Optimality of the C-D Knowledge Production Function Under the Condition of Lognormal Distributions

Cheng Bangwen[1], Liu Shumei[2], Shi Linfen[1], Wang Yali[1], Yang Hongjin[1], He Rong[3], Yang Jun[1] and Xu Ji[1]

[1] *stsic@public.wh.hb.cn*
Huazhong University of Science and Technology, School of Management (China)

[2] *liusm@most.cn*
Department of Development Planning, Ministry of Science and Technology (China)

[3] *Herong128@163.com*
Wuhan University of Science and Engineering, College of Economics and Management (China)

Introduction

The Knowledge production function (KPF) with the C-D form (Griliches, 1979 and Jaffe, 1989) has got wide spread in analyzing knowledge productions and innovations. The C-D KPF is widely used mainly because of its excellent properties. Besides, there must be other reasons. Based on the optimal estimation theory this paper explores the model form of the KPF and proves that under the condition of lognormal distributions of knowledge variables the C-D KFP is the only one that can make the mean square error minimum, while any other forms of the KFP, for example the CES or Translog function, has not this important property.

Lognormal Distribution of scale variables

Usually the activity of knowledge productions and innovations is measured by variables describing the scale of input resources or outputs, such as the numbers of R&D personnel, R&D expenditure, S&T papers and patents etc., which we call scale variables. Cheng (2000, 2003) indicated that social-economic scale variables approximately obey joint lognormal distributions. The input and output of knowledge productions and innovations are scale variables, then their joint distributions over the population composed of knowledge production units are multidimensional lognormal.

Now, we take the three-dimensional distributions of the R&D expenditure, R&D personnel and patent application over 31 provinces in China from 2004 to 2008 as an example to show the validity of the assumption above. The original data are from “China Statistical Yearbook on science and technology 2005-2009”. Whether a distribution is lognormal can be tested with the statistical test of K-S method. For multi-dimensional distributions of random variables, we first need to conduct the logarithmic transformation to each variable, and then do the principal factor analysis decomposing the variables to mutually orthogonal one-dimensional factors, finally test each factor with K-S. That is, testing multidimensional variables is equivalent to test each one-dimensional factor. All calculations above can be done by SPSS software. Table 1 lists the K-S test results of three-dimensional variables by region. The results indicate the significance levels of three factors are all greater than 0.05, therefore the regional distributions of three factors in each year are lognormal, and we can conclude that the regional distribution of R&D expenditure, R&D personnel and patent application can be described by three-dimensional lognormal.

Table 1 K-S test result of 3-dimensional Lognormal distribution

year	significance level α		
	Factor 1	Factor 2	Factor 3
2004	0.729	0.485	0.775
2005	0.741	0.990	0.974
2006	0.856	0.997	0.952
2007	0.887	0.984	0.997
2008	0.899	0.874	0.840

If other scale variables of knowledge productions and innovations are taken into account, their joint distributions will almost be multidimensional lognormal. The empirical analysis indicates that the assumption above is reasonable.

Knowledge Production Function under optimal estimation

In economics, the general form of the production function is

$$y = F(x_1, x_2, \cdots, x_m) \quad (1)$$

For knowledge productions and innovations, the function (1), which regards R&D expenditures and personals as main inputs, and papers, patents as outputs, is named as a KPF. The model form of the KPF is unknown. The goal of this paper is to determine its form, which is not from theoretical deductions, but rather from empirical analyses with the optimal estimation theory based on the output and inputs $y, x_1, x_2, \cdots, x_m$. This paper is based on the assumption: $y, x_1, x_2, \cdots, x_m$ are scale random variables with a joint lognormal distribution.

Since $x_i > 0$, the function (1) is equivalent to

$$\ln y = \ln F(\ln x_1, \ln x_2, \cdots, \ln x_m) \quad (2)$$

We define the mean square error (MSE) e

$$e = E\{[\ln y - \ln F(\ln x_1, \cdots, \ln x_m)]^2\} \quad (3)$$

Obviously, the less is the MSE, the more F can reflect the reality of inputs and output of the knowledge production and innovation. Our problem now is to find the function F such that the MSE is minimum. According to the principle of the mean square estimation (Papoulis, 1991, pp203), $\ln F$ can only be

$$\ln F(\ln x_1, \cdots, \ln x_m) = E\{\ln y | \ln x_1, \cdots, \ln x_m\}. \quad (4)$$

According to the assumption

$$(\ln y, \ln x_1, \cdots, \ln x_m)^T \sim N(\mu, v) \quad (5)$$

$$\mu = E(\ln y, \ln x_1, \cdots, \ln x_m)^T \quad (6)$$

$$v = D(\ln y, \ln x_1, \cdots, \ln x_m)^T. \quad (7)$$

where $N(\cdot)$ denotes normal distribution, μ and v are mean vector and covariance matrix respectively.

The μ and v can be subdivided as

$$\mu = \begin{array}{c} \begin{pmatrix} \mu_1 \\ \mu_2 \end{pmatrix} \end{array} \begin{array}{c} 1 \\ m \end{array} \quad (8)$$

$$v = \begin{array}{cc} 1 & m \\ \multicolumn{2}{c}{\begin{pmatrix} v_{11} & v_{12} \\ v_{21} & v_{22} \end{pmatrix}} \end{array} \begin{array}{c} 1 \\ m \end{array} \quad (9)$$

According to the properties of joint normal distributions (Mardia, 1979) we have

$$E\{\ln y | \ln x_1, \cdots, \ln x_m\} = \mu_1 + v_{12} v_{22}^{-1}[(\ln x_1, \cdots, \ln x_m)^T - \mu_2] = b + \sum_{i=1}^{m} a_i \ln x_i \quad (10)$$

where a_i, b are parameters. From (1), (4) and (10), we obtain

$$y = A x_1^{a_1} x_2^{a_2} \cdots x_m^{a_m} \quad (11)$$

where A is a parameter, α_i is output elasticity of input factor x_i.

The analysis above indicates that if the lognormal distribution assumption of knowledge variables is satisfied, C-D function (11) is the only KPF which minimizes the MSE and can appropriately reflect the quantitative relationship between output and inputs.

Discussion

1. In the C-D KPF (11) the inputs and output are random variables over the population composed of knowledge production units $\omega_1, \omega_2, \cdots \omega_n$, therefore the function (11) can be rewritten as:

$$y(\omega_i) = A x_1^{a_1}(\omega_i) \cdots x_m^{a_m}(\omega_i)$$
$$i = 1, \cdots, n \quad (12)$$

When estimating parameters A and $a_1,\cdots,a_m$ with regression, we should use input and output data of $\omega_1,\omega_2,\cdots\omega_n$ composing the population.

2. The paper is based on the assumptions which can be satisfied in most cases. But it cannot be guaranteed especially in the case of small samples, it is better to test whether the distribution is multidimensional lognormal.

3. Houthakker (1955) analyzed the relation between the aggregate production function F and the production possibilities of the individual cells, and derived the C-D aggregate production function from a Pareto distribution. The function (11) only reflects the relation between input and output of the individual cells, which is different with the result from Houthakker's paper.

4. This paper analyzed the optimality of C-D KPF only from the mathematic perspective. Of course it is not enough. We should further investigate and present the proof of the fact that for lognormally distributed random variables compared with other KFP, the C-D KPF indeed minimizes the MSE.

References

Cheng Bangwen et al(2000), Sci-tech Scale Index Conforming to the Law of Logarithmic Normal Distribution, Science of Science and Management of S&T, Vol.228, No. 9, p9~11.

Cheng Bangwen et al(2003), The Model and Method for Checking Quality of Multidimensional Statistics and Identifying Outliers from the Data, Mathematics in Practice and Theory, Vol.33, No. 4, pp4~7.

Griliches, Z. (1979), Issues in Assessing the Contribution of R&D to productivity Growth, Bell Journal of Economics, Vol.10, pp92-116.

Houthakker, H. S. (1955), The Pareto Distribution and the Cobb-Douglas Production Function in Activity Analysis, The Review of Economic Studies, Vol.23, No. 1, pp27-31.

Jaffe, A. B. (1989), Real Affects of Academic research, American economics review, Vol.79, pp957-970.

Mardia, K. V, et al (1979), Multivariate Analysis, Academic Press Inc, New York.

Papoulis, A. (1991), Probability, Random Variables, and Stochastic Processes, McGrsw-Hill, New York.