Forecasting the Quantity Trend of Human Resources with Higher Education in China Based on Correlation – Regression Analysis

Tang Wenling [1], Wang Huadong[2]
1 Department of Social Sciences, Handan College, Handan , P.R.China ,056005
2 Economics and Management School of Hebei University of Engineering, Handan P.R.China, 056038
(E-mail: whdjennes@126.com, lindaxyh@126.com)

Abstract: Degree education in China has entered a stage of the rapid development, which increases the quality of China's population and its employment capacity greatly; meanwhile, some negative effects are also presented. In order to macroscopically forecast personnel supply and coordinate employment work needs, regression software was employed to forecast the future personnel supply data in this paper. Because variables are so many and radix is so big between them, they were corrected by using the specific actual value comparison in the paper to make the predictive values have more persuasive power.
Key words: Regression analysis; Human resources; Higher Education; Forecast

1 Introduction

Since 1978, graduate education in China has developed more than 30 years, and has trained a large number of highly-educated students including post-graduates, who play an important role in China's social and economic development. China's economy has been badly affected by the impact of the global economic downturn in recent years, and employment of post-graduates is also severely affected under macroeconomic conditions. China has made efforts to make policies to support post-graduate employment, to adjust the number of graduate enrollment, and to make relevant policies to support other employment to reduce the employment pressure. But how to reduce graduated students' employment pressure is related to the number of the real needs of post-graduate, which is a study on macro-demand and it is what the government is doing now - the social demand forecast. Planned enrollment number is mainly considered, and the supply number is made prediction from the planned number. Due to the overall trend of population development, the time considered is a short-term one rather than a long-term one. Meanwhile, the losing situation from planned learning to graduation should be taken into account; eventually a rather rigorous prediction system is reached.

2 Summary of Human Resources Prediction Theories

Prediction theories of human resources include demand forecast of human resources, and supply forecast of human resources. Human resources trend forecast is about study of supply of human resources in the paper.

Supply forecast of human resources is the core of human resources planning, which predicts a certain amount of personnel, its quality, and structure provided in the internal organization and supplied in the external labor market at a time in the future to meet enterprises' staff requirement for achieving their goals. Human resources supply is divided into the external supply and the internal supply from sources of supply. The external supply refers to the study of personnel supply that the external labor market provides to organization. The internal supply refers to predict of human resources in future organization based on development of human resources in the internal organization and its used status. In most cases, the external supply cannot be understood or grasped by the organization, and so it can be achieved through strategic analysis of the local labor market, enterprises' employment conditions and their competitors. Thus, supply forecast mainly focuses on organization's human resources internal supply.

In recent years, rapid development of computer technology has brought changes in the field of modeling, powerful computing and statistical function has made huge-workload and high-precision models' actual operability greatly increased. For this reason, many scholars at home and abroad put forward many new internal supply predict models of human resources on the basis of previous studies. There are: multiple objective embedded network model, Holonic model etc.. In China, "modern control theory model" by Hai-ying Jing to predict human resource supply, and Li lei's "two-story integer programming model." These methods are not very innovative in thought, but have models more complicated, precise and perfect using advanced computer technology based on the original two ways--Markov and objective programming.

Starting from the actual situation, more mature quantitative analysis-- combination of correlation and regression analysis is used to close to predict to the great extent because of the variable selection limited, but the larger amount of vertical in this paper.

3 Summary of Analysis Methods

Two main analysis methods are introduced, one is correlation analysis, the other is regression analysis. Correlation analysis is to draw relativity degree among selected samples in order to do preparation for next correlation analysis, because there is no correlation analysis in typical supply predict methods of human resources. Introduction has to prove even though correlation analysis is only one part (relevant theory and practical test will be done in the following text).

3.1 Correlation coefficient

Correlation study is also known as canonical correlation analysis and it is a multivariate analysis method to study the correlation between two sets of variables. Principal components are extracted respectively from the two sets of variables according to the ideology of dimension reduction of principal components, and relativity degree between principal components is made largest, while each principal component extracted from the same group is independent, whole linear correlation of the two sets of variables is described by relativity degree between principal components. Canonical correlation analysis, first proposed by Hoteling in 1936, has become a widely used technology to analyze the correlation between two sets of variables now. The correlation coefficient between two sets of interrelated variables is called the overall correlation coefficient, which ρ is usually used as, the overall correlation coefficient can be calculated using formula (3.1):

$$\rho = \frac{Cov(X,Y)}{Var(X)Var(Y)} \tag{1}$$

Where: $Var(X)$ is variance of the variable X; $Var(Y)$ is variance of the variable Y; $Cov(X,Y)$ is covariance of the variables X and Y.

The overall correlation coefficients ρ reflects linear correlation degree between two variables X and Y in the whole, to specific entirety, values of X and Y are decided, the overall correlation coefficient ¡s an specific value objectively existing. However, all the values of X and Y in the whole can not be directly observed, so overall correlation coefficient is usually unknown。It is usually possible that certain numbers of samples are taken at random from the whole, and the sample correlation coefficient is estimated through the sample observed values x and y of X and Y. r_{XY} represents the sample correlation coefficient or simplified as r, can be used the following formula (3.2) to estimate:

$$r_{XY} = \frac{\sum (x_i - x)(y_i - \bar{y})}{\sqrt{\sum (x_i - \bar{x})^2 \sum (y_i - \bar{y})^2}} \tag{2}$$

Where: x_i and y_i are respectively the sample observed values of variables X and Y; $\bar{x}$ and $\bar{y}$ are respectively the average of the sample observed values of variables X and Y.

3.2 Regression analysis

Study of the linear relationship among two or more variables of is known as multivariate linear correlation analysis; Study of the linear relationship between the dependent variable and two or more independent variables is known as multiple linear regression analysis; Mathematical formula showing linear relations among several variables is a multiple linear regression model. Overall regression function describes the linear relationship between a dependent variable and several independent variables, so this overall regression function is a multiple linear regression model. General form of multiple overall linear regression function containing the dependent variable Y and $(k-1)$ independent variables such as X_2, X_3,$\cdots$, X_k is as follows:

$$Y_i = \beta_1 + \beta_2 X_{2i} + \beta_3 X_{3i} + \cdots + \beta_k X_{ki} + \mu_i \tag{3}$$

Where: $k-1$ is the number of variables; $(Y_i, X_{2i}, X_{3i}, \cdots, X_{ki})$ is the observed sample value for the i time; $\beta_j (j = 2, \ldots, k)$ is the model parameter; μ_i is the random error term.

All regression coefficients in the overall linear regression function are unknown, and can be only estimated using the sample observed values. If the sample conditional mean of dependent variables is expressed as the linear function of the various independent variables, multiple linear regression function is as follows:

$$\hat{Y}_i = \hat{\beta}_1 + \hat{\beta}_2 X_{2i} + \hat{\beta}_3 X_{3i} + \cdots + \hat{\beta}_k X_{ki} \quad (4)$$

Where: $\hat{\beta}_j (j = 1, 2, ..., k)$ are estimates of overall regression parameters β_j

4 Analysis of Forecast of Trend of Human Resources with Higher Education

Higher education refers to post-graduate education in this paper. Graduate education started in 1978 and has become a large one in China. Here, the data are from 2010 China Statistical Yearbook, in which the data from 1990 to 2009 are used, and the data in 2009 are retained in the calculation to compare for the purpose of logical rigor. Selected indexes of 1990-2009 post-graduate education conditions are number of post-graduate enrollment, school size, number of graduates. Specifications are shown in Table 1.

Table 1 Post-graduate Statistics in 1990-2009 in China

Year	Number of post-graduate enrollment	School size	Number of graduates
1990	29649	93018	35440
1991	29679	88128	32537
1992	33439	94164	25692
1993	42145	106771	28214
1994	50864	127935	28047
1995	51053	145443	31877
1996	59398	163322	39652
1997	63749	176353	46539
1998	72508	198885	47077
1999	92225	233513	54670
2000	128484	301239	58767
2001	165197	393256	67809
2002	202611	500980	80841
2003	268925	651260	111091
2004	326286	819896	150777
2005	364831	978610	189728
2006	397925	1104653	255902
2007	418612	1195047	311839
2008	446422	1283046	344825
2009	510953	1404942	371273

Correlation analysis of the relationship among three indexes in the table is carried on, their correlation coefficients after the software processing are shown in Table 2.

Table 2 Variables among Three Indexes

	Number of post-graduate enrollment	School size	Number of graduates
Number of post-graduate enrollment	1		
School size	0.99575501134	1	
Number of graduates	0.9471580403	0.97118471843	1

As can be seen from the table, there is a strong correlation among number of post-graduate enrollment, school size, number of graduates. In theory, correlation analysis among three indexes can be

carried out, that is regression analysis, and in practice, there actually exists relationship among them.

Multiple correlation regression analysis is made by software to the samples, and regression statistics values, variance analysis and estimated values are calculated as the following table.

Table 3 Regression Statistical Analysis

Regression statistics	
Multiple R	0.99496845315
R Square	0.98996222277
Adjusted R Square	0.98862385247
Standard error	9009.7136582
Observed value	18

Table 4 Analysis of Variance

	df	SS	MS	F	Significance F
Regression analysis	2	1.2008653272E+11	60043266361	739.677371	1.0286832009E-15
Residuals	15	1217624103.1	81174940.203		
Total	17	1.2130415682E+11			

Table 5 Analysis of estimated values

	Coefficients	Standard error	t Stat	P-value	Lower 95%	Upper 95%	Lower limit 95.0%	Upper limit 95.0%
Intercept	6676.1975	3307.0451	2.01878	0.0617	-3726023	13724.997	-372.6023	13724.997
X Variable 1	-1.4315	0.1712	-8.3594	4.9911848648E-07	-1.7965	-1.066	-1.7965	-1.066
X Variable 2	0.7434	0.0631	11.7794	5.5674727989E-09	0.6089	0.878	0.6089	0.878

Therefore, regression results are obtained as follows:

$$Y_i = 6676.19 - 1.4315X_1 + 0.7434X_2 \quad (5)$$

(4.1) is a regression correlation formula among number of graduates and number of enrollment and school size. Number of graduates next year can be predicted after in determining number of enrollment next year and school size that year using the formula (4.1).

5 Theory and Analysis of Practical Tests

Statistical capacity of the parameter t of each index are taken from the software analysis and then look up table

$$t_{(\alpha/2)}(n-k) = t_{0.025}(18-2) = 1.7459$$

Compared with the statistical capacity t in Table 4-5, it is obvious that individual has a significant impact on the regression model. Multiple coefficient of determination and adjusted multiple coefficient of determination is listed in Table 4-3, $R^2 = 0.9899$, $\overline{R}^2 = 0.9886$, it obviously conforms to the case that coefficient of determination is close to 1, and the model is right. Statistical capacity of F by software analysis is 739.677. $\alpha = 0.05$ is made according to software calculation, and then degree of freedom is the critical value $F_{0.05}(3-1, 18-3) = 3.68$ between $(k-1)$ and $(n-k)$ after checking the distribution list of F, and it is clear that the statistical capacity in this case is much greater than the critical value, indicating number of enrollment and school size together has a significant impact on number of graduates next year. The data in 2009 is introduced here (data is from the 2009 China Statistical Yearbook):

$$Y = 6676.19 - 1.4315 * 446422 + 0.7434 * 1283046 \approx 321440$$

According to the data of 2009 actual graduates, after actual comparison, the error is 6% and rather small. Forecasted data is credible through the above theory test--three kinds of parametric tests and the actual data comparison test. The forecasted results in certain years such as 2009, 2010, 2011 are credible in theory and in actual operation.

6 Conclusions

Process of forecasting number of China's human resources with higher education is very complex, and requiring very precise. Non-typical human resources forecasting method (supply forecasting method) -- multiple regression analysis is introduced in this paper, and the test is feasible through pre-test of the correlation. And then multiple regression analysis model is built and the relevant data is calculated in accordance with the existing data. Establishing the multiple regression equation, test is made using the theory parametric test and the actual data and so coupling rate is found very high. Therefore, another method is introduced in forecasting number of China's human resources with higher education, and can be provided as a related support in theory study.

References

[1] Yuan Wei, Pang Hao, Zeng wu-yi. Statistics [M]. Higher Education Press, 2006 (In Chinese)
[2] Zhou Sanduo. Management [M]. Higher Education Press, 2005 (In Chinese)
[3] Li Zinai. Econometrics [M]. Higher Education Press, 2007 (In Chinese)