MAXIMUM PERMISSIBLE ESTIMATES IN THE PROBLEM OF PARAMETRIC IDENTIFICATION OF THE EXTENDED COBB-DOUGLAS FUNCTION

[a]OLGA KANTOR

[a]Department of Corporate Finance and Accounting Technologies, Ufa State Petroleum Technological University, 450062, 1 Kosmonavtov Str., Ufa, Russia

email: [a]o_kantor@mail.ru

Abstract: Cobb-Douglas production functions are one of the most popular instruments for analyzing the relationship between the factors of production-economic activity. Their construction is traditionally performed using mathematical-statistical methods. To a great extent, the problems of identifying models of this type are related to a priori uncertainty of the available data, in particular, their inaccuracy and insufficient depth of presentation. The present paper describes the developed method for the parametric identification of the extended Cobb-Douglas production function, which neutralizes these types of uncertainty and is based on using Kantorovich's idea for calculating interval estimates of the sought-for parameters.

Keywords: : extended Cobb-Douglas function, parametric identification, maximum permissible estimates of parameters, uncertainty region.

1 Introduction

Cobb-Douglas production functions are actively used in research to establish a relationship between the resulting indicators of production activity (Y) and the exogenous factors influencing them (*K* is capital and *L* is labor): (1-2)

$$Y = AK^{\alpha}L^{1-\alpha}. \quad (1)$$

Due to the transparency of interpretation of its parameters, the classical Cobb-Douglas function (1) is widely used in research practice. (3-4) At the same time, there is also a great desire to increase the application scope of this tool of applied analysis by including more factors (3,5) in the model (1). However, for a number of objective reasons, extended Cobb-Douglas functions today remain not a real instrument for studying the relations between production factors, but rather a desirable one.

The main factors restraining the use of extended Cobb-Douglas functions for economic analysis and substantiation of managerial decisions are:

- the scantiness of the available statistical information;
- computational difficulties that arise in the application of mathematical-statistical methods, the main tool for determining estimates of the desired parameters.

In particular, it is well known that, for each exogenous variable, 6–7 data are required to obtain statistically significant estimates of parameters. Therefore, a priori, "short" series of data do not allow obtaining reliable relations, thus limiting the possibility of using mathematical-statistical methods. Moreover, even in the presence of the "long" series of initial data, the use of mathematical-statistical methods does not guarantee to obtain adequate models.

However, there is another objective problem that researchers prefer not to touch upon in practical research, and which, in the author's opinion, deserves serious attention. This problem concerns a priori inaccuracy of the initial data. The data of any observations (both active and passive) contains some errors. On the other hand, statistical data on socio-economic indicators are based on multi-stage procedures for collecting and processing information, as a result of which errors accumulate and can result in significant distortions of the real picture. It is not possible to assess the accuracy of the data obtained due to the lack of reference values and the impossibility of multiple observations under the same conditions.

All of the above necessitates the application of special approaches to the construction of the extended Cobb-Douglas function. Of particular interest in this regard, in the authors' opinion, is the approach, the founder of which is the outstanding scientist Kantorovich, who for the first time formulated in his work (6) the idea of obtaining exact two-sided boundaries for the parameters of models and regions for the sought-for and observed quantities. A distinctive feature of this approach is that it was a new word in the theory of mathematical processing of experimental data, because it does not require knowledge of the statistical properties of the distribution of measurement errors and allows the determination of the desired parameters considering the requirements that are of interest from the point of view of the researcher.

The ideas, expressed by Kantorovich, laid the foundation for interval analysis, which is being actively developed due to the work of foreign and Russian authors. (7-11) They are also successfully used in certain scientific areas, (12-13) in particular, in solving problems of chemical kinetics. (14-15) In the present work, developing the ideas for obtaining precise two-sided boundaries for model parameters under conditions of the initial data uncertainty, the authors present the method of parametric identification of the extended Cobb-Douglas function based on the use of maximum permissible parameter estimates.

2 Materials and Methods

The extended Cobb-Douglas function of the following form will be considered:

$$Y = \alpha_0 \cdot X_1^{\alpha_1} \cdot X_2^{\alpha_2} \cdot \ldots \cdot X_n^{\alpha_n} \quad (2)$$

$$\alpha_j \geq 0, \; j = \overline{0,n},$$

where X_1, …, X_n are exogenous variables of the model, *Y* is the endogenous variable, α_j, $j = \overline{1,n}$ are the parameters that are the elasticity coefficients for the corresponding exogenous factors, α_0 is a parameter characterizing the scale of the economy as a whole.

The initial information is represented by the sets of values:

$$\{X_{1t},\ldots,X_{nt},Y_t\}, \; t = \overline{1,m}, \quad (3)$$

which cannot be considered as absolutely precise. Therefore, it can be assumed that the true values of the initial data belong to some, not always a priori known, intervals:

$$X_{jt} \in \left[\underline{X}_{jt}, \overline{X}_{jt}\right], \; Y_t \in \left[\underline{Y}_t, \overline{Y}_t\right], \; t = \overline{1,m}, \; j = \overline{1,n}. \quad (4)$$

Considering (4), the determination of a single set of parameter values $\overline{\alpha} = \{\alpha_0, \alpha_1, \ldots, \alpha_n\}$ can lead to the fact that an exact solution will be obtained based on a priori inaccurate data. For this reason, it may be more appropriate to search for a region Λ^*, consisting of sets of parameters $\overline{\alpha} = \{\alpha_0, \alpha_1, \ldots, \alpha_n\}$ that ensure acceptable agreement between the experimental and calculated values of the variable *Y* from the positions of the introduced optimality criterion. The region Λ^* will be called the region of uncertainty, implying in this case that the term "uncertainty" reflects the fact according to which each point of the region Λ^* can be selected to specify the final form of the model (2).

The region Λ^* may have a complex structure. Therefore, instead of establishing the exact boundaries of this region, the traditional approach can be used based on the consideration of a set approximating the initial region. To do this, for each of the parameters α_j, a segment is defined:

$$\alpha_j \in \left[\underline{\alpha}_j, \overline{\alpha}_j\right], \; j = \overline{0,n}, \quad (5)$$

consisting of the values α_j, for each of which there exist some values of other parameters that form the set $\overline{\alpha} = \{\alpha_0, \alpha_1, \ldots, \alpha_n\}$: $\overline{\alpha} \in \Lambda^*$, whereas for $\alpha_j \notin [\underline{\alpha}_j, \overline{\alpha}_j]$ $\overline{\exists}\, \overline{\alpha} \in \Lambda^*$.

This interval will be called the uncertainty interval, whereas its boundaries, the maximum permissible estimates of the parameters of the identified dependence (2).

Let us introduce the uncertainty set Λ, given by the direct product of intervals (5):

$$\Lambda = [\underline{\alpha}_0, \overline{\alpha}_0] \times \ldots \times [\underline{\alpha}_n, \overline{\alpha}_n] \qquad (6)$$

It is obvious that $\Lambda^* \subset \Lambda$, by virtue of which Λ is a set approximating the region Λ^*. The set (5) has a simpler structure compared to the region Λ^*. In this case, the estimate (7) is valid, with the help of which conclusions can be drawn about the degree of uncertainty in solving the parametric identification problem for the model (2)

$$diam\, \Lambda^* = diam\, \Lambda = \max_{j=0,n} \left(\overline{\alpha}_j - \underline{\alpha}_j\right) \qquad (7)$$

3 Results and Discussion

The construction of the extended Cobb-Douglas function was carried out in the context of establishing a connection between the gross domestic product (GDP) of the Russian Federation (RF) and the factors characterizing the cost of labor and the composition of labor resources (Table 1).

Tab. 1: Variables of the Model (2)

Notation of the variable	Name of the indicator
Y	GDP (in prices of 2000), billion rubles
X_1	The value of fixed assets (in prices of 2000), billion rubles
	The average annual number of people employed in the economy, thousand people
X_2	▪ with higher education;
X_3	▪ with secondary vocational education;
X_4	▪ with complete secondary education;
X_5	▪ with basic general education;
X_6	▪ the remaining employed people (with other categories of education and not having basic general education)

The values of the variables Y and X_1 were reduced to a comparable form using the deflator indices of the physical volume of GDP and producer prices in the construction industry, respectively (Figure 1). (16)

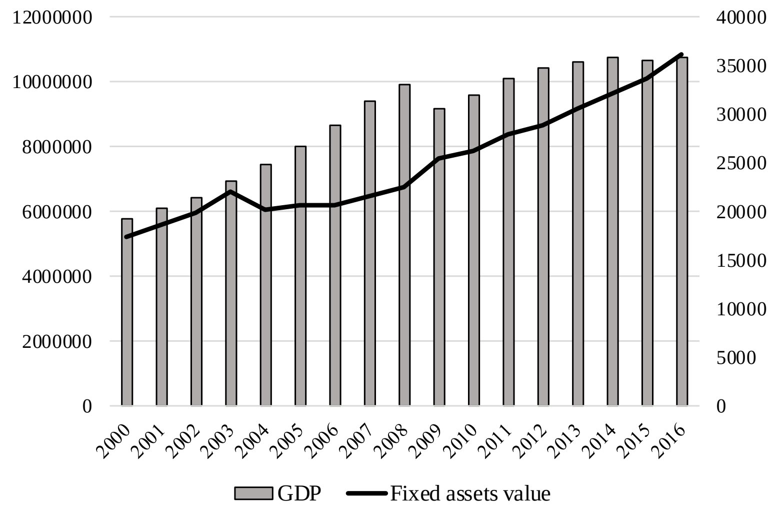


Figure 1: Dynamics of the Cost Indicators of the Model (2)

Data on the average annual number of people employed in the region’s economy (Figure 2) were formed based on official statistical sources. (17)

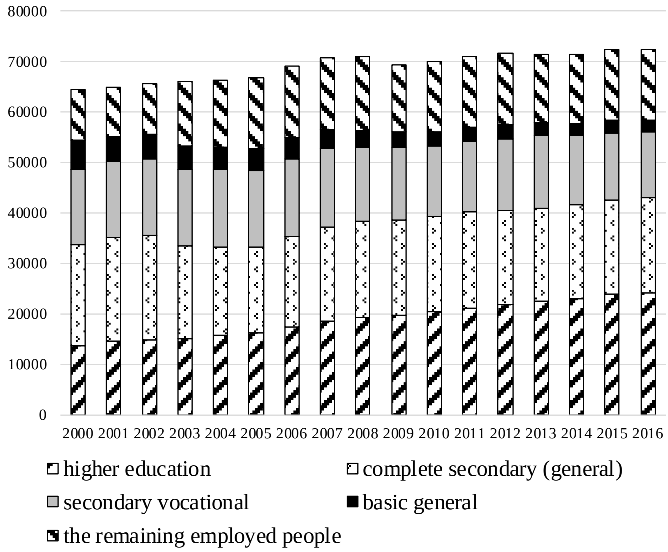


Figure 2: Dynamics of the Average Annual Number of People Employed in the Russian Economy

The parameters of the extended Cobb-Douglas function (2) were determined using the following three-stage procedure.

On the *first stage*, the linearization of the model was carried out:

$$lnY = ln\alpha_0 + \alpha_1 ln X_1 + \ldots + \alpha_6 ln X_6 \quad (8)$$

$\alpha_j \geq 0$, $j = \overline{0,6}$.

On the *second stage*, the maximum permissible approximation error ξ^* was calculated − the indicator of the accuracy of the correspondence of the logarithms of the actual (Y_t) and the calculated ($\hat{Y}_t$) values of the endogenous variable *Y*. The implementation of this stage was carried out based on solving the problem:

$$\xi \to min$$

$$\left| lnY_t - ln\hat{Y}_t \right| \leq \xi, \quad t = \overline{1,17} \quad (9)$$

$\alpha_j \geq 0$, $j = \overline{0,6}$.

Additionally, the accuracy of the model was estimated based on the average approximation error

$$\overline{A} = \frac{1}{17}\sum_{t=1}^{17}\left|\frac{Y_t - \hat{Y}_t}{Y_t}\right| \cdot 100\%.$$

The result of the numerical implementation of the model (9) is the maximum permissible approximation error $\xi^* = 0.029$ and the point estimates of the parameters α_j^*, $j = \overline{0,6}$, which allowed writing the model (2) in the form:

$$Y = 1.0 \cdot X_2^{0.820} \cdot X_3^{0.114} \cdot X_4^{0.111} \cdot X_6^{0.604}. \quad (10)$$

The average approximation error for model (10) was $\overline{A} = 1.74\%$.

In the *third stage*, the maximum permissible estimates were calculated for each of the parameters α_j, $j = \overline{0,n}$ using the models:

$$\alpha_j \to min(max), \quad j = \overline{0,6}$$

$$\left| lnY_t - ln\hat{Y}_t \right| \leq \xi^*, \quad t = \overline{1,17}, \quad (11)$$

$\alpha_j \geq 0$, $j = \overline{0,6}$.

As a result, it was found that all maximum permissible estimates coincided with the corresponding point values of the parameters α_j^*: $\alpha_j^* = \underline{\alpha}_j = \overline{\alpha}_j$, $j = \overline{0,6}$.

Following the assumption of a priori inaccuracy of the initial data, let us pose the problem of determining the value intervals of the parameters $\alpha_j \in \left[\underline{\alpha}_j^{\delta}, \overline{\alpha}_j^{\delta}\right]$, $j = \overline{0,6}$, which ensure the closeness of the logarithms of the actual and calculated values of the variable *Y* at a level not exceeding $\xi^*(1+\delta)$. In essence, this means that a quantitative estimate of the degree of sensitivity of the set of values of the model parameters (10) to the change in the maximum permissible approximation error (i.e., to the accuracy variation) by 100δ percent ($\delta \geq 0$) will be obtained. In this case, additionally, the condition for the maximum discrepancy between the actual and calculated values of the variable *Y* will be set at the level of 5%.

For these purposes, it is necessary to solve 14 problems of the following type:

$$\alpha_j \to min(max), \quad j = \overline{0,6}$$

$$\left| lnY_t - ln\hat{Y}_t \right| \leq \xi^*(1+\delta), \quad t = \overline{1,17} \quad (12)$$

$$\frac{1}{17}\sum_{t=1}^{17}\left|\frac{Y_t - \hat{Y}_t}{Y_t}\right| \cdot 100 \leq 5$$

$\alpha_j \geq 0$, $j = \overline{0,6}$.

The solution of the problems (12) for $\delta = \mathbf{0.02}$ showed (Table 2) that the reaction to the expected changes in accuracy was observed to a greater extent for the parameter α_0. The smallness of the ranges of variation of the parameters α_1 and α_5 is an additional confirmation of the fact that it is not advisable to include the corresponding variables to the Cobb-Douglas production function.

Tab. 2: Investigation of the Sensitivity of the Parameters of the Model (10) to Changes in the Maximum Permissible Approximation Error for $\delta = \mathbf{0.02}$

Parameter	α_0	α_1	α_2	α_3	α_4	α_5	α_6
$\left[\underline{\alpha}_j^{\delta}, \overline{\alpha}_j^{\delta}\right]$	[1, 1.376]	[0, 0.009]	[0.804, 0.835]	[0.090, 0.150]	[0.089, 0.134]	[0, 0.013]	[0.573, 0.613]

The results obtained allow drawing the following conclusions regarding the expanded Cobb-Douglas production function (10):

- the model (10) is characterized by increasing yield from scale ($\sum_{j=1}^{6} \alpha_j = 1.65$);
- the factors X_2 and X_6 have the greatest impact on GDP; the factors X_3 and X_4, to a much lesser extent; the factors X_1 and X_5, very insignificant impact, as a result of which they are recommended to be excluded from consideration;
- the level of technological productivity (parameter α_0) increases the yield from the exogenous factors of the production function.

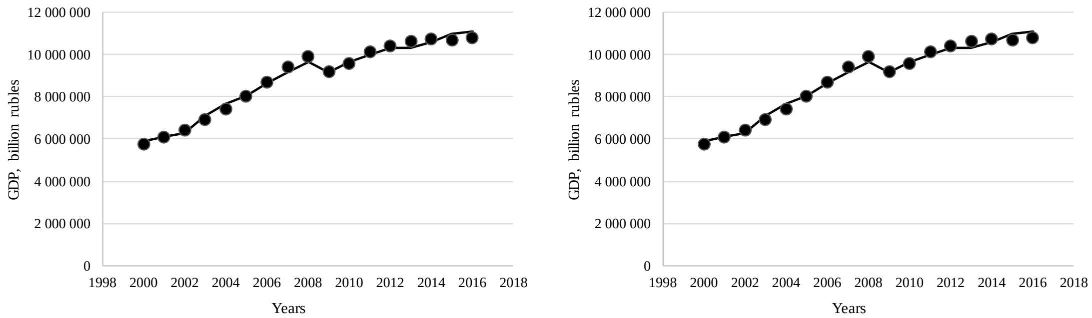
In order to carry out a comparative analysis of the possibilities of the presented method of parametric identification of the Cobb-Douglas production function and the classical mathematical-statistical approach using the same initial data using the least-squares method, the authors estimated the parameters of the regression dependence of the form

$$Y = \alpha_0 \cdot X_2^{\alpha_2} \cdot X_3^{\alpha_3} \cdot X_5^{\alpha_5} \cdot X_6^{\alpha_6}.$$

According to the calculation results, the extended Cobb-Douglas function assumed the following form:

$$Y = 0.0000896 \cdot X_2^{0.978} \cdot X_3^{0.378} \cdot X_5^{0.652} \cdot X_6^{0.602}. \quad (13)$$

The average approximation error for the model (13) was 1.02%. A comparison of the models (10) and (13) showed that each of them describes with high accuracy the behavior of the endogenous variable (Figure 3). However, the smallness of the parameter α_0, characterizing the level of technological productivity in the Russian Federation, in the model (13) does not allow considering it adequate even considering the confidence interval for its values ($\alpha_0 \in [0.00000078, 0.0105]$).



a) model (10) b) model (13)

Figure 3: Graphical Interpretation of the Constructed Cobb-Douglas Functions

4 Conclusion

The paper presents a method for determining the parameters of the extended Cobb-Douglas function. Its distinctive feature is obtaining the interval estimates of values for each parameter being evaluated. For the presented approach, the necessary mathematical bases have been developed, according to which the problem of identifying the extended Cobb-Douglas function is reduced to the successive solving two types of linear programming problems: (9) and (11). Based on the results of the numerical solving of the problem (9), the value of the best approximation of the available data is established – the maximum absolute approximation error; according to the results of solving the problems (11), the maximum permissible estimates of the model parameters are calculated, which specify the ranges of variation of values for each of the parameters and allow estimating the degree of uncertainty of the obtained solution. Also, the procedure is formalized for studying the sensitivity of the obtained solution in the problem of parametric identification for the expected change in the accuracy of the model.

An important specific feature of the presented method is the ability to conduct research for the data that are characterized by the absence of a large number of observations.

The possibilities and advantages of the presented method are demonstrated by the example of a study of the dependence of the GDP of the Russian Federation on the value of fixed assets and the average annual number of people employed in the economy in terms of their education levels based on the data for the years 2000-2016.

Literature:

1. Cobb C, Douglas P. Theory of production. American Economic Review, Supplement. 1928; 18:139-65.
2. Chechulin VL. On the place of the Cobb-Douglas model in the hierarchy of models. Bulletin of Perm University. Series: Mathematics. Mechanics. Informatics. 2013; 1(13):46-9.
3. Koritsky AV. (2011). Dynamics of private and social norms of education output in Russia. Issues of Innovative Economy. 2011; 1:11-29.
4. Gafarova EA. Simulating regional development on the basis of production functions. Internet Journal of Science Studies. 2013; 3(16):10.
5. Buravlev AI. Three-factor Cobb-Douglas production model. Economics and Management: Problems and Solutions. 2012; 3:13-9.
6. Kantorovich LV. On some new approaches to computational methods and processing of observations. Siberian Mathematical Journal. 1962; 5(3):701-9.

7. Moore RE. (1967). Interval analysis. Journal of the Franklin Institute. 1967; 2(284):148-9.
8. Alefeld G, Mayer G. Interval analysis: Theory and applications. Journal of Computational Applied Mathematics. 2000; 1-2(121):421-64.
9. Kalmykov SA, Shokin YuI, Yuldashev ZKh. Methods of interval analysis. Novosibirsk: Nauka; 1986.
10. Jaulin L, Kieffer M, Didrit O, Walter E. Applied interval analysis. Moscow-Izhevsk: Institute for Computer Research; 2005.
11. Sharyi SP. Finite-dimensional interval analysis. Novosibirsk: XYZ Publishing House; 2019.
12. Sukhanov VA. The study of empirical dependencies: a non-statistical approach: a collection of scientific articles. In N.A. Oskorbina, P.I. Kuzmina, Eds. Barnaul: Altay University; 2007: 115-27.
13. Spivak SI, Ismagilova AS, Kantor OG. Uncertainty regions in the mathematical theory of measurement analysis. Control Systems and Information Technology. 2014; 4(58):17-21.
14. Spivak SI, Timoshenko VI, Slinko MG. Methods for constructing kinetic models of stationary reactions. Chemical Industry Today. 1979; 3:33-6.
15. Mustafina SA, Larin OM, Mustafina SI. Study of accuracy in chemical kinetics problems. Entomology and Applied Science Letters. 2018; 1(5):59-63.
16. Rosstat. Official site of the Federal State Statistics Service; 2020. Available at: www.gks.ru.
17. Labor force, employment and unemployment in Russia (based on the results of sample labor force surveys). Rosstat. Statistical Book; 2018.

Primary Paper Section: B

Secondary Paper Section: BA, BB