Empirical Analysis on Regional Innovation: Base on Innovative Enterprises in Fujian Province of China

Lv Linfeng, Chen Yalan
Soft Science Research Institute, Fuzhou University, Fuzhou, P.R.China, 350002
(E-mail: linfeng157@gmail.com, chenyl1965@163.com)

Abstract: This paper analysis the cross-section data of innovative enterprises (pilot) in Fujian Province of China. Firstly, we did stepwise regression to the overall sample and established a model of production function about innovation. Secondly, we divided the sample into 3 clusters through k-mean cluster, and established the model of production function for each cluster. Thirdly, we compared the 4 models and try to causation of differences of them. The main conclusion includes three aspects, one is that the innovative enterprise(pilot) in Fujian province of China are all increasing returns to scale, two is that we found differences between innovative enterprises (pilot) in the 3 clusters, three is we proved the necessity of differential policy for the local government in order to further promote the regional innovation.
Key words: Regional innovation; Innovative enterprises; Stepwise regression; K-mean cluster; Differential policy

1 Introduction

The position of a region in national and global industry chain and regional labor division mainly depends on the regional innovative ability. As subject of technology innovation and market economy, Enterprises play great role in regional innovation and regional economic development, through the system evolution mode of product-Industry-region(Mou 2008)[1].

Large number of scholars has involved the research of innovative enterprise. For the identify features and certification standards of innovative enterprise, scholars and some agencies revealed from different angles. First, the innovation is the core impetus to the enterprises development. (Fontes, Coombs 2001)[2] suggested that the purpose of the establishment of these enterprises is to develop a new invention or conduct technological innovation, which generally have a greater impact to the development of existing enterprises, and promote technological advances throughout the industry. Second, enterprises stand on the strong foundation of technical ability. O' Regana and Simsb (O' Regana, Simsb 2008) [3] showed that the dense of high skilled or higher educated employees, the more fund will be inputted by enterprise to increase the human capital reserve. Thirdly, the output indicators of enterprise's patents and new products are more prominent. Lumpkin(Lumpkin 1996)[4], Hicks (Hicks 2005) [5] found the characteristics of innovation enterprises are that their output indicators of technological innovation activities such as R&D outcome of new product and patents are superior.

2 Source of Data and Method of Analysis

2.1 Source of data

The cross-section data we used is submitted by the innovative enterprises (pilot) in Fujian Province of China. For the need of statistic, we chose 113 enterprises whose data are complete.

2.2 Selection of dependent variable and independent variable

y: Sales revenue of new products, which measure enterprise innovation performance from financial perspective.

x1: Input of research and experiment, which give explain to input of production R&D.

x2: Input in education of staff, reflected the degree of attention paid to staff education by enterprise.

x3: Input in cooperation with scientific research institutions.

2.3 Stepwise regression

Stepwise regression method can choose the most appropriate independent variables to provide a full explanation to the dependent variable. This method is able to ensure the explanation of the model ,in the simplified to get an ideal model(Kramer and Gedeck 2011)[6].

For multiple linear regression model:

$$\mathbf{Y} = \mathbf{1}_n \mathbf{b}_0 + \mathbf{X}\mathbf{b} + \boldsymbol{\varepsilon} \quad (1)$$

Named **A** as a subset of{ 1， 2， …,p}, p is the number of independent variable, A represent a selection of the independent variables set:

$$y = b_0 + \sum_{i \in \mathbf{A}} b_i x_i + \varepsilon \tag{2}$$

We call it as the model to be select. Selection of the independent variable is not to find a true model, but to find a optimal one from all the models to be select. Common standards include MPESS, minimum Cp statistics, AIC criterion, BIC criterion, etc.

2.3 k-mean cluster method

Clustering analysis is a new branch in multiple statistics, and this theory has not been completed. But for the effect of this theory in practice, it is high valued by scholars. Cluster analysis is to establish a classification method, and then classify the sample, according to their degree of similar in nature.

K-mean cluster is a kind of cluster analysis method. Start from the overall samples, then, according to some optimal principle, stop at getting of the required K clusters.

We use k-mean cluster for clustering analysis, in 5 step (N.Bolloju 2001)[7]:

Step 1) Process the raw data.

Step 2) Find rallying point for each of the k cluster.

Step 3) Calculate distance of every sample to the k rallying point in ordered, and then will all the samples into one rallying point according to a nearest criterion, then get a classification results.

Step 4) Recalculate center of the cluster.

Step 5) In order to get the optimal classify, using Calinski and harabaszpseudo-F as the rule of stop.

2.4 Software for analysis

Use Stata measurement analysis software to realize the stepwise regression and the k-mean cluster.

3 Process and Result of Analysis

3.1 Collections and transaction of data

The cross-section data we used is submitted by the innovative enterprises (pilot) in Fujian Province of China. For the need of statistic, we chose 113 enterprises whose data are complete.

In processing of the raw data, due to the input-output model of the innovative enterprises (pilot) in Fujian Province of China are similar to the thought of Cobb-Douglas production function, so we take logarithm to the raw data.

3.2 Description and regression to the overall sample

To totally describe the overall sample, we make a scattergram matrix:

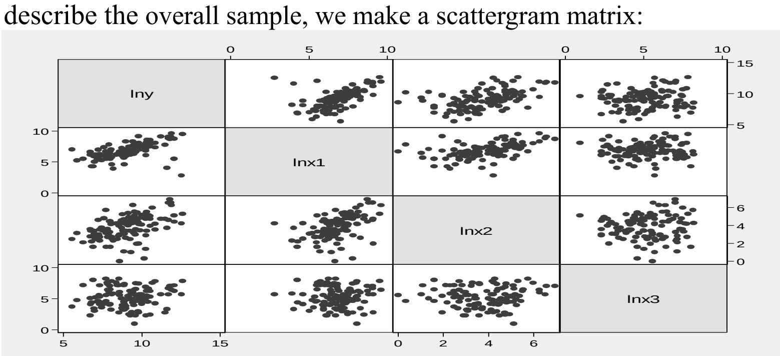

Figure 1 Scattergram Matrix of ln*y*, ln*x1*, ln*x2*, ln*x3*

In the regression process, three independent variables are all into the model. The result of regression listed in table1:

Table 1 Regression Result

Source	SS	df	MS
Model	9428.04193	3	3142.68064
Residual	210.863362	110	1.91693966
Total	9638.90529	113	85.3000468

Number of obs	= 113
F(3, 110)	= 1639.43
Prob > F	= 0.0000
R-squared	= 0.9781
Adj R-squared	= 0.9775
Root MSE	= 1.3845

lny	Coef.	Std. Err.	t	P>|t|	[95% Conf.	Interval]
lnx1	.9247485	.082311	11.23	0.000	.7616274	1.08787
lnx2	.3539772	.1130258	3.13	0.002	.1299866	.5779679
lnx3	.2668396	.0715588	3.73	0.000	.1250268	.4086524

The p-value of coefficient of every variable is less than 0.005, and the R-squared is 0.978. Regression equation is good, so the regression model for overall sample is:

$$\ln y = 0.925 \ln x1 + 0.354 \ln x2 + 0.267 \ln x3 \qquad (3)$$

3.3 K-mean cluster to the overall sample

We use the last k obs as rallying point for each of the k cluster, and use Calinski and harabaszpseudo-F as the rule of stop. The Calinski and harabaszpseudo-F value listed in table2:

Table 2 Calinski and Harabaszpseudo-F Value and Number of Cluster

Number of cluster	2	3	4	5
Calinski and harabaszpseudo-F	**45.80**	**49.27**	**42.33**	**42.72**

Calinski and harabaszpseudo-F get maximum when the number of cluster is 3, so the best number of cluster is 3. Cluster 1 includes 34 enterprises, and cluster 2 includes 40 enterprises, and cluster 3 includes 39 enterprises.

3.4 Regression to 3 clusters

3.4.1 Regression to cluster1

In the regression process, ln*x1* and ln*x3* are into the model. Result of regression listed in table3:

Table 3 Regression Result of Cluster 1

Source	SS	df	MS		
				Number of obs	= 34
				F(2, 32)	= 1203.73
Model	2148.65581	2	1074.32791	Prob > F	= 0.0000
Residual	28.559926	32	.892497688	R-squared	= 0.9869
				Adj R-squared	= 0.9861
Total	2177.21574	34	64.035757	Root MSE	= .94472

lny	Coef.	Std. Err.	t	P>|t|	[95% Conf.	Interval]
lnx1	.8812654	.1340687	6.57	0.000	.6081764	1.154354
lnx3	.4139384	.125014	3.31	0.002	.1592931	.6685837

The p-value of coefficient of every variable is less than 0.005, and the R-squared is 0.987. Regression equation is good, so the regression model for cluster 1 is:

$$\ln y = 0.881 \ln x1 + 0.414 \ln x3 \qquad (4)$$

3.4.2 Regression to cluster 2

In the regression process, ln*x1* and ln*x2* are into the model. Result of regression listed in table4:

Table 4 Regression Result of Cluster 2

Source	SS	df	MS		
				Number of obs	= 40
				F(2, 38)	= 726.86
Model	2959.25245	2	1479.62622	Prob > F	= 0.0000
Residual	77.3539967	38	2.03563149	R-squared	= 0.9745
				Adj R-squared	= 0.9732
Total	3036.60644	40	75.9151611	Root MSE	= 1.4268

lny	Coef.	Std. Err.	t	P>|t|	[95% Conf.	Interval]
lnx1	.8213159	.1392073	5.90	0.000	.5395055	1.103126
lnx2	.8423963	.2373137	3.55	0.001	.3619797	1.322813

The p-value of coefficient of every variable is less than 0.005, and the R-squared is 0.987. Regression equation is good, so the regression model for cluster 2 is:

$$\ln y = 0.821 \ln x1 + 0.842 \ln x2 \qquad (5)$$

3.4.3 Regression to cluster 3

In the regression process, ln*x1* and ln*x3* are into the model. Result of regression listed in table5:

Table 5 Regression Result of Cluster 3

Source	SS	df	MS		
				Number of obs	= 39
				F(2, 37)	= 766.63
Model	4320.81298	2	2160.40649	Prob > F	= 0.0000
Residual	104.26844	37	2.81806594	R-squared	= 0.9764
				Adj R-squared	= 0.9752
Total	4425.08142	39	113.463626	Root MSE	= 1.6787

lny	Coef.	Std. Err.	t	P>|t|	[95% Conf.	Interval]
lnx1	1.054349	.1401729	7.52	0.000	.770332	1.338367
lnx2	.4604257	.2223514	2.07	0.045	.0098991	.9109524

The p-value of coefficient of every variable is less than 0.05, and the R-squared is 0.976. Regression equation is good, so the regression model for cluster 3 is:

$$\ln y = 1.054 \ln x1 + 0.460 \ln x2 \qquad (6)$$

4 Conclusion and Suggestion

Slope coefficient of regression model measure in flexibility of input to output. The flexibility of ln$x1$, ln$x2$, ln$x3$ are 0.925，0.354，0.267 respectively. In addition, the return to scale parameter of overall sample is 1.546, present increasing returns to scale.

Almost all enterprises in cluster 1 are small-scale or new established, and include no state-owned enterprise. And in the regression process, ln$x1$ and ln$x3$ are into the model. The flexibility of ln$x1$, ln$x3$ are 0.881，0.414 respectively. But ln$x2$ has not been including in the independent variable set, and the return to scale parameter of cluster 1 is 1.546, lower than the overall sample. The model imply that enterprises in cluster 1 are lower in the level of innovation, and neglect the education of their staff, while have more cooperation with scientific research institutions.

Enterprises in cluster 2 and cluster 3 have similar scale. Almost all the state-owned enterprises are in these two clusters and lnx3 are out of their independent variable set. In the model of cluster 2 the flexibility of ln$x1$, ln$x2$ are 0.821, 0.842 respectively, while in the model of cluster 3 the flexibility of ln$x1$, ln$x2$ are 1.054，0.460 respectively, and the return to scale parameter is 1.546 lower than cluster 2. The models of cluster 2 and cluster 3 imply that enterprises in these two cluster are higher than cluster 1 in the level of innovation, and they all neglect the cooperation with scientific research institutions, while enterprises in cluster 2 pay more attention to the education of their staff, and their level of innovation is higher than cluster 3.

On the whole, innovative enterprises (pilot) in Fujian Province of China all present increasing returns to scale, but enterprises in different clusters have different tend in their innovation, so, in order to further promote the regional innovation, it is necessary for the local government to execute differential policy.

References

[1] Mou R. Research on the Path of Innovation: Product-industry-regional[D]. 2008, Wuhan University of Technology: Wuhan (In Chinese)

[2] Fontes M., R. Coombs. Contribution of New Technology-based Firms to the Strengthening of Technological Capabilities in Intermediate Economies [J]. Research Policy, 2001, 30(2001): 79-97

[3] O'Regana N., M.A. Simsb. Identifying High Technology Small Firms: A Sectoral Analysis [J]. Technovation, 2008, 28(2008): 408-423

[4] Lumpkin G. Clarifying the Entrepreneurial Orientation Construct and Linking It of Performance. Academy of Management Review [M], 1996: 135-172

[5] Hicks D., D. Hegde. Highly Innovative Small Firms in the Markets for Technology [J]. Research Policy, 2005, 34(5): 703-716

[6] Kramer C., P. Gedeck. Three Descriptor Model Sets a High Standard for the CSAR-NRC HiQ Benchmark[J]. Journal of Chemical Information and Modeling, 2011, 2011(5.30): 35-40

[7] N. Bolloju. Aggregation of Analytic Hierarchy Process Models based on Similarities in Decision Makers' Preferences[J]. European Journal of Operational Research, 2001, 128(2001): 499-508