CONSTRUCTION OF ECONOMIC MODELS

Oliynyk Viktor

Ph.D., Associate Professor

Sumy State University , Department of Finance, 2 Rymskogo-Korsakova St , Sumy, Ukraine, 40007;

oliynyk.viktor@gmail.com

In this work the algorithm of determination two-factor optimum regression equation in the presence of a small amount of observations is considered. Using a method of least squares, there are unknown constants which was most probablly describe input datas. Deriving of mathematical model does not demand bulky evaluations, and all necessary parametres are shown in the table. Correspondence of regression equation and the differential equation with a boundary condition is reduced. The received boundary value problem adequately describes investigated process and considers the discovered factors in the regression equation. Quality of the equation of a regression is characterised by determination factor. Formulas for definition of a significance level of quality of a prediction, and also a confidential zone of a prediction regression models are reduced. As numerical outcomes the discovered optimum models describing Gross domestic product based on purchasing-power-parity (PPP) valuation of country GDP (variable Y) in Sultanate of Oman from similar indicators in United Arab Emirates, Japan and United States (a variable X) are reduced. Also the various economic indicators GDP of Oman are calculated, what depends on the temporary factor. The predicted values on the future phases are reduced. These mathematical models are presented both in an aspect of regression functions, and in the form of the differential equations with boundary conditions.

Keywords: expectation, variance**,** regression analysis, component analysis, determination factor.

JEL classification codes: C01; C50

Introduction

In a modern society there is an association between economic indicators both in the country, and between the different countries. In a more comprehensive sense, economic indicators are influenced by social, cultural, legal and other fields of activity of a society.

For forecasting of economic reforms, it is necessary to make economic-mathematical models. These models help to make an optimum solution, considering many restrictions. There are many various mathematical methods allowing to treat considerable information content and to construct model. It would be desirable to stop on two: a component analysis, the regression analysis.

The component analysis is necessary for handling of great volume of the various information (not necessarily economic), with their subsequent grouping. In this analysis basically are used endogenic variables. From set of various parametres which there can be some tens or even hundreds, by scientifically justified mathematical method, there are some integral indicators which as a whole reflect investigated process. The analysis of the received data, allows to receive the primary factors reflecting the investigated appearance. As an example the state model as a whole or its economic component can be considered. At numerical realisation of this method it is necessary to consider, that the amount of observable factors depends on an amount of investigated variables.

The regression analysis allows to receive model which with certain degree of accuracy, reproduces input datas. There is an one-dimensional and many-dimensional regression, linear and nonlinear. One of models is the square many-dimensional regression. Its essence is that the full square form, and then in the beginning is constructed, using certain criteria, the amount of items in model decreases.

Determination of function of a regression happens on empirical data which contain accidents, therefore in the regression equation association between variables gains stochastic character. Selection of function of a regression which as is possible is better characterised studied regularity of connection between a successful variable and explaining variables, is a basis the regression analysis. We will consider an one-factor problem of determination optimum of regression function in more details.

Problem statement

At modelling of monotone processes when the amount of unknowns is insignificant, as research following nine functions can be used. These associations possess such property, that if separate values of variables X and Y satisfy to one of the equations average values also to it satisfy. For each of functions there are characteristic averages which can be arithmetical, geometrical and harmonic averages in this case. Correspondence of investigated functions and their average magnitudes is reduced in table 1. In this table $a, b = const$.

Table 1. Aspect of the average magnitudes characterising functions of a regression

№	Function aspect	Characteristic averages	
		$\bar{X}$	$\bar{Y}$
1	$Y = a + bX$	$\bar{X} = \sum_{i=1}^{n} X_i / n$	$\bar{Y} = \sum_{i=1}^{n} Y_i / n$
2	$Y = a + b \ln X$	$\bar{X} = \sqrt[n]{X_1 X_2 ... X_n}$	$\bar{Y} = \sum_{i=1}^{n} Y_i / n$
3	$Y = a + b / X$	$\bar{X} = n / \sum_{i=1}^{n} (1 / X_i)$	$\bar{Y} = \sum_{i=1}^{n} Y_i / n$
4	$Y = ab^X$	$\bar{X} = \sum_{i=1}^{n} X_i / n$	$\bar{Y} = \sqrt[n]{Y_1 Y_2 ... Y_n}$
5	$Y = aX^b$	$\bar{X} = \sqrt[n]{X_1 X_2 ... X_n}$	$\bar{Y} = \sqrt[n]{Y_1 Y_2 ... Y_n}$
6	$Y = \exp(a + b / X)$	$\bar{X} = n / \sum_{i=1}^{n} (1 / X_i)$	$\bar{Y} = \sqrt[n]{Y_1 Y_2 ... Y_n}$
7	$Y = 1/(a + bX)$	$\bar{X} = \sum_{i=1}^{n} X_i / n$	$\bar{Y} = n / \sum_{i=1}^{n} (1 / Y_i)$

8	$Y = 1/(a + b\ln X)$	$\overline{X} = \sqrt[n]{X_1 X_2 ... X_n}$	$\overline{Y} = n / \sum_{i=1}^{n} (1/Y_i)$
9	$Y = X/(a + bX)$	$\overline{X} = n / \sum_{i=1}^{n} (1/X_i)$	$\overline{Y} = n / \sum_{i=1}^{n} (1/Y_i)$

Determination of one optimum function happens in some stages. At the first stage the necessary average magnitudes for variables X and Y are calculated. At the second stage, depending on $X_i < \overline{X} < X_{i+1}$, by means of linear interpolation values are calculated

$$\hat{Y}_i = Y_i + \frac{Y_{i+1} - Y_i}{X_{i+1} - X_i} (\overline{X} - X_i) \qquad (1)$$

At the third stage it is defined one of nine functions which in the best way describes input datas. As criterion of selection it is possible to use a condition

$$\left| \frac{\hat{Y} - \overline{Y}}{\hat{Y}} \right| \to \min \qquad (2)$$

The unknown constants, which are in the regression equation, are calculated by means of a method of least squares. This method is a basis of the regression analysis and consists of performance of a following condition for function of errors

$$S = \sum_{i=1}^{n} (Y_i - \hat{Y}_i)^2 \to \min \qquad (3)$$

Determination of an extremum (3) for linear regression functions, is reduced to a solution of linear system of the algebraic equations concerning parametres a and b. It is proved, that this system has a unique solution and function of errors S reaches the minimum. For application of a method of least squares to all regression functions (table 1), it is necessary to transform them beforehand. This transformation consists in their information to a linear aspect. Unknown constants which are calculated from a condition (3) definitively define an aspect of the optimum equation of a regression.

It is possible to continue modelling of initial process and receive the concrete differential equation which maps an investigated appearance. The constants entering into this differential equation are directly connected with the constants entering in the regression equation. In table 2 we will reduce values of constants a and b, which are calculated from condition (3), and also a corresponding boundary value problem for every regression equations. In this table following labels numerical magnitudes are used: $M(X)$ - expectation; $D(X)$ - a variance; $K(XY) = M(XY) - M(X)M(Y)$ - the correlative moment. Thus, knowing the optimum regression formula it is possible to construct corresponding mathematical model in the form of the differential equation.

Table 2. - Correspondence of the regression formula and a boundary value problem of Cauchy

№	Function aspect	Parametres a and b	The differential equation	Boundary condition
1	$Y = a + bX$	$a = M(Y) - bM(X)$ $b = K[XY]/D(X)$	$Y' = b$	$Y(1) = a + b$
2	$Y = a + b\ln X$	$a = M(Y) - bM(\ln X)$ $b = K[\ln(X)Y]/D(\ln X)$	$Y'X = b$	$Y(1) = a$
3	$Y = a + b/X$	$a = M(Y) - bM(X^{-1})$ $b = K[X^{-1}Y]/D(X^{-1})$	$Y'X^2 = -b$	$Y(1) = a + b$
4	$Y = ab^X$	$\ln a = M(\ln Y) - \ln bM(X)$ $\ln b = K[X\ln Y]/D(X)$	$Y'/Y = \ln b$	$Y(1) = ab$
5	$Y = aX^b$	$\ln a = M(\ln Y) - bM(\ln X)$ $b = K[\ln X \ln Y]/D(\ln X)$	$Y'X/Y = b$	$Y(1) = a$
6	$Y = \exp(a + b/X)$	$a = M(\ln Y) - bM(X^{-1})$ $b = K[X^{-1}\ln Y]/D(X^{-1})$	$Y'X^2/Y = -b$	$Y(1) = \exp(a + b)$
7	$Y = 1/(a + bX)$	$a = M(Y^{-1}) - bM(X)$ $b = K[XY^{-1}]/D(X)$	$Y'/Y^2 = -b$	$Y(1) = 1/(a + b)$

8	$Y = 1/(a + b\ln X)$	$a = M(Y^{-1}) - bM(\ln X)$ $b = K[\ln(X)Y^{-1}]/D(\ln X)$	$Y'X/Y^2 = -b$	$Y(1) = 1/a$
9	$Y = X/(a + bX)$	$b = M(Y^{-1}) - aM(X^{-1})$ $a = K[X^{-1}Y^{-1}]/D(X^{-1})$	$Y'X^2/Y^2 = a$	$Y(1) = 1/(a + b)$

Quality of stochastic connection between variables Y and X (quality of the regression equations) can be estimated by means of factor of determination which is calculated under the formula

$$R^2 = 1 - \sum_{i=1}^{n}(Y_i - \hat{Y}_i)^2 / \sum_{i=1}^{n}(Y_i - \overline{Y})^2 \qquad (4)$$

The formula (4) shows, how percent from the general variance of variable Y explains investigated the regression equation. It is possible to make the deduction, that the problem of minimisation of function of errors (3) using the method of least squares, is equivalent to problem of maximisation of factor of determination (4). The more value R^2 is closer to 1, the quality of the received model is better.

Check of the importance of the equation of a regression as a whole it is possible to realise using various methods, for example with the help of F - Fisher's distributions. Some parametre F is discovered for this and compared it to table values of distribution of Fisher $F^t(m, n - m - 1)$ at set level of significances α . At condition performance

$$F = \frac{R^2}{1 - R^2}\frac{n - m - 1}{m} > F^t \qquad (5)$$

it is possible to make a conclusion on the importance of the equation of a regression.

Check of the significance of factors of the equation of a regression can be made by means of a t - distribution, comparing the received values with table values.

In the supposition, that investigated magnitudes are distributed under the normal law, we will calculate a variance of magnitude $\hat{Y}_i = \hat{Y}(X_i)$

$$D(\hat{Y}_i) = \left[\frac{1}{n} + \frac{(X_i - \overline{X})^2}{\sum_{i=1}^{n} (X_i - \overline{X})^2} \right] \frac{1}{n-m-1} \sum_{i=1}^{n} (Y_i - \hat{Y}_i)^2 \qquad (6)$$

The confidential zone of a prediction regression model can be received approximately in an aspect

$$\hat{Y} - t_{1-\alpha} \sqrt{D(\hat{Y})} \le \tilde{Y} \le \hat{Y} + t_{1-\alpha} \sqrt{D(\hat{Y})} \qquad (7)$$

where $t_{1-\alpha}$ - a quantile of a distribution with n-m-1 degree of freedoms at significance level α .

Numerical examples

As an example, using data of International Monetary Fund for 2000-2007, we will consider following two-factor regression models:

i) association Gross domestic product based on purchasing-power-parity (PPP) valuation of country GDP (variable Y) in Sultanate of Oman from similar indicators in United Arab Emirates, Japan and United States (a variable X). These associations can use in that case when United Arab Emirates, Japan and United States have predicted the values on the future phases. For determination of the inverse relationships, these regression models cannot be used and it is necessary to make additional researches.

Table 3. - Optimum models GDP.

	United Arab Emirates, billions US dollars (Z=X/100)	Japan, billions US dollars (Z=X/3000)	United States, billions US dollars (Z=X/10000)
Oman, billions US dollars, regression model	Y=1/(0,046-0,0168*Z) R^2=0,986	Y=26,6578*Z^1,6878 R^2=0,984	Y=Z/(0,0366-0,0058*Z) R^2=0,992
Oman, billions US dollars, Boundary value problem	$Y'/Y^2 = 0{,}0168$ $Y(1) = 34{,}2465$	$Y'Z/Y = 1{,}6878$ $Y(1) = 26{,}6578$	$Y'Z^2/Y^2 = 0{,}0366$ $Y(1) = 32{,}4675$

ii) determination of economic indicators GDP of Oman (variable Y) depending on the temporary factor (t=1,2 … 8; accordingly 2000-2007) and forecasting of these indicators for the future phases. Analyzing by means of a method of least squares investigated economic indicators, it is received, that for all of them the optimal regression function is $U = 1/(a + bt)$. According to table 2 of this function there corresponds following boundary value problem $U'/U^2 = -b$; $U(1) = 1/(a + b)$. In table 4 regression models of various economic indicators GDP of Oman with their prognosis on following phases are reduced.

Table 4. - Some economic indicators GDP of Oman

№	Economic indicator	Parametre a	Parametre b	The prognosis 2008 (Y)	The prognosis 2009 (Y)	R^2
1	Gross domestic product based on purchasing-power-parity (PPP) per capita GDP, US dollars; U=Y/10000	0,758268	-0,028418	19900,16	21093,01	0,988
2	Gross domestic product based on purchasing-power-parity (PPP) valuation of country GDP Billions , US dollars ; U=Y	0,034264	-0,001665	51,867	56,769	0,995
3	Gross domestic product per capita, constant prices, National currency: U=Y/1000	0,399035	-0,008454	3096,421	3179,651	0,948
4	Gross domestic product, constant prices, Billions, National currency U=Y	0,181012	-0,006225	8,000	8,420	0,982
5	Gross domestic product, current prices, Billions, National currency U=Y	0,150585	-0,009958	16,402	19,604	0,953

Check of the significance of the received factors of determination by means of a relation (5), confirms the supposition that the discovered models was probablly describe input datas and can be used in practical calculations.

Conclusion

Thus, the outcomes received in this work, allow to choose the optimal function from offered nine. The given algorithm well works for monotone input datas and when an amount of observations not so big. Checking of the importance of the received equations, shows their practical working capacity. On the basis of the received associations it is possible to predict economic indicators GDP of Oman on the future phases. It is necessary to pay special attention to mathematical models which are described by the differential equations. It is possible to continue the research with use of the theory of a component analysis, and also many-dimensional nonlinear regression analysis.

References

1.Dougherty C.(1992). Introduction to Econometrics Oxford University Press

2.Goldberger A.A. (1990) Course in Econometrics Cambridge, MA: Harvard University Press

3.Green W. (1993) Econometric Analysis, second edition/ Macmillan Publishing Company

4.Pindyck R., Rubinfeld D. (1991) Economics Model and Economic Forecasts, third edition. McGraw

5.Harman H.H. (1968) Modern factor analysis. Chicago and London: The University of Chicago press

Oliynyk, V. M. Construction of economic models [Text] / V. M. Oliynyk // The1 st International Conferenceon Applied Businessand Economics (ICABE 2009), Faculty of Business, Sohar Universyity, Sohar, Sultanate of Oman, 30-31 March, 2009.- P. 258-263.