MINISTRY OF EDUCATION AND SCIENCE OF UKRAINE
WEST UKRAINIAN NATIONAL UNIVERSITY

Methodical instructions and recommendations

for solution of training tasks in discipline

«ECONOMETRICS»

Тернопіль – 2022

UDK 519.2

Reviewers:

S.V. Martyniuk – PhD, Candidate of physical and mathematical sciences, associate professor of the Department of Mathematics and Teaching Methods at Volodymyr Hnatyuk Ternopil National Pedagogical University.

O.S. Bashutska – PhD, Candidate of Economics, associate Professor of the Department of Economic Cybernetics and Informatics of the West Ukrainian National University;

approved at the meeting of the Department of Applied Mathematics, protocol № 1 of 26.08.2022.

Examples of solution of the training tasks in the course «Econometrics» are considered in methodical instructions on the base of necessary theoretical knowledge. Methodical instructions to study the discipline «Econometrics» are proposed for students for helping them in solution of training tasks.

Plaskon S.A., Dzyubanovska N.V. Methodical instructions and recommendations for solution of training tasks in discipline «Econometrics». - Ternopil: WUNU, 2022.- 17 p.

UDK 519.2

Responsible for release: O.M. Martyniuk, PhD, Candidate of physical and mathematical sciences, associate professor of the Department of Applied Mathematics,

Head of the Department of Applied Mathematics WUNU

Training tasks give for students the next opportunity:

analyze economic objects and processes;

build and analyze econometric models, make sound economic conclusions and calculate forecast indicators;

use advanced information technologies and software systems for modelling economic phenomena and processes;

form a complete system of theoretical knowledge from the course "Econometrics";

learn the methodology and technique of calculating economic indicators, the conditions for using certain econometric methods for a comprehensive analysis of socio-economic processes;

learn to use the results of econometric research in practical management activities;

learn more deeply and consolidate the theoretical knowledge obtained at the lectures.

Topic of training in the discipline "Econometrics": Application of econometrics methods for solving economic problems.

Stages of training:

1. Create a database of the main macroeconomic indicators of Ukraine for the years 2000-2021: GDP, export and import of goods and services, inflation, the exchange rate of the US dollar to the hryvnia, investments, the volume of deposits of the population in banks, etc. For work, choose 3 time series (each student his data). Series must contain at least 20 observations. To determine the necessary information, use the following sites:

https://www.me.gov.ua,
https://www.bank.gov.ua/,
https://knoema.com/.

2. With the help of statistical packages, conduct a graphical analysis of data series. Determine the type of econometric dependence and build a model.

3. For the selected series of data, conduct a variance analysis and test the adequacy of the model.

4. For the selected series of data, check the presence of multicollinearity, heteroskedasticity, and autocorrelation.
Make the necessary conclusions.

Task. For ten enterprises of the region, for a certain conditional period, the numerical values of two economic indicators are known: gross production ***y*** (million dollars) and the cost of the main production assets ***x*** (million dollars), (Table 1). To study the characteristics of the influence of the cost of the main production assets (***x***) on the output of gross products (***y***) of the enterprise:

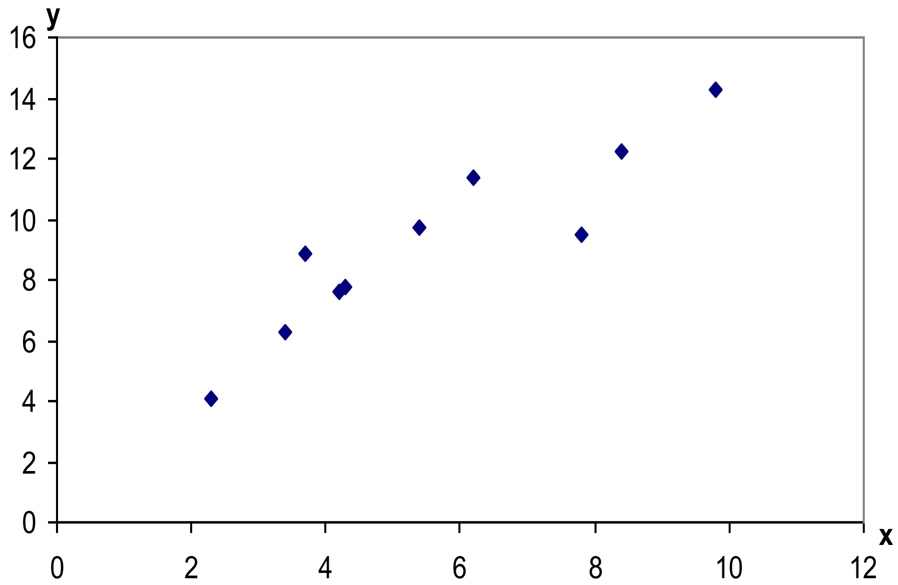
1. Make a model specification.
2. Find statistical estimates of the parameters of the linear regression equation and construct an estimation line.
3. Calculate the value of the point estimate of the dispersion of disturbances.
4. For the level of significance $\alpha = 0{,}05$ check the significance of the regression coefficients α_0 and α_1.
5. Find the confidence intervals of the regression coefficients with reliability $\gamma = 0{,}95$.
6. Calculate the values of the sample: coefficient of determination, coefficient of correlation.
7. Find and construct the confidence interval of the regression function with reliability $\gamma = 0{,}95$.
8. Check the adequacy of the constructed econometric model. If the model is adequate, then find the forecast value of the gross output for the value of the main production assets in the amount of UAH 12 million, with reliably $\gamma = 0{,}95$ build a confidence interval for this forecast value.

Table 1

№ firms	Gross output, mln. dol., y_i	Main production assets, mln. dol., x_i
1	4,1	2,3
2	6,3	3,4
3	7,6	4,2
4	8,9	3,7
5	7,8	4,3
6	9,7	5,4
7	11,4	6,2
8	9,5	7,8
9	12,2	8,4
10	14,3	9,8

♦ ***Solution.***

1. Let's build a scatter diagram of the dependence of gross output (y) on the cost of the enterprise's main production assets (x):



The placement of points on the scatter diagram makes it possible to make an assumption about the existence of a linear form of the relationship in the form of a function:

$$\hat{y} = a_0 + a_1 x, \tag{1}$$

where $\hat{y}$ – estimated volume of gross output, mln. dol.; x – the cost of the main production assets, mln. dol.

2. Statistical estimates of the parameters of the linear regression equation can be found by the method of least squares (LSM), which is based on the requirement to minimize the sum of squares of the deviations of the empirical values of the variable y from the values calculated by the equation of the straight line. With its help, such estimates of the parameters of the regression equation are found that minimize the selected measure of dispersion. These estimates can be found by the system of normal

$$\begin{cases} na_0 + \left(\sum_{i=1}^{n} x_i\right) a_1 = \sum_{i=1}^{n} y_i, \\ \left(\sum_{i=1}^{n} x_i\right) a_0 + \left(\sum_{i=1}^{n} x_i^2\right) a_1 = \sum_{i=1}^{n} x_i y_i. \end{cases} \tag{2}$$

Dividing each of the equations by n, we will get:

$$\begin{cases} a_0 + \overline{x} a_1 = \overline{y}, \\ \overline{x} a_0 + \overline{x^2} a_1 = \overline{xy}, \end{cases} \tag{3}$$

where $\overline{x} = \sum x_i / n$, $\overline{y} = \sum y_i / n$, $\overline{xy} = \sum x_i y_i / n$, $\overline{x^2} = \sum x_i^2 / n$.

Solution of the system (3) find by Kramer's rule:

$$a_1 = \frac{\overline{xy} - \overline{x} \cdot \overline{y}}{\overline{x^2} - \left(\overline{x}\right)^2}, \quad a_0 = \overline{y} - a_1 \overline{x}. \qquad (4)$$

The value of the same estimates can be found using formulas due to the deviation from the average:

$$a_1 = \frac{\sum_{i=1}^{n} (\Delta x_i \Delta y_i)}{\sum_{i=1}^{n} (\Delta x_i)^2}. \qquad (5)$$

The value of the assessment *a* we find from the formula (1.7):

$$a_0 = \bar{y} - a_1 \bar{x}, \qquad (6)$$

where $\Delta y_i = y_i - \bar{y}$ and $\Delta x_i = x_i - \bar{x}$ – this is the variance of the variables *y* and *x* from their average values.

To simplify calculations when finding estimates a_0 and a_1 let's build a table of parameters of the econometric model:

Table 2

№	y_i	x_i	x_i^2	y_i^2	$xi \cdot yi$	Δx_i	$(\Delta x_i)^2$	Δy_i	$\Delta x_i \cdot \Delta y_i$
1	4,1	2,3	5,29	16,81	9,43	-3,25	10,56	-5,08	16,51
2	6,3	3,4	11,56	39,69	21,42	-2,15	4,62	-2,88	6,19
3	7,6	4,2	17,64	57,76	31,92	-1,35	1,82	-1,58	2,13
4	8,9	3,7	13,69	79,21	32,93	-1,85	3,42	-0,28	0,52
5	7,8	4,3	18,49	60,84	33,54	-1,25	1,56	-1,38	1,73
6	9,7	5,4	29,16	94,09	52,38	-0,15	0,02	0,52	-0,08

7	11,4	6,2	38,44	129,96	70,68	0,65	0,42	2,22	1,44
8	9,5	7,8	60,84	90,25	74,1	2,25	5,06	0,32	0,72
9	12,2	8,4	70,56	148,84	102,48	2,85	8,12	3,02	8,61
10	14,3	9,8	96,04	204,49	140,14	4,25	18,06	5,12	21,76
Сума	**91,8**	**55,5**	**361,71**	**921,94**	**569,02**	**0**	**53,69**	**0**	**59,53**

Let's calculate the average values:

$$\bar{x} = \frac{\sum_{i=1}^{n} x_i}{n} = \frac{55,5}{10} = 5,55; \qquad \bar{y} = \frac{\sum_{i=1}^{n} y_i}{n} = \frac{91,8}{10} = 9,18.$$

$$\overline{x^2} = \frac{\sum_{i=1}^{n} x_i^2}{n} = \frac{361,71}{10} = 36,17; \qquad \overline{xy} = \frac{\sum_{i=1}^{n} x_i y_i}{n} = \frac{569,02}{10} = 56,9.$$

Let's write down the system (3):

$$\begin{cases} a_0 + 5,55a_1 = 9,18, \\ 5,55a_0 + 36,17a_1 = 56,9, \end{cases}$$

and solve it using the formulas (4):

$$a_1 = \frac{\overline{xy} - \bar{x} \cdot \bar{y}}{\overline{x^2} - \left(\bar{x}\right)^2} = \frac{56,9 - 5,55 \cdot 9,18}{36,17 - 5,55^2} = \frac{5,95}{5,37} = 1,11,$$

$$a_0 = \bar{y} - a_1 \bar{x} = 9,18 - 1,11 \cdot 5,55 = 3,02.$$

Let's find the same estimates using the formulas due to the deviation from the average. First, let's calculate the deviation of the variables from their average values:

$$\Delta x_1 = x_1 - \bar{x} = 2{,}3 - 5{,}55 = -3{,}25;$$
$$\Delta x_2 = x_2 - \bar{x} = 3{,}4 - 5{,}55 = -2{,}15;$$
....
$$\Delta x_{10} = x_{10} - \bar{x} = 9{,}8 - 5{,}55 = 4{,}25.$$
$$\Delta y_1 = y_1 - \bar{y} = 4{,}1 - 9{,}18 = -5{,}08;$$
$$\Delta y_2 = y_2 - \bar{y} = 6{,}3 - 9{,}18 = -2{,}88;$$
....
$$\Delta y_{10} = y_{10} - \bar{y} = 14{,}3 - 9{,}18 = 5{,}12.$$

Then
$$a_1 = \frac{\sum_{i=1}^{n}(\Delta x_i \Delta y_i)}{\sum_{i=1}^{n}(\Delta x_i)^2} = \frac{59{,}53}{53{,}69} = 1{,}11,$$
$$a_0 = \bar{y} - a_1\bar{x} = 9{,}18 - 1{,}11 \cdot 5{,}55 = 3{,}02, \qquad .$$

So, we obtained the estimation equation of the econometric model
$$\hat{y} = 3{,}02 + 1{,}11x.$$

Let's construct this estimation straight line.

3. Unbiased point estimate S_u^2 of the unknown variance of disturbances σ_u^2 we find by the formula:
$$S_u^2 = \frac{1}{n-2}\sum_{i=1}^{n}(y_i - \hat{y}_i)^2.$$

First, let's calculate the estimated values $\hat{y}_i$ according to the estimation equation of the econometric model $\hat{y} = 3{,}02 + 1{,}11x$, then deviation $u_i = y_i - \hat{y}_i$ and $u_i^2 = (y_i - \hat{y}_i)^2$.

Table 3

№	y_i	x_i	$\hat{y}_i$	$y_i - \hat{y}_i$	$(y_i - \hat{y}_i)^2$
1	4,1	2,3	5,58	-1,48	2,18
2	6,3	3,4	6,80	-0,50	0,25
3	7,6	4,2	7,68	-0,08	0,01
4	8,9	3,7	7,13	1,77	3,14
5	7,8	4,3	7,79	0,01	0,00
6	9,7	5,4	9,01	0,69	0,47
7	11,4	6,2	9,90	1,50	2,25
8	9,5	7,8	11,67	-2,17	4,73
9	12,2	8,4	12,34	-0,14	0,02
10	14,3	9,8	13,89	0,41	0,17
Сума	**91,8**	**55,5**	**91,8**	**0**	**13,2**

Using the obtained calculations, we get:

$$S_u^2 = \frac{1}{10-2} \cdot 13{,}2 = 1{,}65.$$

4. Significance of regression coefficients a_0 та a_1 we check using the null hypothesis $H_0(\alpha_m = 0)$, the content of which and alternative to it $H_1(\alpha_m \neq 0)$ is as follows: if the inequality holds

$$\left| \frac{a_m}{S_{a_m}} \right| > t_{кр.},$$

where $t_{кр.} = t_{двост.кр.}(\alpha, n-2)$, $\alpha = 1 - \gamma$, – critical point of the Student's distribution, then the hypothesis is accepted at the level of significance α the hypothesis H_1 is accepted, that is, it is considered that $\alpha_m \neq 0$.

Values S_{a_0} and S_{a_1} we will find using the formulas:

$$S_{a_0}=\sqrt{\frac{S_u^2\overline{x^2}}{n\sigma_x^2}}=\frac{S_u}{\sigma_x}\sqrt{\frac{\overline{x^2}}{n}}=\frac{\sqrt{1{,}65}}{\sqrt{\overline{x^2}-\left(\overline{x}\right)^2}}\sqrt{\frac{\overline{x^2}}{10}}=\frac{\sqrt{1{,}65}}{\sqrt{36{,}17-(5{,}55)^2}}\sqrt{\frac{36{,}17}{10}}=1{,}05$$

;

$$S_{a_1}=\sqrt{\frac{S_u^2}{n\sigma_x^2}}=\frac{S_u}{\sigma_x\sqrt{n}}=\frac{\sqrt{1{,}65}}{\sqrt{10[36{,}17-(5{,}55)^2]}}=0{,}17.$$

Then the empirical values of the criterion:

$$\left|\frac{a_0}{S_{a_0}}\right|=\frac{3{,}02}{1{,}05}=2{,}88, \qquad \left|\frac{a_1}{S_{a_1}}\right|=\frac{1{,}11}{0{,}17}=6{,}53.$$

Critical point for the two-sided critical region

$t_{кр.}=t_{двост.}(\alpha,k)$ at the values $\alpha=0{,}05$, $k=n-2=8$ we find from the table of critical values of the Student's distribution $t_{кр.}=2{,}306$.

Since $2{,}88>t_{кр.}=2{,}306$ and $6{,}53>t_{кр.}=2{,}306$, then at the level of significance $\alpha=0{,}05$ we conclude that $\alpha_0\neq 0$ and $\alpha_1\neq 0$.

5. Confidence intervals with reliability γ for unknown regression parameters a_0 and a_1 have the appearance:

$$a_m-t_m(\gamma,k)S_{a_m}<\alpha_m<a_m+t_m(\gamma,k)S_{a_m},$$

where $m=0{,}1$, $t_m=t_m(\gamma,k)$ – root of equation $P(|t_m|<t)=\gamma$, t_0 and t_1 – random variables distributed according to Student's law.

In our case $\gamma=0{,}95$, the number of degrees of freedom $k=n-2=8$.

We find by the table $t_0(0{,}95;8)=t_1(0{,}95;8)=2{,}306$. Then, taking into account the found values $S_{a_0}=0{,}1055$, $S_{a_1}=0{,}1132$ we will get:

$$3{,}02-2{,}306\cdot 1{,}05<\alpha_0<3{,}02+2{,}306\cdot 1{,}05,$$

$$1{,}11 - 2{,}306 \cdot 0{,}17 < \alpha_1 < 1{,}11 + 2{,}306 \cdot 0{,}17$$

or

$$0{,}6 < \alpha_0 < 5{,}44,$$

$$0{,}72 < \alpha_1 < 1{,}5.$$

6. Coefficient of determination R^2 we find by the formula:

$$R^2 = 1 - \frac{\sum_{i=1}^{n} (y_i - \hat{y}_i)^2}{n\sigma_y^2}.$$

From the table 3 $\sum_{i=1}^{10} (y_i - \hat{y}_i)^2 = 13{,}02$.

Calculate σ_y^2, using the calculations of item 2:

$$\sigma_y^2 = \overline{y^2} - (\overline{y})^2 = 921{,}94/10 - (9{,}18)^2 = 7{,}92.$$

Then

$$R^2 = 1 - \frac{13{,}02}{10 \cdot 7{,}92} = 0{,}836.$$

Thus, the variation of the dependent variable Y on 83,6% explained by the variation of the explanatory variable.

The sample correlation coefficient can be found using the formula:

$$r = \sqrt{R^2} = \sqrt{0{,}836} = 0{,}914.$$

At the same time, the positive sign of this number is chosen due to the fact that $a_1 > 0$.

7. In order to construct a confidence interval for the regression function, it is necessary to place points with coordinates on the coordinate plane $\{x_i; \hat{y}_i - t(\gamma, n-2)S_{\hat{y}_i}\}$, $i = \overline{1,n}$ and connect adjacent (by index) points with straight lines, and then carry out a similar procedure for points $\{x_i; \hat{y}_i + t(\gamma, n-2)S_{\hat{y}_i}\}$.

Let's calculate the value $S_{\hat{y}_i}$ according to the formula:

$$S_{\hat{y}_i} = S_u \sqrt{\left[1 + \frac{(x_i - \overline{x})^2}{\sigma_x^2}\right]\frac{1}{n}}.$$

We will use the value found in point 3

$S_u = \sqrt{1{,}65} = 1{,}28$ та $\sigma_x^2 = \overline{x^2} - (\overline{x})^2 = 36{,}17 - 5{,}55^2 = 5{,}37$, we will get:

$$S_{\hat{y}_1} = 1{,}28\sqrt{\left[1 + \frac{(2{,}3 - 5{,}55)^2}{5{,}37}\right]\frac{1}{10}} = 0{,}698;$$

$$S_{\hat{y}_1} = 1{,}28\sqrt{\left[1 + \frac{(3{,}4 - 5{,}55)^2}{5{,}37}\right]\frac{1}{10}} = 0{,}552;$$

$S_{\hat{y}_3} = 0{,}469$; $S_{\hat{y}_4} = 0{,}518$; $S_{\hat{y}_5} = 0{,}46$; $S_{\hat{y}_6} = 0{,}406$;

$S_{\hat{y}_7} = 0{,}421$; $S_{\hat{y}_8} = 0{,}564$; $S_{\hat{y}_9} = 0{,}462$; $S_{\hat{y}_{10}} = 0{,}846$.

We find by the table $t(0{,}95;8) = 2{,}306$. Using the value $\hat{y}_i$ from table 3 and values $S_{\hat{y}_i}$, we will get the ordinates of the points of the lower limit of the confidence zone:

$$\hat{y}_1 - tS_{\hat{y}_1} = 5{,}58 - 2{,}306 \cdot 0{,}698 = 3{,}97;$$

$$\hat{y}_2 - tS_{\hat{y}_2} = 6{,}8 - 2{,}306 \cdot 0{,}552 = 5{,}52;$$

$$\hat{y}_3 - tS_{\hat{y}_3} = 7{,}68 - 2{,}306 \cdot 0{,}469 = 6{,}6;$$

$$\hat{y}_4 - tS_{\hat{y}_4} = 7{,}13 - 2{,}306 \cdot 0{,}518 = 5{,}93;$$

$$\hat{y}_5 - tS_{\hat{y}_5} = 7{,}79 - 2{,}306 \cdot 0{,}46 = 6{,}73$$

$$\hat{y}_6 - tS_{\hat{y}_6} = 9{,}01 - 2{,}306 \cdot 0{,}406 = 8{,}08;$$

$$\hat{y}_7 - tS_{\hat{y}_7} = 9{,}9 - 2{,}306 \cdot 0{,}421 = 8{,}93;$$

$$\hat{y}_8 - tS_{\hat{y}_8} = 11{,}67 - 2{,}306 \cdot 0{,}564 = 10{,}97;$$

$$\hat{y}_9 - tS_{\hat{y}_9} = 12{,}34 - 2{,}306 \cdot 0{,}462 = 10{,}86;$$

$$\hat{y}_{10} - tS_{\hat{y}_{10}} = 13{,}89 - 2{,}306 \cdot 0{,}846 = 11{,}94.$$

Then the ordinates of the points of the upper limit of the confidence zone take the following:

$$\hat{y}_1 + tS_{\hat{y}_1} = 5{,}58 + 2{,}306 \cdot 0{,}698 = 7{,}18;$$

$$\hat{y}_2 + tS_{\hat{y}_2} = 6{,}8 + 2{,}306 \cdot 0{,}552 = 8{,}07;$$

$$\hat{y}_3 + tS_{\hat{y}_3} = 7{,}68 + 2{,}306 \cdot 0{,}469 = 8{,}76;$$

$$\hat{y}_4 + tS_{\hat{y}_4} = 7{,}13 + 2{,}306 \cdot 0{,}518 = 8{,}32;$$

$$\hat{y}_5 + tS_{\hat{y}_5} = 7{,}79 + 2{,}306 \cdot 0{,}46 = 8{,}86$$

$$\hat{y}_6 + tS_{\hat{y}_6} = 9{,}01 + 2{,}306 \cdot 0{,}406 = 9{,}95;$$

$$\hat{y}_7 + tS_{\hat{y}_7} = 9{,}9 + 2{,}306 \cdot 0{,}421 = 10{,}87;$$

$$\hat{y}_8 + tS_{\hat{y}_8} = 11{,}67 + 2{,}306 \cdot 0{,}564 = 12{,}98;$$

$$\hat{y}_9 + tS_{\hat{y}_9} = 12{,}34 + 2{,}306 \cdot 0{,}462 = 13{,}98;$$

$$\hat{y}_{10} + tS_{\hat{y}_{10}} = 13{,}89 + 2{,}306 \cdot 0{,}846 = 15{,}84.$$

8. The adequacy of the constructed econometric model can be checked using the coefficient of determination. If its value is close to one, it can be considered that the obtained econometric model is adequate. In this case, the change in the value of the outcome variable y depends linearly on the change in the explanatory variable x, and not due to the influence of random factors. If the value of the coefficient of determination is close to zero, then the model is considered inadequate, that is, there is no linear relationship between y and x. If the value of the coefficient of determination is unclear, that is, close to 0.5, then the Fisher F. test is used to check the adequacy of the econometric model.

We calculate the empirical value of the F. parameter using the formula:

$$F_{емп} = \frac{R^2(n-m-1)}{m(1-R^2)} ,$$

where m – the number of independent variables (for simple regression m=1).

After that, we find the value from the table $F_{кр}$ – critical value F-Fisher distribution with з (k_1=m=1, k_2=n-m-1) degrees of freedom and level of significance α. For example, α=0,05, then we can be wrong in 5 % cases, and in 95 % of cases our conclusions will be correct.

If the value calculated by us $F_{емп}>F_{кр}$, then the econometric model built by us is adequate to reality.

In the opposite case, that is, if calculated $F_{емп} \leq F_{кр}$, then in this case the econometric model we built is inadequate to the real reality.

Let's calculate:

$$F_{емп} = \frac{R^2(n-m-1)}{m(1-R^2)} = \frac{0,836\cdot(10-1-1)}{1\cdot(1-0,836)} = 40,78.$$

Let's find the tabular value of this criterion (*$F_{кр.}$)* for the reliability level p=0,95 and the number of degrees of freedom $k_1=m=1$, $k_2=n-m-1=10-1-1=8$:

$$F_{кр} = 5,32.$$

Since $F_{емп.} > F_{кр.}$, then the estimation equation of the econometric model obtained by us

$$\hat{y} = 3,02 + 1,11x$$

adequate to the real reality and on its basis it is possible to make forecasts, i.e. predict the ways of development of the studied phenomena and processes for the near future. The forecast can be point or interval.

Point forecast for the next one n+1 we get the period when we substitute the value of the explanatory variable x_{n+1} into the estimation equation of the econometric model. The forecast value of the gross production output for the cost of the main production assets in the amount of 12 million dollars will be:

$$\hat{y}_{n+1} = 3,02 + 1,1\cdot 12 = 16,22.$$