TELKOMNIKA Indonesian Journal of Electrical Engineering
Vol.12, No.3, March 2014, pp. 1933 ~ 1938
DOI: http://dx.doi.org/10.11591/telkomnika.v12i3.4538

On Data Mining Technology to the Quantitative Efficiency Assessment using SBM Model: An Empirical Study on Education Efficiency in Jiangxi Province

Xiaodong Zhu
School of Management, Wuhan University of Technology
Luoshi Road No 117, Wuhan 430070, China, Telp/fax: +86-027-86572231
e-mail: zhuxiaodong_whut@163.com

Abstract

Data mining technology has been widely used in various applications. Important information could be discoveried by data mining technology. In the education efficiency analysis, it is essential to use data mining technology to analyze histrical data. However, little work has been done to the quantitative analysis on education efficiency. To address this issue, this paper applies the slacks based measure (SBM Model) to objectively assess the input-output eficiency of education efficiency in Jiangxi province, China. In the design of the input-output data of the SBM model, this paper employed the Fuzzy principal component analysis (Fuzzy PCA) to analyze historical data sets in an itteligent manner. Three input indexes and three output indexes have been chosen by the Fuzzy PCA. The empirical analysis finds that the change of relative education efficiency in Jiangxi from 2000 to 2012 falls into 5 stages – being relatively efficient; relative efficiency decrease; relative efficiency increase; relative efficiency decrease and being relatively efficient. The analysis results show that the changes of national policies and economy have great influence on the education efficiency in Jiangxi. For the first time, this paper puts forward useful suggestions to maitain education efficiency in Jiangxi, which could provide valuable reference for similar studies over the country.

Keywords: *data mining, quantitative analysis, efficiency assessment, SBM model*

1. Introduction

With highly development of the data warehouse and internet application, the data produced by the information technological system explode dramatically, from MB and GB in 1980s to TB or even PB in the current [1]. The explosion of data and the lagging analysis methods conflict more fiercely. It is really a hard work to find valuable information and knowledge in a pile of complex data. Fortunately, data mining technology provide powerful support to discovery hidden information from large amount of data [2, 3].

Education is an important power to support social and industrial development. It cultivates high-quality personnel for the society through knowledge inheritage and innovation. From the angle of efficiency, education itself is an effective way to promote economic efficiency and that is why it is attached importance to. Nevertheless, as part of social life, educaiton has problem of efficiency as well [4]. In 1966, James and Coleman first put forward educational efficiency in their book Report on Equality of Educational Opportunity. Then, educational efficiency began to be paid attention to and studied [5]. From 1990s, colleges and universities began to increase enrollmnet and the government began to increase investment in higher education in China. From then on, efficiency problem of higher education began to be attended to by macroscopic decision-makers and educational departments of all levels [6].

After decades of years, Chinese research on higher educational efficiency problem turns from qualitative research into quantitative research. At present, Data Envelopment Analysis – DEA is frequently used to study input-output efficiency of higher education. For example, Chen et al [7] applied DEA in studying input-output efficiency of west higher education. Guo [8] and Xu et al [9] applied CCR and BCC Models of DEA in studying input-output efficiency of higher education. Liang et al [10] applied DEA in studying operational efficiency of 28 colleges directly under the Ministry of Education. Zhang and Hu [11] applied

Received August 25, 2013; Revised October 6, 2013; Accepted October 26, 2013

SBM model in assessing higher educational efficiency of 31 provinces and cities. To sum up, DEA, espeically CCR and BCC are the most frequently used methods in studying higher educational efficiency. While, the obove models fail to fully consider slacking problem of inut-output. Meanwhile, the efficeincy values are not exact or even partial [12]. The SBM can perfectly solve this problem; however, little work has been done to the quantitative analysis on education efficiency using SBM model.

Therefore, to objectively assess the education efficiency in Jiangxi, the SBM model based data mining method has been employed to analyze historical data from 2000 to 2012. Empiriacl analysis shows that the SBM model can effectively solve the problem of slacking and may provide scientific support for related educational decision-making departments.

2. Research Method

Traditional measure models are generally based on Farrell's ideas of efficiency measurement which belongs to radial and piecewise linear measure theory. CCR and BCC are the case. Their strong operability may gurantee the convexity of effiiciency margin or indifference curve, but also may result in crowding or slacking of inputs [11]. To solve the problem, Tone in 2001 brought forward a better way to take slack variables into consideration of efficiency calculation, that is, SBM [13], whose expression is as follows:

$$\text{Min} \quad \rho = \frac{1-(\frac{1}{m})\sum_{i=1}^{m}\frac{s_i^-}{x_{io}}}{1+(\frac{1}{k})\sum_{r=1}^{k}\frac{s_r^+}{y_{ro}}} \quad , \tag{1}$$

$$\begin{cases} \text{s.t. } x_0 = X\lambda + s^- \\ y_0 = Y_\lambda + s^+ \\ \lambda \geq 0, s^- \geq 0, s^+ \geq 0 \end{cases} . \tag{2}$$

In which, ρ is efficiency assessment standard; m and k are respectively sorts of inputs and outputs; λ is column vector; x_0 and y_0 are respectively input vector and out vector of estimated decision-making-unit; x_{i0} and y_{r0} are respectively elemnets of vector x_0 and y_0; s_i^- is the element of slack input s-; , s_i^+ is the element of slack output s^+.

For the unit to be assessed, when and only when ρ = 1, namely, s^- = 0, s^+ = 0, the decisioan making unit (DMU) is thought to be effective for DEA. When the minimum value is obtained, if $\Sigma\lambda$ = 1, then the DMU's scale economy is thought to be unchanged; if $\Sigma\lambda > 1$, then the DMU's scale economy is thought to be increasing progressively; if $\Sigma\lambda > 1$, then the DMU's scale economy is thought to be decreasing progressively. For the convenience of solution, the above model is transformed into linear programming problem as follows:

$$\text{Min} \quad \tau = t - (1/m)\sum_{i=1}^{m} s_i^- / x_{i0} \tag{3}$$

$$\begin{cases} \text{s.t. } t + (1/k)\sum_{r=1}^{k} s_r^+ / y_{r0} = 1 \\ tx_0 = Y\Lambda + S^+ \\ ty_0 = Y\Lambda + S^+ \\ \Lambda \geq 0, \text{S}^- \geq 0, \text{S}^+ \geq 0, t > 0 \end{cases} . \tag{4}$$

Where, τ is the value of higher educational efficiency; $S^- = ts^-$, $S^+ = ts^+$, $\Lambda = t\lambda$.

3. Results and Analysis

3.1. Index Selection Using Fuzzy PCA

Educaion has strong nature of public goods. As a result, educational section is not one of material manufacturing, not an economic one either but non-profiting one whose total mission is to cultivate qualified personnel for the society. Thus, research on input-output efficiency of higher education is intended to guarantee reasonable allocation of educational resources.

However, how to select the inputs and outputs of the SBM model is still a challenge. Zhang and Hu [11] in 2010 took education as a manufacturing or economic activity, which inputs human, material and financial resources to output qualified personnel as well as scientific research fruits to promote imporvement of national quality and progress of the scoiety. They selected the following input indexes: number of staff; educational funds as well as educational fixed assets; output indexes: number of postgraduates, number of patents and number of graduates [11]. Other researchers consider the college's investment in scientific programs, college's social reputation, employment condition of graduates, etc. Hence, data mining technology is required to analyze the original data to determine the most important indexes. In this paper, we employ the intelligent method, Fuzzy PCA [14], to select the inputs and outputs of the SBM model.

The original data is taken from Jiangxi Statistical Yearbook during 2000-2012. The Fuzzy PCA was used to evaluate the importance degree of the indexes. Herein, we have compared the performance of the Fuzzy PCA and PCA. Table 1 lists the comparison results.

Table 1. Evaluation on the Importance Degree of the Indexes Using Fuzzy PCA and PCA

Index	Importance degree using Fuzzy PCA	Importance degree using PCA
Number of staff	51.3%	36.2%
Educational funds	22.7%	44.8%
College's investment in scientific programs	12.4%	5.5%
Educational fixed assets	5.6%	1.1%
College's social reputation	4.3%	1.3%
Employment condition of graduates	3.7%	11.1%

Table 2. Input-output Data of Jiangxi's Regular Higher Education of 2000-2012

	professional staff	Educational funds (100 millions)	Scientific programs (ten thousands)	Undergraduates	Patents	postgraduates
2012	50205	216.73	138633	851119	2140	25209
2011	49970	171.81	113108	828599	1275	23824
2010	49028	136.20	85548	816484	855	21313
2009	48637	107.97	70615	793488	587	17990
2008	47510	106.80	61901	764182	270	15304
2007	45153	94.22	51518	781686	179	13688
2006	42227	62.52	41135	770525	146	12149
2005	38587	51.04	30422	646086	62	9860
2004	30419	41.55	22661	489854	32	7483
2003	21938	35.74	10196	358622	25	5711
2002	15934	26.19	6220	266251	16	4123
2001	12176	17.04	3795	196455	10	2972
2000	10380	14.39	2315	144293	6	2118

It can be seen in Table 1 that the importance degrees of the number of staff and the educational funds are comfirmed by both the Fuzzy PCA and PCA. These results agree well with the discussion in [11], where the numbers of staff and the educational funds have been proven to be the most important indexes among the others. However, in the Fuzzy PCA, the scientific programs ranked in the third importance position while the employment condition of graduates was the third important index in PCA. According to the work repoted in [1], the employment condition may be important but its foundational support is the scientific programs. Hence, in fact, the index of scientific programs is more important than the employment condition. The comparison results indicate that the Fuzzy PCA can provide more accurate analysis than the PCA.

Based on the Fuzzy PCA result, in this paper number of staff, the educational funds, and the scientific programs are selected as the inputs of the SBM model. The outputs are number of postgraduates, number of patents and number of graduates.

Then, the SBM model is analyzed and the analysis results are given in Table 2.

3.2. Results and dicussions

The paper constructs SBM model in light of related data of Jiangxi's regular higher education from 2000 to 2012. It applies maxdea software and obtains relative efficiency as shown in Table 3.

Table 3. Relative Efficiency of Jiangxi's Higher Education of 2000-2012

Year	ρ	$\Sigma\lambda$	S_1^-	S_2^-	S_3^-	S_1^+	S_2^+	S_3^+
2000	1	1	0	0	0	0	0	0
2001	1	1	0	0	0	0	0	0
2002	1	1	0	0	0	0	0	0
2003	0.96	1.1	0	-2	-898	0	17	0
2004	0.87	1.7	-907	0	-8380	0	9	0
2005	1	1	0	0	0	0	0	0
2006	1	1	0	0	0	0	0	0
2007	0.88	1.2	0	-13	-11967	0	49	0
2008	0.85	2.3	0	-13	-19219	12464	92	0
2009	1	1	0	0	0	0	0	0
2010	1	1	0	0	0	0	0	0
2011	1	1	0	0	0	0	0	0
2012	1	1	0	0	0	0	0	0

Note: ρ is the efficiency value; S_1^-, S_2^- and S_3^- are respectively slack inputs of professional staff of regular higher education, higher educational funds and college's investment in scientific programs; S_1^+, S_2^+ and S_3^+ are respectively slack outputs of undergraduate number of regular higher education, number of licensed patents and postgraduate number;$\Sigma\lambda$ is the return of scale of DMUs.

It can be seen in Table 3 that:

(1) The efficiency value of Jianxi's higher education from 2000 to 2002 is 1, which shows that from 2000 to 2002 Jiangxi's higher education had been efficient. That is, professional teschers of regular higher education, higher educational funds and college's investment in scientific programs were without overplus; undergraduate number of regular higher education, number of licensed patents and postgraduate number were without slack output. Meanwhile, Σλ is 1, which shows that the scale economy of Jiangxi's higher education investment kept unchanging from 2000 to 2002.

(2) The efficiency value of Jianxi's higher education from 2003 to 2004 is less than 1, which shows that from 2003 to 2004 Jiangxi's higher education was not efficient and tended to decreasing progressively. The result basically agrees with Xiong Li's assessment in 2010 who applied the principal component analysis method in assessing input-output efficiency of Jiangxi's higher education from 2000 to 2007 [15]. Nevertheless, Xiong Li only found that the general efficiency of Jiangxi's higher education tended to decrease but she failed to sort out the indexes with slack inputs or slack outputs. According to Table 2, in 2003, higher educational funds and college's investment in scientific programs were not enough. As for output, colleges' licencesed patents were somewhat slack, that is, compared with those efficient years, the number in 2003 was relativley smaller. In 2004, the professional staff of regular higher education and college's investment in scientific programs were not enough. As a result, the output of licensed patents was somewhat slack, that is, compared with those efficient years, the output of patents in 2004 was somewhat fewer. In addition, the value of Σλ in the two years was bigger than 1, which shows that the scale economy of Jiangxi's higher educational investment was decreasing.

(3) The efficiency value of Jianxi's higher education from 2005 to 2006 is 1, which shows that Jiangxi's higher education was efficient from 2005 to 2006, both inputs and outputs were not slack and the scale economy of Jiangxi's higher educational investment was

unchanging. It shows also that the efficiency of Jiangxi's higher education decreased from 2003 to 2004 and then stayed stable, which may resulted from increase of higher educational enrollment in 2003 and 2004. Undergraduates of regular higher education zoomed but professional staff and educational funds did not increase correspondantly. After 2005, the enrollment stayed stable, professional staff and educational funds increased rapidly and higher education efficiency rose as result.

(4) The efficiency value of Jianxi's higher education from 2007 to 2008 was less than 1 and was decreasing, which shows that from 2007 to 2008 Jiangxi's higher education was not efficient. In 2007 and 2008, higher educational funds and college's investment in scientific programs were not enough, especially college's investment in scientific programs was seriously inenough. As for higher education output, licensed patents were somewhat slack in 2007 and 2008. Compared with those efficient years, undergraduates of regular higher education were not enough. In addition, the value of $\Sigma\lambda$ in the two years is bigger than 1, which shows that the scale economy of Jiangxi's higher educational investment was decreasing.

(5) The efficiency value of Jianxi's higher education from 2009 to 2012 is 1, which shows that from 2009 to 2012 Jiangxi's higher education was efficient; both input and output indexes arenot slack and the scale economy of Jiangxi's higher educational investment remained unchanging. It shows that as a result of increase of Jiangxi's hgiher education investment, higher educational efficiency tends to be stablized.

4. Conclusion

According to the emperical analysis, from 2000 to 2012, the change of Jiangxi's higher educational efficiency falls into 5 stages – being relatively efficient; decrease of realtivel efficienncy; increase of realtivel efficienncy; decrease of realtivel efficienncy and being relatively efficient, which shows that the changes of national policies and economy have great influence on Jiangxi's higher education. In recent years, as a result of steady increase of Jiangxi's hgiher education investment, higher education has been efficient. Nevertheless, as a result of decrease of students taking part in college entrance examination, increase of undergraduates of Jiangxi's regular higher education will surely slow down or even begin t decrease. Then, when Jangxi's education investment increases annually, to keep higher education's ralative efficiency, we should change investment forms, that is, investment in higher education shouldbe centered on educational funds and scientific research programs to promote outputs of licensed patents and postgraduates; and change investment mechanism. Compared with developed provinces, Jiangxi's investment in higher education is still inenough. Meanwhile, invesnment in higher education is easy to be influenced by outside economic enviroment. Therefore, we should change the existing mechansim of investment to open diversified financing channels and increase investment. We also need seek for more scientific evaluating mechansim of higher educational inputs and outputs to obtain more effective outputs from the same inputs.

References

[1] Wu X, Wang L, Yan L, Xue F. Simulation of Radar Track Based on Data Mining Techniques. *TELKOMNIKA Indonesian Journal of Electrical Engineering*. 2013; 11(7): 3780–3788.
[2] Boris M. Prediction and decision making in Health Care using Data Mining. *International Journal of Public Health Science*. 2012; 1(2): 69-78.
[3] Wu X, Wang L, Yan L, Xue F. Simulation of Radar Track Based on Data Mining Techniques. *TELKOMNIKA Indonesian Journal of Electrical Engineering*. 2013; 11(7): 3780–3788.
[4] Chen D, Zhou A. Research on Higher Education's Efficiency. *Heilongjiang Higher Education Research*. 2006; 10: 2-3.
[5] Tong H. *Research on School Management Efficiency*. PhD Thesis. Beijing: Beijing Normal University, 2001.
[6] Chen X, Zhan L. Reflections on the Development of Concept of Chinese Higher Education. *Higher Education Research*. 2009; 10: 12-17.
[7] Chen T, Bai J. Research on Evaluation of Relative Efficiency of China's Higher Education Input and Output. *China Higher Education Evaluation*. 2002; 2: 13-16.
[8] Guo X. Evaluation of Subjects Efectiveness Based on DEA. *Chinese Journal of Management Science*. 2003; 6: 77-80.
[9] Xu J, Wang S. Efficiency Evaluation of Chinese Regional Higher Education Based on DEA. *Research on Higher Engineering Education*. 2009; 4: 81-85.

[10] Liang Q, Peng X. Efficiency Evaluation of Research Universities. *Research on Higher Engineering Education*. 2009; 2: 83-86.

[11] Zhang M, Hu Y. Study on Evaluation of Efficiency of Chinese Higher Education in China Based on SBM. *Statistical Education.* 2010; 7: 22-30.

[12] Xu C. Efficiency Assessment and Cost Control of Higher Education Based on DEA. *Heilongjiang Higher Education Research*. 2012; 8: 17-18.

[13] Kaoru T. A Slacks Based Measure of Efficiency in Data Envelopment Analysis. *European Journal of Operational Research*. 2001; 130: 498-509.

[14] Heo G, Gader P, Frigui H. RKF-PCA: robust kernel fuzzy PCA. *Neural networks.* 2009; 22: 642-650.

[15] Xiong L. Comprehensive evaluation of Higher Education's Input and Output Efficiency. *China Science and Technology Innovation Herald*. 2010; 31: 13-17.