©2019 International Transaction Journal of Engineering, Management, & Applied Sciences & Technologies

International Transaction Journal of Engineering, Management, & Applied Sciences & Technologies

http://TuEngr.com

TuEngr Group

PAPER ID: 10A19F

DATA PREPARATION FOR CREATION OF PRODUCTION FUNCTION OF CORPORATION IN THE PEOPLE'S REPUBLIC OF CHINA IN A VOLATILE ECONOMIC ENVIRONMENT

Akhmetova Irina A. [a*], **Shikhalev Anatoly M.** [a], **Demyanova Olga V.** [a], **Sirazieva Zarina N.** [a]

[a] *Institute of Management, Economics and Finance, Kazan Federal University, Kazan, RUSSIA*

ARTICLEINFO

Article history:
Received 08 April 2019
Received in revised form 31 May 2019
Accepted 07 June 2019
Available online 25 October 2019

Keywords:
Component Analysis; Statistical Population; Multiple Regression; Chi-Squared Test; Production Function.

ABSTRACT

In a volatile economic environment, a special calculation is necessary for developing mathematical models of the examined processes and the phenomena for which there was a typical use of a sample statistical population of small size. In similar circumstances the following tasks are objectively updated: 1) what means (methods) we should use and how it is necessary to operate with limited size of the examined statistical population; 2) in what degree results of the available homogeneous data of the examined sample population (SP) can be extended for their statistical and probabilistic generalization in the form of the parent population (PP). As the main mathematical model, we offer to apply the production functions (PF) model which various elements are investigated previously within the component analysis and the corresponding tests.

Disciplinary: Economic Sciences (International Economics).

1. INTRODUCTION

Nowadays modern development of economy and economic relations are volatile because national economies have tendencies to transfer a part of the capital in so-called multinational corporations. In this regard researchers start to use the term "economic turbulence" [1], in particular for the separate enterprises and corporations. Therefore, the developed models of enterprise management have to consider current trends of economic development in general. China National Petroleum Corporation (CNPC) from 2010 to 2018 is an object of the research.

2. METHODS

Tthe production function model connecting production of X (billion CNY) can be the most effective means of development of the chosen enterprise as an endogenous (independent) variable depending on two cause variables – volume of fixed assets (capital) K (billion CNY) and human resources (labor) of enterprise L (thousand people) that in general, it is possible to present as [3]

$$X = f(K, L) \qquad (1).$$

According to [3], function (1) in the required multiplicative version has to be:

$$X = A \cdot K^{\alpha 1} \cdot L^{\alpha 2} \qquad (2),$$

where A – index of technical progress; α1 and α2 – coefficient of elasticity.

As multiplicative PF of the example (2) is determined by time series of releases and expenses of resources (see Table 1), in relation to it the multiple regression model based on Least Square (LS) method can be used [3]. It is evident that the example (2) in logarithms on any basis (here – on the basis of exponents) will have a linear (additive) structure concerning indeterminate (ln A, α1 and α2) [3, p. 16]:

$$\ln X = \ln A + \alpha_1 \ln K + \alpha_2 \ln L \qquad (3),$$

which easily is solved by means of standard computer application programs on the input given in Table 1:

Table 1: Input for data modeling of CNPC [4].

i	Year	Revenue, X1, billion CNY	Profit, X2, billion CNY	Fixed Assets (FA), K, billion CNY	Labor (L), (approx.) thousand people
1	2	3	4	5	6
1	2011	1714.4	12.4	555.3	1290
2	2012	2376.6	13.1	619.7	1350
3	2013	2678.6	13.9	725.4	1380
4	2014	2259.3	14.1	766.6	1300
5	2015	2720.7	12.4	832.8	1300
6	2016	1998.0	56.2	891.0	1590
7	2017	1855.3	26.8	876.6	1400
8	2018	2319.3	17.6	894.4	1360

According to Table 1 for the realization of the Equations (3) and (2), dependent variable X can act as X1 (the 3rd column of Table 1) and X2 (the 4th column). As on schedule (Figure 1) dynamics of X2 is more stable than X1, we will accept X = X2 [5].

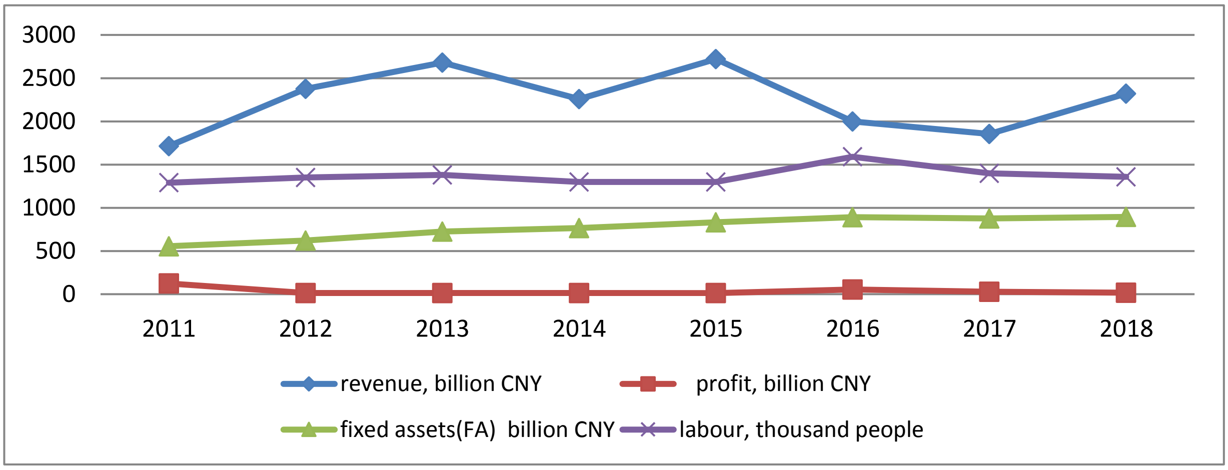

Figure 1: Diagrams of input for data modeling of CNPC

The number of elements of basic data N at number of parameters v = 3 (X, K and L) is necessary to be:

$$N - v \geq 2 \qquad (4),$$

and it's true, because according to (4) $N - v = 8 - 3 = 5 > 2$.

We should note the fact that in general Table 1 and Figure 1 have parameters that are rather stable for eight years.

3. RESULTS AND DISCUSSION

To obtain the equation of linear multiple regression (Equation (3)) previously it is necessary to realize the following intermediate procedures:

1. To carry out the component analysis of all variables in Equations (1) and (2): resulting function X has to correlate closely with independent variables K and L (the correlation coefficients ρXK, ρXL ≥ 0.7 – threshold value [6]), and arguments among themselves (that is correlation coefficients between the explicative variables) have to correlate as little as possible (ρKL <0.7).

2. To be convinced that all elements of the chosen variables were distributed under the normal law and the regression model was used correctly.

We will make calculations of coefficients of Spearman Rank Correlation taking into account groups of connected ranks for carrying out the component analysis [7], and we will have

$\rho XK = 0.69 \approx 0.7$; $\rho XL = 0.75 \geq 0.7$; $\rho KL = 0.55 < 0.7$.

Therefore, variables in Equations (1) and (3) meet the requirements of the component analysis.

On the accepted classification given in [7], the number of variables belongs to midget samples (N <25 units; we have N = to 8 units – see Table 1; at 25 <N <50 – small samples; in case of N> 50 – ordinary statistical samples). Further check of variables X, K, and L on the normal law of distribution which we will carry out in three steps have to follow:

1) creation of variational series (VS) with the finding of values of empirical frequencies of f_i, $i = 1, n$ where n is number of elements of VS (see the number of lines of Table 2) with the calculation of the weighted average x_a^w and mean square deviation σ_x;

2) creation of theoretical distribution (operation of graduation) with the finding of theoretical frequencies f_i^t, $i = 1, n$;

3) application to the received frequencies of f_i and f_i^t of the chi-squared test (χ^2) and, as specified, Rom test (usually Kolmogorov test is used only at ordinary samples). The general for all tests is that the results received for concrete samples with the known share of confidence extend also to their VS.

We carry out a further check of the examined elements from the realization of the general part for all variables – the creation of VS on the example of variable X with N = 8. VS is a means of statistical data grouping and distribution of population units on quantitative criterion [8, p. 7].

The examined population У = {y_j}, j = 1, N = 8 will have elements of Table 1 column 4 which need to be displayed in VS, having the appearance of future Table 3 containing n lines that can be

$$\tau: Y \rightarrow Z \qquad (5),$$

where Z = {z_i}, i = 1, n – number of lines of future VS. At the same time the purpose of such construction is obvious: to make N much more surpassed number of lines n that it is possible to write down as (n <<N). The number of "n" is calculated on Sturges' rule [7, p. 95]:

$$n = 1 + 3.222 \log N \qquad (6)$$

For our example N = 8 and log 8 = 0.9 and on formula (6) we have $n = 1 + 3.222 \cdot 0.9 = 3.9 \approx 4$ (number of intervals).

For interval VS constructing it is necessary to calculate step size:

$$h = \frac{R}{n} \qquad (7),$$

where $R = (y_{max} - y_{min})$ is the range of grouping SP. For chosen output variable Z sample range (see Table 1) $R_Z = 52.6 - 12.4 = 43.8$ (billion CNY). According to the formula (7): $h = 43.8 / 4 \approx 11.0$ (billion CNY), and VS for variable X, presented in columns 1–4 Table 3.

Table 2: Worksheet for calculating empirical and theoretical frequencies for variable X = Profit of CNPC

i	Profit. billion CNY $Z_i^s - Z_i^f$	Empirical frequencies. f_i	Class mark Z_i^a	$Z_i^a - Z_a^w$	$\frac{Z_i^a - Z_a^w}{\sigma} = t$	$\varphi(t)$	Theoretical frequencies $f_i^t = (\frac{N \times h}{\sigma}) \times \varphi(t)$
1	2	3	4	5	6	7	8
1	12.4–23.4	6	17.9	-5.5	-0.50	0.3521	2.82≈3
2	23.4–34.4	1	28.9	5.5	0.50	0.3521	2.82≈3
3	34.4–45.4	0	39.9	16.5	1.50	0.1295	2.03≈2
4	45.4–56.4	1	50.9	27.5	2.50	0.0175	0.14≈1
Total		8	-	-	-	-	Total: 9

We will record only the following results:

$Z_a^w = X_a^w = 23.4$ billion CNY

$\sigma_X = 11.0$ billion CNY

$\frac{N \times h}{\sigma} = Const_X = 8$ (non-dimensional value)

Weighted average and mean square deviation allow to conclude that profit of the company can be as

$$X_a^w \pm \sigma = 23.4 \pm 11 \quad (12.4 \sim 34.4) \text{ billion CNY.} \qquad (8)$$

Equation (8) presents examined population in the form of closed interval. It means that variable $X = (Net\ Profit\ \times billion\ CNY)$.

In Equation (1) variable X with empirical measures from 12.4 to 34.4 billion CNY that according to mean square deviation occupy 68% of all distribution [9].

Empirical frequency (column 3 in Table 2) and theoretical frequency (column 8 in Table 2) allow to use chi-squared test [6]:

$$X^2 calc = \sum \frac{(f_i - f_i^t)^2}{f_i^t} \qquad (9)$$

Table 3: Data calculation by chi-squared test

i	Empirical frequencies f_i	Theoretical frequencies. f_i^t	$f_i - f_i^t$	$(f_i - f_i^t)^2$	$(f_i - f_i^t)^2/f_i^t$
1	2	3	4	5	6
1	6	3	3	9	3
2	1	3	- 2	4	1.33
3	0	1	- 1	1	1
4	1	1	0	0	0
-	8	8	-	-	$\chi_c^2 = 5.33$

It is necessary to compare value of chi-squared test $\chi_c^2 = 5.33$ with table value χ_t^2 on degrees of freedom $(k = n - 3 \rightarrow k = 4 - 3 = 1)$ and significance level $\alpha = 0.05$.

Using Application 4 find table value of chi-squared test $\chi_t^2 = 3.84$ for $k = 1$ and $\alpha = 0.05$. As $\chi_c^2 = 5.33 > \chi_t^2 = 3.84$ (or differ from boundary value 3.84 approximately on 39%) [8].

Divergences between empirical frequencies (f_i) and theoretical frequencies (f^t) cannot be considered accidental. Then previously our null hypothesis about proximity of empirical distribution to normal one has to be formally disproved. That is elements distribution of statistical population of X is not subordinated to the normal distribution law (NDL).

Besides, N = 8 is less than size of even small samples, when $N \leq 25$. However, there is an opportunity to address one more test with an opportunity to be convinced not of so strict requirement in a deviation of the accepted earlier null hypothesis about accessory of distribution of the studied samples under the normal law. Let's calculate by Romanovsky test (10):

$$Rom_c = \frac{|\chi_c^2 - k|}{2k^{0.5}} = \frac{|5.33 - 1|}{2^{0.5}} = 3.07 > 3 \qquad (10)$$

If $Rom_{crit} = 3$ and $Rom_c = 3.07$; We have $3.07 > 3$.

Normal distribution law confirms the discrepancy of distribution of the same population of X. However, this strict inequality differs from demanded one only for 2.3% (a divergence with normal distribution law is insignificant). but not for 39% (χ_c^2 in chi-squared test). Therefore, we consider that elements of variable X in the equations (1) and (2) are distributed "quasinormally". that is close to normal distribution law. In our opinion for the purpose of at least almost acceptable streamlining's of the received excesses in their relative expression (in %). it is necessary to use the Table of Yadov [9, 10]:

Table 4: Reliability assessment of sample observation [9, 10]:

№	Level of reliability	Tolerable error of sample. %
1	High	to 3
2	Ordinary	3 – 10
3	Approximate	10 - 20
4	Assessed	20 - 40
5	Rapid	more than 40

If in Table 4, «Reliability» is measure of deviation of calculated values in tests and "Degree of tolerable sample deviation from the normal distribution law". measure of exceedance of table value calculated by chi-squared test (39%) can be "assessed" reliability of distributed elements of random variable with generic name X according to the normal distribution law. Thus, according to Romanovsky test it should be included in low ordinary (i.e. in the range from 0 to 3% in order to avoid the term high), as in practice 3.07 ≈ Rom_{crit}= 3. However. one of possible reasons of deviation from the normal distribution law may be conditions of unstable economy to 2016. when profit was 56.2 billion CNY (see Table 1): to leave or to exclude it in the studied sample? Grubbs test can help to answer this question [7].

For test of observation equal to 52.6 billion CNY in variable $X = Profit$ for 2015. being N, which is allocated in sample of volume of approximately normally distributed set, is calculated in the following way

$$\text{Кcalc} = \sum_{i=1}^{N-1}(y_i - y_N^a)/(\sum_{i=1}^{N}(y_i - y^a) \quad (11)$$

In Equation (11), y_i is current values of population elements in column 4 of Table 1. Average values of y^a and y_N^a for rather high values are calculated on formulas (12) and (13) respectively:

$$y_N^a = \sum_{i=1}^{N-1}(y_i)/N \quad (12)$$

$$y_N^a = \sum_{i=1}^{N-1}(y_i)/(N-1) \quad (13).$$

Calculations on formulas (12). (13) shows that $y^a = 20.813$; $y_N^a = 15.757$. In this case the formula will be following Equation (11):

$$Кcalc = \frac{161.14}{1592.3} = 0.101$$

Calculated values $Кcalc$ is necessary to compare with table ones. The fragment from Grubbs table is presented in Table 5 [8].

Table 5: Values of $К_{tab}$ in Grubbs table

i	α = 0.01	α = 0.05	I	α = 0.01	α = 0.05	I	α = 0.01	α = 0.05
3	0.0001	0.0027	9	0.2411	0.3742	…..	……...	……...
4	0.0100	0.0494	10	0.2831	0.4154	20	0.5393	0.6379
5	0.0442	0.1270	……	………	……...	…..	……...	……...
6	0.0928	0.2032	……	………	……...	…..	……...	……...
7	0.1447	0.2696	15	0.4401	0.4401	25	0.6071	0.6923
8	0.1948	0.3261	……	……..	……...	…..	……...	……...

Notes: designated by dots at the set number of observations of N is defined by simple interpolating.

Calculated value of ratio Equation (11) $Кcalc = 0.101$ is compared with table value $К_{tab}$ (see Table 5) at certain number of observations N and significance level α and characterizes that extreme size which with probability $(1 - \alpha)$ can be explained with the accidental reasons. If $Кcalc < К_{tab}$. probability of divergences in the sums of mean square deviation Equation (11) can be explained with the accidental reasons. and it is equal to significance level α and owing to small probability it is impossible. In this case the observation (52.6 billion CNY) should be cancelled and for further calculation we will use rest $(N - 1 = 7)$ of observation for variables K and L.

Then we should confirm the normality of elements distribution of the same updated set of X and other sizes in the Equation (1)

For this purpose. calculations are necessary on Equations (5)–(10), and also in Tables 1, 2 and 4 for variable X and then for its arguments with an exception of data number for 2016 from Table 1. As a result, we will have:

1. For variable $X = \text{Profit}$: $\chi_c^2 = 2.83 < \chi_t^2 = 3.84$ according to chi-squared test with the same significance level $\alpha = 0.05$ elements of the statistical population are distributed by the normal

distribution law. as in Rom test $Rom_p = 1.30 < Rom_{bound} = 3$. Thus. picture without profit value of 52.6 billion CNY for 2016 for variable X significantly improved: the normality of distribution of its elements by two criteria does not raise doubts. Besides comparison Equation (8) $X_a^w \pm \sigma = 23.4 \pm 11$ billion CNY with $X_a^w \pm \sigma = 16.2 \pm 3.8$ billion CNY shows that the lower bound practically did not change. But average values and top limits change considerably and their reasons are quite clear.

2. For variable $K = Capital: \chi^2_{calc} = 6 > \chi^2_{tab} = 3.84$ (deviation with boundary value is 56% and according to Table 5, the nature of reliability on normality of distribution is "rapid"; according to Rom test $Rom_c = 3.54 > Rom_{bound} = 3$ (deviation boundary value is 18%: elements are distributed with approximate reliability). Concerning N = 8, the picture worsened a little.

3. For variable $L = Quantity\ of\ Labour: \chi^2_{calc} = 7 > \chi^2_{tab} = 3.84$ (excess of boundary Table value at the same $\alpha = 0.05$ for 82% is too large); according to Rom test $Rom_c = 4.24 > Rom_{bound} = 3 - excess\ \ for\ \ 41\% \approx 40\%$.

The nature of reliability of elements distribution under the normal law is close to "assessed". Concerning N = 8 the picture worsened a little.

Thus, we chose as an object of research CNPC corporation which is objectively reflecting in data activity in the conditions of turbulent economy. This circumstance directly affects as on rather small period of flashback (from 2011 to 2018; on N = 8 values of the examined data). The nature of basic data. In parameter $X = "Profit"$ arose need of removal from further data consideration for 2016 in 52.6 billion CNY by Grubbs test. The PF models which remained for construction are population on N = 7.

Creation of theoretical laws of distribution with the subsequent use of criteria of chi-squared test and Rom test showed that only elements of variable X after use of Grubbs test began to satisfy to the normal law of distribution whereas elements of the K and L variables at N = 7 are distributed "quasinormally" in gradation of the table of Yadov (see Table 5).

4. CONCLUSION

Results of the carried-out calculations give the opportunity to come to the following conclusion: on basic data of Table 1 with removal from data consideration for 2015 creation of production function as multiplicative equation of multiple regression (3) with the subsequent potentiation and receiving production function (2) by LS method is necessary to find possible.

5. DATA AND MATERIAL AVAILABILITY

Relevant information regarding this study is available by request to the corresponding author.

6. ACKNOWLEDGEMENT

The work is performed according to the Russian Government Program of Competitive Growth of Kazan Federal University. Financial supports are gratefully acknowledged.

7. REFERENCES

[1] Burlachov, A. (2018). Turbulentnost' ekonomicheskih processov: teoreticheskie aspekty / V. Burlachov. *Voprosy ekonomki*, 11: S. 90.

[2] Pavlova, Yu.N. (2016). Upravlenie predpriyatiem v usloviyah nestabil'nosti vneshnej sredy. *Social'no-ekonomicheskie yavleniya i process*, 10(2): S. 84.

[3] Schneeweiss, H., and Augustin, T. (2006). Some recent advances in measurement error models and methods. Department of statistics, thom: 90 number, 1: 183-197

[4] Statistics of the Shanghai stock exchange *http://globalcor porations.ru /CNPC.html*

[5] Zhai, Y., and Fang, Z. (2018). Locally optimal designs for binary dose-response models. *Canadian journal of statistics : thom:46: number*, 2: 336-354

[6] Studenmund, A.H. Using Econometrics: A Practical Guide (7th Edition) *https://www.amazon.com/gp/aws/cart/add.html?ASIN.1=013418274X&Quantity.1=1&AWSAccessKeyId=AKIAJBMCXM6F64NRT62Q&AssociateTag=wallstreetmoj-20*

[7] Obshchaya teoriya statistiki: Uchebnik. (2013). T.V. Ryabushkin, M.R. Efimova, I.M. Ipatova, N.I. Yakovleva. – M.: Finansy i statistika: 279 s.

[8] Gromyko, G.L. (2012). Obshchaya teoriya statistiki: Praktikum. – M.: INFRA-M, 139 s. (Vysshee obrazovanie).

[9] 3-sigma rule *https: //wiki.loginom.ru/articles/3-sigma-rule.html*

[10] Yadov, V.A. (1987). Sociologicheskoe issledovanie: metodologiya, programma, metody. – M.: Nauka, -248.

Akhmetova Irina A is an Associate Professor at Department of Production Economics, Institute of Management, Economics and Finance, Kazan Federal University. She is a Candidate of Economic Sciences. She had her Higher Education: KFEI, Labor Economics, Economics and Planning. She is interested in international Finance and Analysis.

Dr.Shikhalev Anatoly M. is an Associate Professor at the Institute of Management, Economics and Finance, Kazan Federal University, Head of the group for the collection, processing and systematization of information of the Institute of Social, Economic and Legal Sciences of the Academy of Sciences of the Republic of Tatarstan. He holds a PhD in Economics. He is interested in Financial information and System.

Dr.Demyanova Olga V. is an Associate Professor Institute of Management, Economics and Finance, Kazan Federal University, Russia. She holds a Doctor of Economic Sciences degree. She is interested in Strategic Analysis, Modelings of the Efficiency of the Modern Economy.

Sirazieva Zarina N. is an Assistant Professor at the Institute of Management, Economics and Finance, Kazan Federal University. She is a Candidate of Philology. She had her higher education from ChOU VPO "Institute of Economics, Management and Law (Kazan)", Management and Marketing, State and Municipal Administration, and Higher Professional Education from Tatar State Humanitarian and Pedagogical University.