

Research on the Effectiveness of the Confucius Institute (Classroom) Based on Linear Regression Models

Fei Song

School of Chinese, Beijing International Studies University, China

Minghui Xu

School of Chinese, Beijing International Studies University, China

Abstract—China has invested heavily in development of the Confucius Institute (Classroom) for the going-global of Chinese language and culture. And its effectiveness evaluation is an essential reference to the reinvestment on the Confucius Institute (Classroom) development as well as to its budget reallocation. Considering that there were basically no such researches in this field ever before, linear regression models (LRMs) were employed in this paper to research the effectiveness of Confucius Institute and establish fitting function models between inputs and outputs, which could provide a tool to quantitatively evaluate the effectiveness in the future. And in this way, the conclusions could be more objective and bases of resource redistribution more scientific. Based on current data, it is found that the growth of the number of government-sponsored Chinese teachers and volunteers lags behind that of the "rigid Chinese language learners" for more than 2 years; If China invests 1,000 Yuan in the project of "the Salary and Training Fee of Chinese Deans and Teachers (including Volunteers) ", the number of the "rigid Chinese language learners" will rise by 10 to 13; if there is an additional overseas test center in the Confucius Institute (Classroom), the "rigid Chinese language learners" of this semester will rise by more than 5,927; and if there is an additional registered student in the Confucius Institute (Classroom), it will rise by more than 6.

Index Terms—Confucius Institute (Classroom), inputs, outputs, effectiveness

I. RESEARCH THINKING

Linear regression (LR) analysis, as an important statistical analysis method widely applied into the studies of variables correlations both in society and economy, is used to analyze the quantitative correlation among objects. Multiple linear regression models (MLRMs), in which two or more independent variables are employed to explain dependent variables, require the least square method (LSM)^① to calculate parameters on the premise of the minimized residual square sum (RSS) as a unary linear regression equation does to estimate parameters.

The matter of the Confucius Institute (Classroom) effectiveness belongs to the category of both sociology and economics, and it is closely related to the international popularization of Chinese language. Various inputs into the Confucius Institute (Classroom), acting as independent variables, have their own different influence on the going-global of Chinese language and culture, so MLRMs are applicable to the quantitative evaluation of the Confucius Institute (Classroom) effectiveness.

Since the first Confucius Institute was established in 2004, China has conducted enormous investment in the Confucius Institute (Classroom). But there is still lack of feasible researches on the specific, arguable and quantitative correlation between the investment and Chinese international popularization. Thus, MLRM is employed to analyze the quantitative correlation among the national funds, resource inputs and "rigid Chinese language learners"^② in this paper to initially study the effectiveness of the Confucius Institute (Classroom). Furthermore, the issue of how to build fitted function between the inputs and outputs is discussed to assess the input-output efficiency and to allocate funds and resources rationally.

II. INPUT-OUTPUT DATA RESOURCES AND REVISION OF THE CONFUCIUS INSTITUTE (CLASSROOM)

A. Inputs

China invests in many aspects of the Confucius Institute (Classroom) such as capitals, personnel, arenas and devices. And the related public data is limited and mainly shown in the "Hanban/Confucius Institute Headquarters Annual Report" (referred to as the "Report"). At present, the "Report" from 2007 to 2015 are available, but it is hard to conduct

^①LSM is a kind of mathematical optimization technique to find out the optimum function matching of data by minimizing the error sum of squares.

^②Referring to "Chinese tests examinees" from Hanban/Confucius Institute Annual Report over the years and the reason for such selection will be specified thereafter.

comparative studies due to vast difference from the previous fund allocation to projects in the “Report” from 2013 to 2015. Therefore, all the data related to fund are selected from capital expenditures of the key projects published in the “Reports” from 2007 to 2012. During these 5 years, there was some slight difference among the items of expenditures successively published in the “Report”. Nevertheless, in consideration of scarce data so far since its establishment, corresponding revisions are made to add the credible observed values as many as possible:

Firstly, project “the Confucius Institute (Classroom) Start-up” and “the Confucius Institute (Classroom) Operation” from 2009 to 2012 is merged into “the Confucius Institute (Classroom) Construction”;

Secondly, project “the Training and Dispatching of Volunteer Chinese Teachers” in 2007 is included in “the Salary and Training Fee of Chinese Deans and Teachers (including Volunteers)”;

Thirdly, project “Sending Great Wall Chinese Soft Hardware and Training to the Confucius Institute” and “Subsidizing the Confucius Institute to Build Experience Center of Chinese Culture in 2009, as well as “Confucius Institute Online” from 2010 to 2012 are merged into “Multimedia Construction”;

Fourthly, project “Chinese Bridge Competition and Fund”, “Chinese Experience Activities for Foreigners” in 2007 and 2008 as well as “Chinese Teaching Materials Exhibition” and “Literature and Art Tour” in 2009 are merged into “the Tour of Show, Exhibition and Speech”;

Fifthly, project “the Examination Project Development” and “Others” in 2007 and “the Examination Project Development”, “Chinese Popularization Study and Practice Base Construction of Chinese Colleges and Universities” and “Others” in 2008 and “the Confucius Institute Conference” in 2009 and 2010 are merged into “Domestic Infrastructure Construction”;

Sixthly, project “International Conferences” in 2007 and 2008, as well as “Regional Joint Conference such as in North America, Asia, Oceania, Africa, and Europe, Eurasia, and Ibero-America” and “the Confucius Institute Conference” in 2009 and 2010 are merged into “the Regional Joint Conference and the Confucius Institute Conference”;

Seventhly, as for the unit of measurement, unit “Thousand Yuan” was adopted for capital expenditure data in the Reports from 2007 to 2009; then it became “Thousand Dollar” from 2010 to 2012. And, it is converted into “Thousand Yuan” uniformly based on the RMB/USD exchange rates^③ from 2010 to 2012.

Eighthly, the data is converted into comparable prices with the year 2007 as the base period according to the domestic inflation rates^④ in China from 2007 to 2012 to ensure the comparability.

Fund investments of China in the Confucius Institute (Classroom) over the years after adjustment are shown in Table 1.

TABLE 1
CLASSIFICATION OF FUND INVESTMENTS OF CHINA IN THE CONFUCIUS INSTITUTE (CLASSROOM) OVER THE YEARS

Unit: Thousand Yuan	2007	2008	2009	2010	2011	2012
The Confucius Institute (Classroom) Construction	213,370	368,297	516,963	295,686	290,289	393,864
Salary and Training Fee of Chinese Deans, Teachers and Volunteers	98,620	197,925	268,098	222,935	347,427	412,292
Scholarship of the Confucius Institute			143,673	202,469	165,088	139,419
Multimedia Construction			82,409	52,806	38,198	30,977
The Tour of Show, Exhibition and Speech	93,820	62,955	78,700	24,527	28,361	36,775
Textbook development and popularization	28,780	40,648	47,179	33,836	42,223	20,428
The Regional Joint Conference and the Confucius Institute Conference	10,400	13,481	9,994	23,584	19,495	
Chinese and Foreign Experts as Field Supervisor			7,570	7,367	7,067	8,318
<i>The Confucius Institute</i> (Magazine)			1,236	2,715	10,890	10,860
Domestic Infrastructure Construction	14,850	90,285				
Total	459,840	773,600	1,168,004	865,926	949,045	1,052,933

Other resource inputs and relevant variables except funds are also subject to the published “Report” and will be listed thereafter.

B. Outputs

There are many outcomes of the Confucius Institute (Classroom). Constitution and By-Laws of the Confucius Institute clearly stipulates that the Confucius Institutes are non-profit educational institutions, and it devotes to enhancing understanding of the Chinese language and culture by these peoples who learn the Chinese language in the world, to deepening friendly relationships with other nations, to facilitating the development of multi-culturalism, and to constructing a harmonious world. It mainly provides the following services for people from all walks of life around the world: Teaching Chinese language; Training Chinese language instructors; Holding the HSK examination and tests for the Certification of the Chinese Language Teachers; Providing information and consultative services concerning

^③RMB/USD exchange rate: 6.8281 in 2010; 6.6215 in 2011; and 6.3001 in 2012.

^④Domestic inflation rate in China: 5.90% in 2008; -0.70% in 2009; 3.30% in 2010; 5.40% in 2011; and 2.60% in 2012.

China's education, culture, and economy, society and so forth; Conducting studies about contemporary China.⁶ In general, its primary task is to contribute to the global dissemination of Chinese language and culture, and the main approach is the Chinese language teaching. That's to say, the effectiveness of the Confucius Institute (Classroom) can be evaluated according to the increase of Chinese language learners to a large extent. However, it is hard to explicitly define "the Chinese language learners". Undoubtedly, HSK examinees and those foreign college students who minor in Chinese are Chinese language learners. But as for such foreigners that pick up several Chinese daily dialogues on-line on a sudden pulse and then have no further understanding of Chinese culture actively, or even worse, forget all the stuff, it is hard to define whether they are "Chinese language learners" or not. Even if all of them are counted as "Chinese language learner", their effects on the dissemination of Chinese language and culture are not the same, either.

Thus, the third group of people abovementioned is excluded as "the rigid Chinese language learners" in this paper. And it is worth noting that "the rigid Chinese language learners" should refer to the examinees attending any Chinese proficiency test, including HSK examinees. Namely, the research objects refer to the Chinese test examinees as "the rigid Chinese language learners", HSK examinees included. The reason is that formal Chinese language learners almost have a certain pattern of Chinese test whether they attend Chinese proficiency tests like HSK or not, which makes it more comprehensive and objective. It is more a kind of research thinking than a perfect way, and it will play a greater role when the segmentation researches on its effectiveness are conducted and a certain amount of data is accumulated.

III. LRM ANALYSIS OF EFFECTIVENESS OF THE CONFUCIUS INSTITUTE (CLASSROOM)

A. Variable Names

Relevant variable names in this research and the respective item of fund investment in the Confucius Institute by the corresponding country are shown in the table below:

TABLE 2
SPECIFICATION OF VARIABLES

Variable Type	Variable Name	Project Name	
Explained Variable	Examinee	Chinese Test Examinees	
Explanatory Variables	Fund Variables	Confound	The Confucius Institute (Classroom) Construction
		Teacher	Salary and Training Fee of Chinese Deans, Teachers and Volunteers
		Scholarship	Scholarship of the Confucius Institute
		Multimedia	Multimedia Construction
		Exhibition	The Tour of Show, Exhibition and Speech
		Textbook	Textbook development and popularization
		Conference	The Regional Joint Conference and the Confucius Institute Conference
		Expert	Chinese and Foreign Experts as Field Supervisor
		Magazine	<i>The Confucius Institute</i> (Magazine)
		Basis	Domestic Infrastructure Construction
	Sum	Total investment	
	Non-fund Variables	Stud	Registered Students in the Confucius Institute
		Teavol	The Number of Government-sponsored Chinese Teachers and Volunteers
		Test Center	The Number of Overseas Test Centers

B. Correlation between the Rigid Chinese Language Learners and Funds

1. Unary regression analysis

There are at most 6 observed values for each variable but 10 variables, so MLRM is inapplicable due to limited observed values. Instead, correlation between Chinese language learners and various investments is analyzed by way of stepwise regression. LSM is required to conduct the unitary regression for the explanatory variables (investment items of China) first. To save pace, key reference values, instead of all the regression results, are shown in the table below:

⁶Hanban, Constitution and By-Laws of the Confucius Institutes [EB/OL]. http://world.jyb.cn/gjzl/200903/t20090316_255438.html.2013/12/2

TABLE 3
KEY REFERENCE VALUES OF UNARY REGRESSION ANALYSIS

Variable	Effective Observed Value	Coefficient	\bar{R}^2 [®]	T Statistic [®]	F Statistic [®]	p Value [®]
Confund	6	1.831303	-0.22321	0.295974	0.087601	0.78198
Teacher	6	10.69194	0.777277	4.295277	18.4494	0.012693
Scholarship	4	-24.05158	-0.12453	-0.81718	0.667779	0.499687
Multimedia	4	-50.44274	0.530542	-2.09532	4.390354	0.171128
Exhibition	6	-27.33616	0.192848	-1.48142	2.194618	0.212616
Textbook	6	-69.80915	0.090492	-1.22371	1.497477	0.288213
Conference	5	68.16858	0.065673	1.131881	1.281154	0.33998
Expert	4	1653.628	0.107895	1.167405	1.362833	0.363397
Magazine	4	238.9231	0.694173	2.794541	7.809462	0.107747
Basis	2	2.280109	NA	NA	NA	NA

It is clearly shown in Tab. 3 that Teacher has the highest \bar{R}^2 of 0.777277, indicating its good fitting with the sample; its P value is less than the significant level of 0.05 after T test and F test, without regards to the insufficient observed values of Basis. And no other variables pass T and F tests except Teacher and Magazine, but P value of Magazine exceeds 0.05 with fewer observed values, so only Teacher ("Salary and Training Fee of Chinese Deans, Teachers and Volunteers ") can be counted as the first explanatory variable for the unary LRM.

2. Binary regression analysis

Corresponding explanatory variables of the rest projects are incorporated into LRM successively to get regression results, and key reference values are shown in the table below:

TABLE 4
KEY REFERENCE VALUES OF BINARY REGRESSION ANALYSIS

Note: In the columns of Coefficient and T statistics, the former value represents the coefficient of the first variable "Chinese Dean, Teacher and Volunteer's Salary and Training Fee" in LRM, and the latter one represents the coefficient of the corresponding variable on the left.

Variable	Effective Observed Value	Coefficient		\bar{R}^2	T Statistic		F Statistic	p Value
Confound	6	12.29997	-3.88946	0.833173	5.131167	-1.52978	13.48559	0.031668
Scholarship	4	19.00571	13.76373	0.879481	4.202558	1.044003	11.94618	0.200432
Multimedia	4	12.17046	-19.1138	0.900067	2.897495	-1.23307	14.5101	0.182513
Exhibition	6	10.7498	0.340742	0.703088	2.806055	0.022841	6.91999	0.075192
Textbook	6	10.01494	-53.5136	0.964322	9.948291	-4.68725	68.57118	0.003132
Conference	5	3.774059	110.4346	0.573509	1.63896	2.013204	11.05513	0.082952
Expert	4	14.42827	381.2142	0.792483	2.756427	0.462349	6.728317	0.263006
Magazine	4	12.95354	51.59157	0.771153	1.293353	0.317231	6.054587	0.276193
Basis	4	NA	NA	NA	NA	NA	NA	NA

It can be known from Tab. 4 that only LRM incorporated with Textbook passes both T and F tests, and its P value is less than 0.05. Besides, its \bar{R}^2 is 0.964322 and higher than that of Teacher in the unary LRM, also with a large amount of F statistics, showing LRM fits the sample very well and that the regression equation is significant. Meanwhile, their T values are also large and the corresponding P value is less than 0.05, indicating significant influence of the explanatory variable on the explained variable. Thus, Teacher ("the Salary and Training Fee of Chinese Deans and Teachers and Volunteers" project confund) and Textbook ("Textbook development and popularization" project confund) should be reserved in LRM in a method of the stepwise regression.

3. Ternary regression analysis

Corresponding explanatory variables of the rest projects are incorporated into LRM continuously to get regression results, and key reference values are shown in the table below:

[®] \bar{R}^2 , also known as coefficient of determination, referring to the proportion of ESS in TSS, can be used as the goodness of fit measurement index of comprehensive measurements regression model for the sample observed values. It illustrates that the greater it is, the larger the LRM explained proportion in TSS is and the better the LRM goodness of fit is and vice versa.

[®] Referring to the test statistic of T test that is targeted at the significance test of regression coefficient, and it shows that the test is statistically significant and rejects the null hypothesis when T statistics obviously exceeds the corresponding critical value in the table.

[®] F statistic refers to the ratio between variance and residual of LRM, indicating that the greater it is, the smaller the residual is and the higher the simulative precision is. It shows good fitting between LRM and the sample and significant regression equation when F statistics exceeds the required critical value in the significance test.

[®] P value refers to the probability occurrence of observed or extreme result acquired from the sample in case the null hypothesis is true. The smaller P value is, the smaller the probability occurrence of the null hypothesis is, and in that case, the null hypothesis can be rejected on the fuller grounds of the small probability principle. In a word, the smaller P value is, the more significant the result is.

TABLE 5

KEY REFERENCE VALUES OF TERNARY REGRESSION ANALYSIS

Note: They represent "the Salary and Training Fee of Chinese Deans and Teachers and Volunteers " project, "Textbook development and popularization" project and the corresponding confound from top to bottom for the three lines in the column of Coefficient and T statistics

Variable	Confound	Scholarship	Multimedia	Exhibition	Conference	Expert	Magazine	Basis
Effective Observed Value	6	4	4	6	5	4	4	4
Coefficient	10.72363	14.69519	11.7421	10.34539	15.31535	13.14286	10.08411	NA
	-46.2187	-32.87809	-31.5313	-53.8037	-113.795	-49.8387	-39.3368	NA
	-1.490935	6.224683	-8.46491	1.96778	-53.3405	-342.723	48.38248	NA
\bar{R}^2	0.967883	NA	NA	0.949055	0.906896	NA	NA	NA
T Statistic	9.444728	NA	NA	6.50572	4.436546	NA	NA	NA
	-3.685539	NA	NA	-3.93499	-2.85685	NA	NA	NA
	-1.154409	NA	NA	0.317736	-1.52061	NA	NA	NA
F Statistic	51.22744	NA	NA	32.04811	13.98753	NA	NA	NA
p Value	0.019208	NA	NA	0.030411	0.193496	NA	NA	NA

It can be known from Tab. 5 that excluding those variable without sufficient observed values, the ternary LRM incorporated with Confund has the highest \bar{R}^2 of 0.967883 and is higher than that of Textbook in the binary LRM; F statistics is the largest and the corresponding P value is 0.019208 which is less than 0.05, indicating that the combined explanatory variables have significant effects on the number of Chinese language learners,. However, as for the sole Confund, its T statistics is -1.154409 and the P value is 0.3676, showing less significant influence on the explained variables. On the whole, LRM incorporated with Confund ("the Confucius Institute (Classroom) Construction" project) is still the optimum choice. And the less significant influence might be caused by multicollinearity, which can be verified through correlation coefficient test methods. At present, relevant coefficient matrix of these three variables is listed as below:

TABLE 6
CORRELATION COEFFICIENT MATRIX

Correlation Matrix			
	TEACHER	TEXTBOOK	CONFUND
TEACHER	1	-0.143474262301891	0.438507281278756
TEXTBOOK	-0.143474262301891	1	0.385268638519961
CONFUND	0.438507281278756	0.385268638519961	1

It shows a relatively close correlation between Confund and the other two variables in Tab. 6. Thus there is a certain kind of multicollinearity that is a very common and concerns a matter of degree. But a stepwise regression method is adopted to reduce its severity relatively, so Confund is used as the third variable to be incorporated into LRM for the present.

4. Quaternary regression analysis

Corresponding explanatory variables of the rest projects are incorporated into LRM continuously to get regression results, and key reference values are shown in the table below:

TABLE 7
KEY REFERENCE VALUES OF QUATERNARY REGRESSION ANALYSIS

Variable	Scholarship	Multimedia	Exhibition	Conference	Expert	Magazine	Basis
Effective Observed Value	4	4	6	5	4	4	4
Coefficient	NA	NA	12.76466	11.32882	NA	NA	NA
	NA	NA	-40.97447	-29.10264	NA	NA	NA
	NA	NA	-2.81592	-3.254836	NA	NA	NA
	NA	NA	8.403682	-33.54625	NA	NA	NA
\bar{R}^2	NA	NA	0.995768	NA	NA	NA	NA
T Statistic	NA	NA	18.74605	NA	NA	NA	NA
	NA	NA	-8.607439	NA	NA	NA	NA
	NA	NA	-4.80395	NA	NA	NA	NA
	NA	NA	3.76551	NA	NA	NA	NA
F Statistic	NA	NA	295.1391	NA	NA	NA	NA
p Value	NA	NA	0.043626	NA	NA	NA	NA

It can be known from Tab. 7 that owing to limited observed values, the regression analysis is not applicable to Scholarship, Multimedia, Conference, Expert, Magazine and Basis; and \bar{R}^2 , T statistic, F statistics and P value of Conference cannot be solved, which is also not authentic. And only Exhibition is an exception that can be utilized to finish the quaternary regression analysis. That's to say, Exhibition might be the fourth explanatory variable for LRM,

and based on its reference values, it is certainly a wise choice: R^2 is 0.999154 and \bar{R}^2 0.995768, with an obvious increase, and the fitting value and the actual observed value is very close, indicating high fitting degree; F statistic is 295.1391 and P value is 0.043626, less than 0.05, suggesting significant regression equation; besides, all the explanatory variables including Confund pass T test, on which Confund has a significant influence.

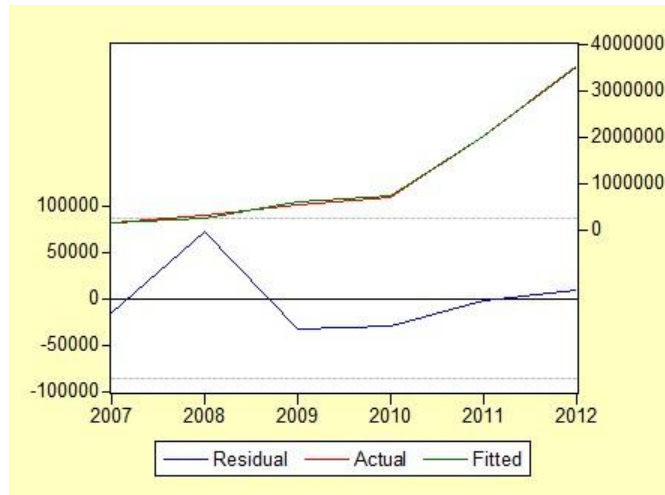


Fig. 1 Observed value, fitted value and residual plot

Examinee, Teacher, Textbook, Confund and Exhibition are represented by y , x_1 , x_2 , x_3 , x_4 , respectively, and MLRM obtained to reflect the effectiveness of the Confucius Institute is shown as below:

(Equation 1)

$$Y = 12.76465745X_1 - 40.97446804X_2 - 2.815920437X_3 + 8.403681619X_4 - 113975.9637$$

$$(0.680925) \quad (4.760355) \quad (0.586168) \quad (2.231751) \quad (261342.3)$$

$$t = (18.74605) \quad (-8.607439) \quad (-4.80395) \quad (3.76551) \quad (-0.436118)$$

$$R^2 = 0.999154 \quad \bar{R}^2 = 0.995768 \quad F = 295.1391 \quad p = 0.043626$$

It is worth noting that this MLRM is obtained on the premise that Scholarship, Multimedia, Conference, Expert, Magazine, Basis and Conference are excluded from the multiple regression equation due to their insufficient observed values instead of multicollinearity. Thus, it is hard to say that the present model can be instantly applied to predict the correlation between input and outputs of the Confucius Institute. There are two negative coefficients in the equation, so investment will be guided to flow to those effective and beneficial projects by the equation; or otherwise, the budgets of these projects would be tight subject to definite funds investment from China each year, and consequently, the overall effectiveness will be reduced. Nevertheless, owing to insufficient observed values, it may also happen when the explanatory variables are not thorough and comprehensive. There are two solutions: I, by accumulating data to complete the entire regression analysis; II, in a mathematical method to adjust symbols if the aforementioned method is proved useless (Xie, 2006). Thus, no further explanations on the equation adjustment or the symbolic matter are made; instead, conclusions are continued with the enlarging observed data.

C. Correlation between the Rigid Chinese Language Learners and Non-fund Variables

The comparatively complete annual non-fund variables based on data from the “Report” are shown as below:

TAB. 8
ANNUAL NON-FUND INVESTMENT OF CHINA

	Chinese Test Examinees	Confucius Institutes	Registered Students	Countries	Government-sponsored Chinese Teachers and Volunteers	Trainers of Foreign and Local Chinese Teacher	Overseas Test Centers
2005	36462	NA	NA	NA	1035	10597	
2006	72924	122	NA	49	2054	15896	106
2007	138000	202	NA	66	2977	16782	116
2008	310000	249	130000	78	4000	16512	128
2009	548000	282	260000	88	4800	23000	161
2010	689000	322	360000	96	6099	NA	204
2011	2010000	358	500000	105	6815	32319	488
2012	3520000	400	655000	108	7982	11527	640

Unary regression analysis is conducted for the variables in Tab. 8, and it is found that there is a high degree of regression fitting between the explanatory “Overseas Test Centers” (Test Centers) and the explained “Chinese Test Examinees”, of which the \bar{R}^2 reaches 0.973835, the corresponding T statistics 14.97714, F statistics 224.3146, and p Value only 0.000024, with a very significant regression equation:

(Equation 2)

$$\begin{aligned} \text{EXAMINEE} &= 5927.512541 * \text{EXAMPOINT} - 519497.3733 \\ &\quad (395.7708) \quad (130094.3) \\ t &= (14.97714) \quad (-3.993239) \\ R^2 &= 0.978196 \quad \bar{R}^2 = 0.973835 \quad F = 224.3146 \quad p = 0.000024 \end{aligned}$$

Then, the best regression fitting is the explanatory “Registered Students of the Confucius Institute (Classroom)” (Stud), of which the \bar{R}^2 is 0.896053, the corresponding T statistics 5.085360, F statistics 25.86089, and p Value only 0.014694, with a high degree of fitting and significance for the equation:

(Equation 3)

$$\begin{aligned} \text{EXAMINEE} &= 6.246013627 * \text{STUD} - 964331.1917 \\ &\quad (1.228234) \quad (519101.6) \\ t &= (5.085360) \quad (-1.857693) \\ R^2 &= 0.896053 \quad \bar{R}^2 = 0.861404 \quad F = 25.86089 \quad p = 0.014694 \end{aligned}$$

According to these two regression analyses, it can be clearly concluded that, as the separate explanatory variable, “Overseas Test Centers” and “Registered Students of the Confucius Institute (Classroom)” have the most significant influence on “the Chinese Test Examinees”. With one additional overseas test center, the number of Chinese test examinees will increase by more than 5,927; and with one additional registered student in the Confucius Institute (Classroom), it will increase by more than 6, which is completely compatible with the estimations and further verifies that the effectiveness researches of the Confucius Institute are feasible, scientific and objective.

To confirm the correlation between “the Salary and Training Fee of Chinese Deans and Teachers (including Volunteers)” and “the Chinese Test Examinees” above, the government-sponsored Chinese teachers and volunteers are merged into a new variable Teavol, and its regression result to “the Chinese Test Examinees” is also satisfactory.

(Equation 4)

$$\begin{aligned} \text{EXAMINEE} &= 435.6703433 * \text{TEAVOL} - 1032007.102 \\ &\quad (108.9934) \quad (545712.0) \\ t &= (3.997216) \quad (-1.891121) \\ R^2 &= 0.726996 \quad \bar{R}^2 = 0.681496 \quad F = 15.97774 \quad p = 0.007142 \end{aligned}$$

Hereinto, the \bar{R}^2 is 0.726996, T statistics 3.997216, F statistics 15.97774, and p Value 0.007142, with a high degree of fitting and significance for the equation, which fully demonstrates the positive influence of government-sponsored Chinese teachers and volunteers on the rigid Chinese language learners.

Besides, by analyzing the condition that government-sponsored Chinese teachers and volunteers lag behind the rigid Chinese language learners, it is found that:

(Equation 5)

$$\begin{aligned} Y &= 1560.670902 - 0.0006649876738X + 0.0007414071131X_{-1} + 0.004779515108X_{-2} \\ &\quad (125.5573) \quad (0.000353) \quad (0.000428) \quad (0.001070) \\ t &= (12.42995) \quad (-1.886218) \quad (1.731901) \quad (4.467799) \\ R^2 &= 0.985150 \quad \bar{R}^2 = 0.962874 \quad F = 44.22556 \quad p = 0.022193 \end{aligned}$$

Of which Y represents government-sponsored Chinese teachers and volunteers and X represents the rigid Chinese language learners. In Equation 6, with the increase of lag order, coefficient of every lag phase for X also increases, showing it will take some time to gradually affect the number of government-sponsored Chinese teachers and volunteers when the number of Chinese language learners changes. In the meantime, corresponding T statistics at every lag phase also keeps enlarging and exceeds 0.05 since the delay of 2 years. The degree of fitting for the whole equation reaches up to 0.985150 with p value less than 0.05 and it passes F test.

It indicates that the growth of government-sponsored Chinese teachers and volunteers lags behind that of the rigid Chinese language learners for more than 2 years, and the lagging intervals cannot yet be judged.

IV. CONCLUSIONS

Despite some problems for objective reasons of data in this research, several conclusions have been made as below:

First, compared with the rest 6 variables, four variables (fund investment of “the Salary and Training Fee of Chinese Deans and Teachers (including Volunteers)”, “Textbook development and popularization”, “the Confucius Institute (Classroom) Construction”, and “The Tour of Show, Exhibition and Speech”) in the equation are combined to leave a significant influence on the explained “Chinese test examinees” at present, which is confirmable even irrespective of these 6 variables.

Second, whether in the unary regression analysis or multiple regression analysis, “the Salary and Training Fee of Chinese Deans and Teachers (including Volunteers)” by China maintains a stable correlation with “Chinese test examinees” and its coefficient lies within the interval [10.01494, 12.76466]. Thus, it can be concluded that: effects of this input on the rigid Chinese language learners are less affected by other variables; based on the former deduction, there will be another 10 to 13 “rigid Chinese language learners” every time China invests 1,000 Yuan in this project.

Third, “Overseas Test Centers” and “Registered Students of the Confucius Institute (Classroom)” significantly affect the rigid Chinese language learners, respectively. It is estimated that the number of the “rigid Chinese language learners” will rise by more than 5,927 with an additional overseas test center; or that it will rise by more than 6 with one additional registered student in the Confucius Institute (Classroom).

Fourth, there is a lag phase for the growth of government-sponsored Chinese language teachers and volunteers to that of the rigid Chinese language learners for at least 2 years based on data, and more accurate delay intervals remain to be inspected when the observed value is complete. To a certain degree, it reflects a lack of Chinese language teachers around the world, which should be solved attentively.

Fifth, a method of LRM is put forward to research the effectiveness of the Confucius Institute (Classroom). But subject to insufficient observed values at present, it is unlikely to build a model with high robustness[®] and complete fitting of the reality. From the perspective of statistics, various data is reaching the critical point for some types of mathematical modeling according to the observed values. Therefore, it has great significance to accumulate operational data as planned for economic analyses. As the data related to the Confucius Institute development is kept accumulating and improving, the final goal is predictable, which will offer a solid theoretical basis for the rapid, scientific and sustainable development of the Confucius Institute.

REFERENCES

- [1] Hanban. Constitution and By-Laws of the Confucius Institutes [EB/OL] http://world.jyb.cn/gjzl/200903/t20090316_255438.html. (accessed 2/12/2013).
- [2] Hanban. Hanban/Confucius Institute Headquarters Annual Report (2006) [EB/OL] http://www.hanban.edu.cn/report/pdf/2006_final.pdf. (accessed 2/12/2013).
- [3] Hanban. Hanban/Confucius Institute Headquarters Annual Report (2007) [EB/OL] http://www.hanban.edu.cn/report/pdf/2007_final.pdf. (accessed 2/12/2013).
- [4] Hanban. Hanban/Confucius Institute Headquarters Annual Report (2008) [EB/OL] http://www.hanban.edu.cn/report/pdf/2008_final.pdf. (accessed 2/12/2013).
- [5] Hanban. Hanban/Confucius Institute Headquarters Annual Report (2009) [EB/OL] http://www.hanban.edu.cn/report/pdf/2009_final.pdf. (accessed 2/12/2013).
- [6] Hanban. Hanban/Confucius Institute Headquarters Annual Report (2010) [EB/OL] http://www.hanban.edu.cn/report/pdf/2010_final.pdf. (accessed 2/12/2013).
- [7] Hanban. Hanban/Confucius Institute Headquarters Annual Report (2011) [EB/OL] http://www.hanban.edu.cn/report/pdf/2011_final.pdf. (accessed 2/12/2013).
- [8] Hanban. Hanban/Confucius Institute Headquarters Annual Report (2012) [EB/OL] http://www.hanban.edu.cn/report/pdf/2012_final.pdf. (accessed 2/12/2013).
- [9] Xie Zhongqiu. (2006). Construction of LRM with Non-negative coefficients. *Statistical Education*, 1, 22-26.

Fei Song was born in Linyi, China in 1986. He received his PH.D. degree in linguistics from Minzu University of China in 2015.

He is currently a lecturer in the Chinese Academy, Beijing International Studies University, Beijing, China. His research interests include natural language processing and cognitive linguistics.

Dr. Song is a member of the Association for Modernization of Chinese Language Education.

Minghui Xu was born in Cangzhou, China in 1991. He is currently a postgraduate in the School of Chinese, Beijing International Studies University, Beijing, China.

[®]Robustness: generally used to describe the stability of a model or system; namely, a model or system is relatively stable even if encountered with a certain distraction.