\sum_{i=1}^{n} \sum_{i}^{i} x_{i'} = \sum_{i=1}^{n} \sum_{i'}^{i} {}^{(1)}x_{i'} = \sum_{i=1}^{n} {}^{(2)}X_i,$$

we called

$$\frac{1}{1+2+...+n} \sum_{i=1}^{n} {}^{(2)}x_i$$

as two-order barycentre operator.
Let

$$\sum_{j=1}^{n} \sum_{i=1}^{j} \sum_{i'=1}^{i} x_{i'} = \sum_{j=1}^{n} \sum_{i'=1}^{i} {}^{(2)}x_{i'} = \sum_{i=1}^{n} {}^{(3)}x_i,$$

we called

$$\frac{1}{1 + (1+2) + ... + (1 + 2 + ... + n)} \sum_{i=1}^{n} {}^{(3)}x_i$$

As three-order barycentre operator.

Theorem 3: Let

$$\sum_{i=1}^{n} {}^{(k)}x_i = \frac{1}{(k-1)!}$$
$$\sum_{i=1}^{n} (n - t + 1)(n - t + 2)...(n - t + k - 1) x_i$$

Then the k-order barycentre is:

$$\frac{\sum_{i=1}^{n} {}^{(k)}x_i}{\frac{1}{(k-1)!} \sum_{i=1}^{n} (n-t+1)(n-t+2)...(n-t+k-1)}$$

MODELING

In this section, the correlated relationship between Chinese steel consumption and GDP is formulated by following model, which is acknowledged by many scholars of China.

Assumed that the relationship of steel consumption and GDP could be described by the following model (Wang *et al*., 1999):

$$y_t = \alpha_0 + \alpha_1 x_t + \alpha_2 x_t^2 + \varepsilon_t \quad (1)$$

where y_t is represented steel consumption, x_t is represented GDP, ϵ_t is a random error term, t refer to the year. (t = 1985, 1986, ..., 2006).

The real historic data of GDP and Chinese steel consumption from 1985 to 2006 is used to estimate the parameters of model (1). Then the fitting and regressing result of model (1) is obtained as following using data barycenter method based on Matlab procedure. The fitting curve used by this method and LS method in Fig. 1, which show that the fitting results of new method is better than LS method:

$$\hat{y}_t = 4.69686 + 0.0174984 x_t + 0.001000387 x_t^2 \quad (2)$$

TRIANGLE RECURSIVE GROUOING DATA BARACENTRE

This new forecasting method including following process:

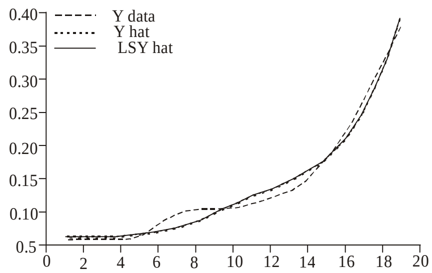

Fig. 1: The fitting curve of steel consumption model

Step 1: Recursive grouping data: Assumed that there are n group real data (x_i, y_i) (I = 1, 2, ..., n), which are recursive grouped many groups, and every group includes three data. These data are recursive grouped like that, from the first data to third data is grouped named first group; then from the second data to forth data is grouped named second group; recursive like that, lastly, from the n-3+1data to n data is grouped named nth group. As a result, the real n group data are recursive grouped n-2 group new data group.

Step 2: Find the data barycenter of the grouping data group and built up econometric dynamic model for the new data of the data barycenter.

Named the data barycenter of n-2 groups as $(x_{c1},y_{c1}),(x_{c2},y_{c2}),...,(x_{c(n-2)},y_{c(n-2)})$, set y_{ck} as new dependent variable and x_{ck} as new independent variable, built up following new econometric dynamic model:

$$y_t = \beta_0 + \beta_1 x_{t1} + \beta_2 x_{t2} + \ldots + \beta_p x_{tp} + \varepsilon \tag{3}$$

Using historical data to fit and regress the above model, obtain following forecasting model:

$$y_t = \hat{\beta}_0 + \hat{\beta}_1 x_{t1} + \hat{\beta}_2 x_{t2} + ... + \hat{\beta}_p x_{tp} \tag{4}$$

Now assume x_{n+1} as known variable, y_{n+1} as unknown variable.
How to forecast y_{n+1}?

Step 3: The abscissa $x_{c(n-1)}$ of the data barycenter, which include the data (x_{n+1},y_{n+1}), using following formula, which is the formula of coordinate of triangle gravity:

$$x_{c(n-1)} = \frac{x_{n-1} + x_n + x_{n+1}}{3} \tag{5}$$

Table I: Chinese steel consumption forecasting result

Year	Forecasting value (0.1 billion)	Year	Forecasting value (0.1 billion)
2007	4.92343	2014	8.00630
2008	5.32753	2015	8.53850
2009	5.91930	2016	8.89970
2010	6.29350	2017	9.04060
2011	6.71630	2018	9.28070
2012	7.12930	2019	9.40350
2013	7.63020	2020	9.73580

Step 4: Substitute $x_{c(n-1)}$ to model (4), then obtain the estimated value of $\hat{y}_{c(n-1)}$.

$$\hat{y}_{c(n-1)} = \hat{\beta}_0 + \hat{\beta}_1 x_{c(n-1)} + \hat{\beta}_2 x^2{}_{c(n-1)}$$

Step 5: Then obtain forecasting value of y_{n+1} using following formula, which is conclude from the coordinate of gravity of triangle:

$$\hat{y}_{n+1} = 3\hat{y}_{c(n-1)} - y_{n-1} - y_n$$

Empirical analysis: Now the author use the Chinese GDP and steel consumption historical data from 1985 to 2006, total 22 years data, to apply to triangle recursive grouping data barycenter forecasting method. In order to use this new forecast method conveniently, the author developed a Matlab procedure for the method.

The forecasting result is as follows in Table I, compare to the forecasting result from 2007 to 2009, which is near to the real dada:

Using the above empirical model and least squares method in model (1), we could obtain:

$$\hat{y}_t = 3.7323 + 0.0583x_t + 6.55\times10^{-6}x_t^2 + \varepsilon_t.$$

The sum of absolute error of (y_t-$\hat{y}_t$), (t = 1985, 1986, …, 2002), using least square method was 22.7325, that was:

$$\sum_{t=1985}^{2002} \left| y_t - \hat{y}_t \right| = 22.7325$$

And the sum of absolute error of (y_t-$\hat{y}_t$) ,using data barycentre method was 16.3249, that was:

$$\sum_{t=1985}^{2002} \left| y_t - \hat{y}_t \right| = 16.3249.$$

So the absolute error of barycentre method was smaller than that of least squares method, which showed that the fitting and forecasting results of using the data barycentre method was more stable than that of using Ls method Wen (1992).

CONCLUSION

In this study, the data barycentre method was presented. The steel consumption volume fitting and forecasting model was set up and the steel consumption volume from 2010 to 2020 was forecasted. It is shown that the result of fitting and forecasting was satisfactory. From the comparison between the two methods, the fitting and forecasting results of using the data barycentre method was more stable and the computation was simpler than the Ls method. Especially, the data barycentre method based on small sample was more suitable to use to forecast the economy variables in the profound changing economy society.

REFERENCES

Labson, B.S. and P.L. Crompton, 1993. Common trends in economic activity and metals demand: Cointegration and the intensity of use debate. J. Environ. Econ. Manage., 25(2): 147-161.

Liu, F., and Q.L. Yang, 2003. A method of regression model. J. Yang Zhou Univ., 6(4): 121-125.

Meadows, D.H., D.L. Meadows, J. Randers and W.W. Behrens, 1972. The Limits to Growth. Universe Books, Nork.

Paul, C., 1999. Forecasting steel consumption in South-East Asia. Res. Policy, 25: 111-123.

Wang, S.G., M. Cheng and L.P. Cheng, 1999. Linear Regression and Variance Analysis. Advanced Education Publication.

Wen, F.R., 1992. A Conservative method of regression lines. Math. Pract. Theory, 22(2): 115-121.