ISSN : 0974 - 7435 Volume 10 Issue 7

2014

BioTechnology

An Indian Journal

FULL PAPER

BTAIJ, 10(7), 2014 [1695-1701]

Moment of inertia based simple linear regression

Xu Li-Li
School of Economics and Management, Henan Polytechnic University, Jiaozuo
Henan 454000, China

ABSTRACT

Variables on economic activities are of random and correlation generally, while ordinary least square regression (OLSR) can not reflect such data. The reason is OLSR defines only the dependent variable as the one-dimensional random variable, without considering the random variable, and the regression results are affected by the chosen coordinate. To improve it and reflect real economic variables, a new independent-coordinated linear regression method -- the rotating inertia method is presented, which is based the dynamics nature of rigid plate fixed-axis rotation. The simulation example verified and compared with the least squares method to conventional method regression system deviation is small, the regression accuracy. Moment of inertia method is a numerical solution, the computation of the advantages make it has a broad application prospect.

KEYWORDS

Simple linear regression (SLR); Ordinary least squares regression (OLSR); Moment of inertia, independence of coordinates.

INTRODUCTION

Linear regression analysis[1,2] is one of the most basic research method in mathematical statistics, and usually used to study the variables correlation. In the field of social economy, the relationship between many variables is not linear even on the macro level, while can be approximated as linear processing on the micro level. Moreover, sometimes if the variables are pretreated logarithmically, the nonlinear relationship between variables can be transformed to the linear relationship. At present, the main analysis of the statistical numerical calculation software based on matrix computation. Therefore, it's useful to make linear regression variables with high precision.

Based on number of variables dependent variables linear regression can be divided into simple linear regression (one-element linear regression) and multiple regression analysis, in which, simple linear regression is one of the most simple and the most basic questions, summarized as follows.

Assume x, y as variables existing a linear relationship

$$y = \beta_0 + \beta_1 x + \varepsilon \tag{1}$$

In which, $\beta_i (i = 0,1)$ is constant, ε is a random error. N observations on variables are made, observations are following.

$$X = (x_1, x_2, ..., x_n)'$$
$$Y = (y_1, y_2, ..., y_n)'$$

The above data and scattered point sets $S = \{(x_i, y_i) | i \in [1, n]\}$ are equivalent. The regression line between x and y based on above observation data is[3]

$$\hat{y} = \hat{\beta}_1 x + \hat{\beta}_0 \tag{2}$$

Matrix form of equation (1) is

$$Y = (1, X)\mathrm{B} + \mathrm{E} \tag{3}$$

In which

$$\mathrm{B} = (\beta_0, \beta_1)' , \quad \mathrm{E} = (\varepsilon_1, ..., \varepsilon_n)'$$

Most commonly used solution of a linear regression is the linear regressions based on ordinary least squares regression (OLSR), y is regarded as the dependent variable, the only random variable, while x is as the independent variable and is not as random variables. Maximum likelihood estimation for the parameter matrix B is[4]

$$\hat{\mathrm{B}} = \left((1, X)'(1, X)\right)^{-1} (1, X)'Y \tag{4}$$

In which $\hat{\mathrm{B}} = (\hat{\beta}_0, \hat{\beta}_1)'$

Results of least squares regression (OLSR) are not of coordinate independence. The coordinate independence refers to the operation location coordinates orthogonal transformation (translation and / or rotation) does not affect the result of the operation. As shown in Figure 1, the lines L and L ' are respectively for the same set of data the regression results in the least squares coordinates xOy and x'O'y'. L and L 'are apparently not coincidence.

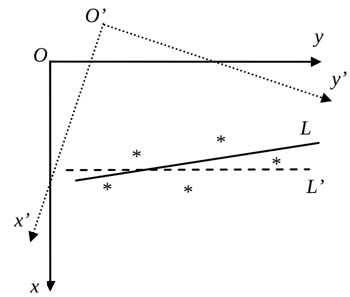

Figure 1 : OLSRs in different coordinate systems

The author thinks, social economic variables value is seldom "pure" argument without random. When the angle of observation, observation equipment, data definition and summarization method is different, the same economic phenomenon may have very different observational data form. But by a linear transformation and even simple coordinate transformation, data often exhibit obvious equivalence. Therefore, to get the results of regression data set having the same equivalence relation, it is necessary to develop the linear with coordinate invariance regression method.

A LINEAR REGRESSION METHOD BASED ON MOMENT OF INERTIA

The following two-element linear regression method is inspired by a classical dynamics nature that as rotating shaft is axis of symmetry, the moment of inertia of rigid plate with uniform thickness symmetric shape is to get the minimum value.

Definition1: The paired observation value of variable x, y is expressed as a set of points $P = \{(x_i, y_i) | i \in [1, n]\}$. Regard a given point x_i, y_i as the particle whose mass is 1, then with respect to the line $y = kx + b$ in any plane the moment of inertia of the point is

$$J = \sum_{i=1}^{n} \frac{(y_i - kx_i - b)^2}{1 + k^2} \qquad (5)$$

Get the minima of J (k,b)

$$\begin{cases} k = \dfrac{-G + \sqrt{G^2 + 4F^2}}{2F} \\ b = \overline{Y} - k\overline{X} \end{cases} \qquad (6)$$

In which

$$\overline{X} = \frac{1}{n} \sum_{i=1}^{n} x_i \qquad (7)$$

$$\overline{Y} = \frac{1}{n} \sum_{i=1}^{n} y_i \qquad (8)$$

$$F = \sum_{i=1}^{n} (x_i - \overline{X})(y_i - \overline{Y}) \qquad (9)$$

$$G = \sum_{i=1}^{n} [(x_i - \overline{X})^2 - (y_i - \overline{Y})^2] \qquad (10)$$

In the formula (2), assume

$$\begin{cases} \hat{\beta}_1 = k \\ \hat{\beta}_0 = \overline{Y} - \hat{\beta}_1 \overline{X} \end{cases} \quad (11)$$

The above one-element linear regression method is called "one-element linear regression method based on moment of inertia", referred to as the Moment of Inertia Based Regression, MIBR. According to the definition of MIBR, results with coordinate independence can be obtained.

THE SIMULATION EXPERIMENT

Comparing MIBR to least square method can prove its advantages. To simplify the calculation, the mathematical model[5] is constructed, as shown in Figure 2.

a. Assume $D_1,...,D_5$ independent each other, and $D_i : N(0,0.4), d_i$ is the observed values of random variables $D_i, i = 1,...,5$.

b. Assume x, y as the variables, the observation vector respectively $X = (x_1, x_2, x_3, x_4, x_5)'$, $Y = (y_1, y_2, y_3, y_4, y_5)'$

In which

$$\begin{cases} x_i = 5 + i - \dfrac{\sqrt{2}}{2} d_i \\ y_i = 6 + i + \dfrac{\sqrt{2}}{2} d_i \end{cases} \quad (12)$$

Obviously, the theoretical relationship of variables x, y is considered

$$y = 1 + x + \varepsilon \quad (13)$$

The above data were calculated respectively by MIBR and the least square regression method, and the corresponding regression errors were got by comparing calculated results with the formula (13). When the regression line angle is large, small changes in the position of the line will cause dramatic changes of the slope, so defining β_1 as a measurement variable of error is not appropriate, this paper adopts $\Delta\alpha$ (error of regression line angle) as regression error metrics. Additionally, since β_0 (the intercept of the regression line) and β_1 (or α) is not independent, it's not of practical significance to analysis their errors separately. Therefore, only one index is defined to measure the performance of each regression method. Making (X, Y) 30 independent observations and linear regression with the two methods, the regression line is obtained, whose angle is as shown in TABLE 1, $\overline{\Delta\alpha}$ (the average error of three kinds of regression methods) and $\overline{|\Delta\alpha|}$ (the mean absolute error) are as shown in TABLE 2, among them

$$\overline{\Delta\alpha} = \frac{1}{30}\sum_{i=1}^{30}(\alpha_i - 45^\circ) \quad (14)$$

$$\overline{|\Delta\alpha|} = \frac{1}{30}\sum_{i=1}^{30}|\alpha_i - 45^\circ| \quad (15)$$

TABLE 1 : Obliquity data of unary linear regressions (°)

OLSR	MIBR	OLSR	MIBR	OLSR	MIBR
36.4614	36.7430	51.1583	52.5040	34.3672	36.5952
44.9712	46.0789	42.2432	43.4865	44.1101	45.4543
39.6035	47.3888	57.4840	57.6469	52.4824	57.6470
29.3945	30.1362	36.8264	38.4114	48.5504	50.4036
53.9081	54.6807	49.1087	55.8194	33.3938	35.8162
41.9061	44.3928	46.9750	49.2606	42.2317	44.1155
47.1428	50.2315	43.0093	44.6703	26.8489	27.5804
31.7569	33.8584	35.8210	36.1897	44.7187	47.1474
42.4422	43.0127	46.0340	52.6152	40.3706	41.4210
51.9599	52.1456	35.0894	36.0983	48.0692	51.4395

TABLE 2 : Angle error data of unary linear regressions (°)

	OLSR	MIBR
$\overline{\Delta\alpha}$	-2.3854	-0.2336
$\overline{\lvert\Delta\alpha\rvert}$	6.2436	6.5979

The absolute value of $\overline{\Delta\alpha}$ (mean angle error) by MIBR is smaller than by OLSR, as shown in TABLE 2, which means MIBR has unbiased advantages.

APPLICATION EXAMPLES

Data on the annual reports of the listed coal enterprises in China are collected, in which those non-fixed assets value being between RMB25 000 000 000 yuan and RMB150 000 000 000 yuan are shown in TABLE 3.

TABLE 3 : Data in some China listed coal corporates'annual reports

Corporate name	Year	Main business income *p*(RMB)	Non-fixed capital *k*(RMB)
Shenhuo shares	2012	27985000000	26417400000
Lu'an Environmental Energy	2011	22426300000	27582110000
Lu'an Environmental Energy	2012	13980400000	32268490000
Jizhong energy	2012	30072400000	26613100000
Yanzhou coal	2009	21500352215	45172821500
Yanzhou coal	2010	34844400000	54495300000
Yanzhou coal	2011	48768300000	76592900000
Yanzhou coal	2012	59673500000	96623500000
China Coal Energy	2006	28346700000	36712800000
China Coal Energy	2007	36823300000	41069800000
China Coal Energy	2008	52282566000	72945635000
China Coal Energy	2009	53729503000	83103856000
China Coal Energy	2010	71268400000	92494400000
China Coal Energy	2011	88872400000	129312600000
China Coal Energy	2012	87291700000	143319700000
China Shenhua Energy	2007	82107000000	121975000000
China Shenhua Energy	2008	107133000000	146466000000
China Shenhua Energy	2009	121312000000	164152000000

Scatter diagram of data shown in TABLE 3 is expressed in Figure 2. All data were divided into two groups, of which, scatters are marked "*" and made regression analysis if the non-fixed asset value is less than the RMB100 000 000 000 yuan, the others are marked "O" and used for verification of the results of the regression.

Data of group "*" are made least squares regression, the result is

$$\hat{p}_1 = 58.3537 + 0.5980k \tag{16}$$

The unit of variable in above formula is RMB100 million yuan. Regression line of OLSR is dotted line as shown in Figure 2.

While data of group "*" are made regression by MIBR, the result is

$$\hat{p} = 41.4890 + 0.6288k \tag{17}$$

The unit of variable in above formula is RMB100 million yuan. Regression line of MIBR is the solid line as shown in Figure 2.

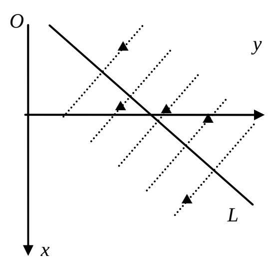

Figure 2 : Scatter of random data

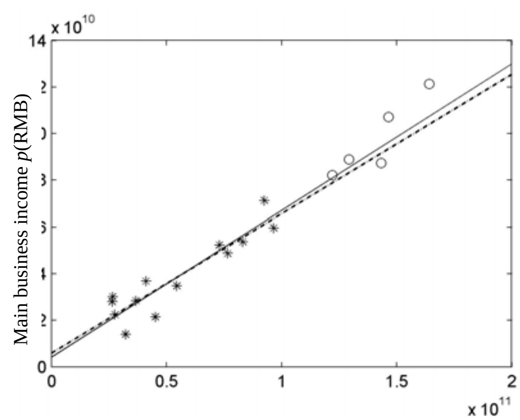

Figure 3 : Simple linear regressions

To compare regression quality of two kinds of regression method, data of non-fixed assets value being more than RMB100 000 000 000 yuan in TABLE 3 are used to validate regression effect. Based on the maximum likelihood hypothesis, considering main business income of group "O" as theoretical value, the error of the predictive value compared with the theoretical value is calculated on the basis of the two kinds of regression. The regression error of OLSR is recorded as ε_1, while the regression error of MIBR is taken as ε_2, the mean regression error is denoted as $\overline{\varepsilon_2}$, and the mean absolute value of regression errors is written as $\left|\varepsilon_2\right|$. The calculation of regression error results as shown in TABLE 4.

The mean regression error of OLSR

$\overline{\varepsilon_1} = -71.609$

The absolute mean regression error of OLSR

$\left|\overline{\varepsilon_1}\right| = 88.612$

The mean regression error of MIBR

$\overline{\varepsilon_2} = -45.048$

The absolute mean regression error of MIBR

$\left|\overline{\varepsilon_2}\right| = 72.956$

The mean regression error $\overline{\varepsilon}$ can be considered system deviation, according to $\overline{\varepsilon_2} < \overline{\varepsilon_1}$, the system deviation by MIBR is narrower than by OLSR.

TABLE 4 : Errors of regressions

***p*(亿元)**	**888.724**	**872.917**	**821.070**	**1071.330**	**1213.12**
$\hat{p}_1$	831.662	915.426	787.782	934.241	1040.006
ε_1	-57.062	42.509	-33.288	-137.089	-173.114
$\hat{p}_2$	854.610	942.687	808.471	962.471	1073.681
ε_2	-34.114	69.770	-125.990	-108.859	-139.439

The absolute mean regression error can be regarded as standard deviation of the regression error, according to $\overline{\varepsilon_2} < \overline{\varepsilon_1}$, standard deviation of the regression error by MIBR is smaller than by OLSR. In another word, MIBR is more stable than OLSR.

CONCLUSION

Both the simulation experiment and regression examples have proved MIBR is more precise and stable than OLSR in one-element linear regression. Solution of MIBR is analytical and demands less computational complexity, so MIBR could be used broadly.

This paper only discusses the application of MIBR in one-element linear regression, as for its application in a multiple regression will be discussed in another paper.

ACKNOWLEDGEMENTS

This paper is funded by national natural sciences fund youth fund of the People's public of China (51274087), Projects of national natural science foundation of the People's public of China (51104055), Doctor fund of Henan Polytechnic University (B2012-086), the views expressed are authors' alone.

REFERENCES

[1] Xu Wei-wei; Regression analysis of correlation of debt maturity structure of coal enterprises[J], Friends of accounting, **31,**116-118 **(2011)**.

[2] Ning Yun-cai; Risk Recognition and Management on Margin Trading in China[C]. ICMSIS09 PROCEEDINGS, **09**,157-159 **(2009)**.

[3] Sheng Ju, Xie Shi-qian, Pan Cheng-yi; Probability theory and mathematical statistics (Third Edition) [M].Beijing: Higher Education Press, **12,** 297 **(2001)**.

[4] Yu Xiu-lin, Ren Xue-song; Multivariate statistics analysis. Beijing: China Statistics Press, 05,156-161,239 **(1999)**.

[5] Xu Wei-wei; Linear regression method based on sparse data minimum axial symmetric envelope domain[J], Journal of Hangzhou Dianzi University, **34(2)**, 36-40 **(2014)**.

[6] Xu Li-li, Liu Shao-wei; Establishing Prime System of Financial Management in Rural Enterprise[C]. Proceedings of 2007 International Conference on Agriculture Engineering, **11**, 58-62 **(2007)**.

[7] Xu Wei-wei; Risk Conversion of Debt Financing in the Coal Company[C]. Zhengzhou:Artificial Intelligence, Management Science and Electronic Commerce (AIMSEC), 5142-5145 **(2011)**.