MINISTRY OF EDUCATION AND SCIENCE OF UKRAINE

SIMON KUZNETS KHARKIV NATIONAL UNIVERSITY OF ECONOMICS

Guidelines to laboratory sessions on the academic discipline "MATHEMATICAL MODELLING IN ECONOMICS AND MANAGEMENT: ECONOMETRICS"

for full-time students of training direction 6.030601 "Management"

Kharkiv
S. Kuznets KhNUE
2016

Затверджено на засіданні кафедри економічної кібернетики.
Протокол № 7 від 14.01.2016 р.

Compiled by: L. Guryanova
S. Prokopovych
O. Sergienko

G 94

Guidelines to laboratory sessions on the academic discipline "Mathematical Modelling in Economics and Management: Econometrics" for full-time students of training direction 6.030601 "Management" / compiled by L. Guryanova, S. Prokopovych, O. Sergienko. – Kharkiv : S. Kuznets KhNUE, 2016. – 68 p. (English)

The basic issues of analysis and forecasting of socio-economic and financial processes and systems through the application of econometric methods and models are presented. Guidelines for doing laboratory work on the academic discipline using the software STATISTICA are provided.

Recommended for students of economic specialities.

General provisions

Laboratory work is intended for students to assimilate the theoretical and practical materials, acquire skills in the use of application packages to ensure the construction and study of different types of models, and expand students' knowledge in the application of mathematical modelling to economic calculation, prediction, and analysis of economic systems.

Statistica 6.0 is proposed to be used for laboratory work. This package contains a set of statistical methods that support solutions to various econometric problems. *Statistica* was developed to work in Windows. The labs were developed on the assumption that students are familiar with the basic principles and methods of work in Windows.

Each laboratory work is considered as an example for solving some problems with detailed comments and pictures. It is recommended that laboratory work should be performed consistently as the steps and techniques are common and will be described only once. In addition, consistent performance helps better learn the material and consolidate the knowledge of the academic discipline.

Laboratory work deals with the main themes and subjects based on the theoretical material of the relevant themes as well as previous issues. Each work contains goals and tasks to be performed and guidelines for doing them.

To confirm the results of the laboratory work students should prepare individual reports that include: the basic data for solving the problem, formulation of the problem, printing the main results of building a model, analysis of calculations and findings. The variant number, the full name of the student who performed the work and the full name of the teacher who collected the report should be indicated on the title page.

The mark for the work depends on the laboratory work and its presentation. Special attention is paid to the knowledge of the theory, correctness and completeness of the findings of the economic interpretation of the results.

Lab session 1. The Simple Linear Regression

The goal is to assimilate the theoretical and practical material and acquire skills in the construction and analysis of simple econometric models in the *Multiple Regression* module.

The task is to find out if there is a linear relation between the respective indices in the *Multiple Regression* module of the application *Statistica*:

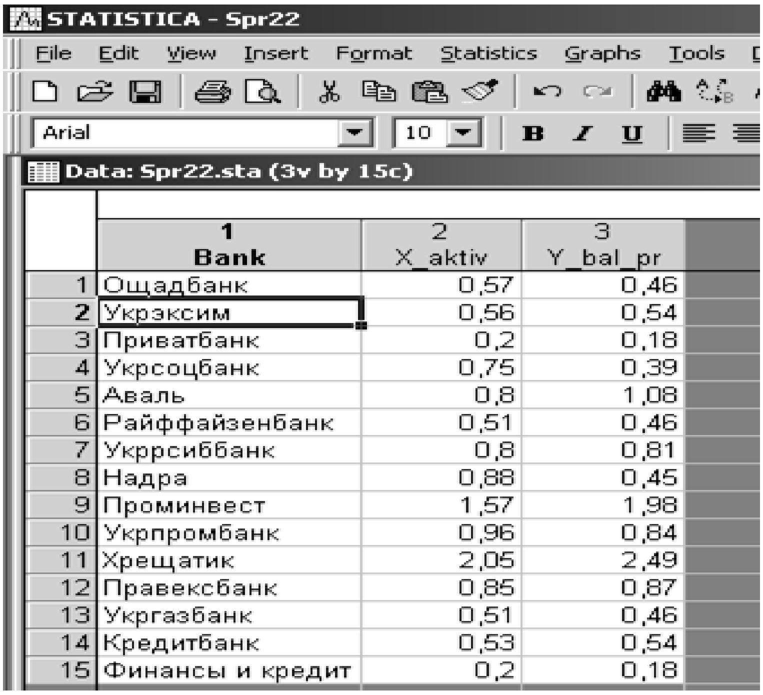
1. Construct a linear econometric model and define all its characteristics (model parameters, standard deviation of the model parameters, variance and standard deviation of models residuals and coefficients of correlation and determination).
2. Check the statistical significance of parameters and the correlation coefficient using the Student's t-test. Check the adequacy of the model by Fisher criterion.
3. Calculate the theoretical value of the dependent variable and model residuals, draw a graph of a linear function with the confidence intervals, construct a histogram and a graph of the distribution of residuals, group the data values of residuals, and interpret the grouping.
4. Calculate the predicted values of the dependent variable and confidence intervals if the value of the independent variable is known.
5. Draw conclusions about the adequacy of the constructed model, provide an economic interpretation of the dependence and possibility of its theoretical use.

Guidelines

Statistica provides a *Multiple Regression* module in order to construct and analyze simple and multiple linear econometric models. Let's consider the algorithm of work in this module.

1. Starting Statistica and preparing data.

Refer to the applications menu and select the program *Statistica*. Then click *File*/*New* in the menu to prepare data. You will see a dialog box in which you must specify the number of variables *(Number of Variables)* and incidence (*Number of Cases*). Click OK. If all the cells of the data are filled you will get a table (Fig. 1.1).



STATISTICA - Spr22

Data: Spr22.sta (3v by 15c)

	1 Bank	2 X_aktiv	3 Y_bal_pr
1	Ощадбанк	0,57	0,46
2	Укрэксим	0,56	0,54
3	Приватбанк	0,2	0,18
4	Укрсоцбанк	0,75	0,39
5	Аваль	0,8	1,08
6	Райффайзенбанк	0,51	0,46
7	Укррсиббанк	0,8	0,81
8	Надра	0,88	0,45
9	Проминвест	1,57	1,98
10	Укрпромбанк	0,96	0,84
11	Хрещатик	2,05	2,49
12	Правексбанк	0,85	0,87
13	Укргазбанк	0,51	0,46
14	Кредитбанк	0,53	0,54
15	Финансы и кредит	0,2	0,18

Fig. 1.1. **The initial data**

2. Calculations.

To start the computational procedures, select the menu item *Statistics / Multiple Regression* (Fig. 1.2). After confirming the selection module you will see the starting panel of the module where you want to set variables for analysis (Fig. 1.3).

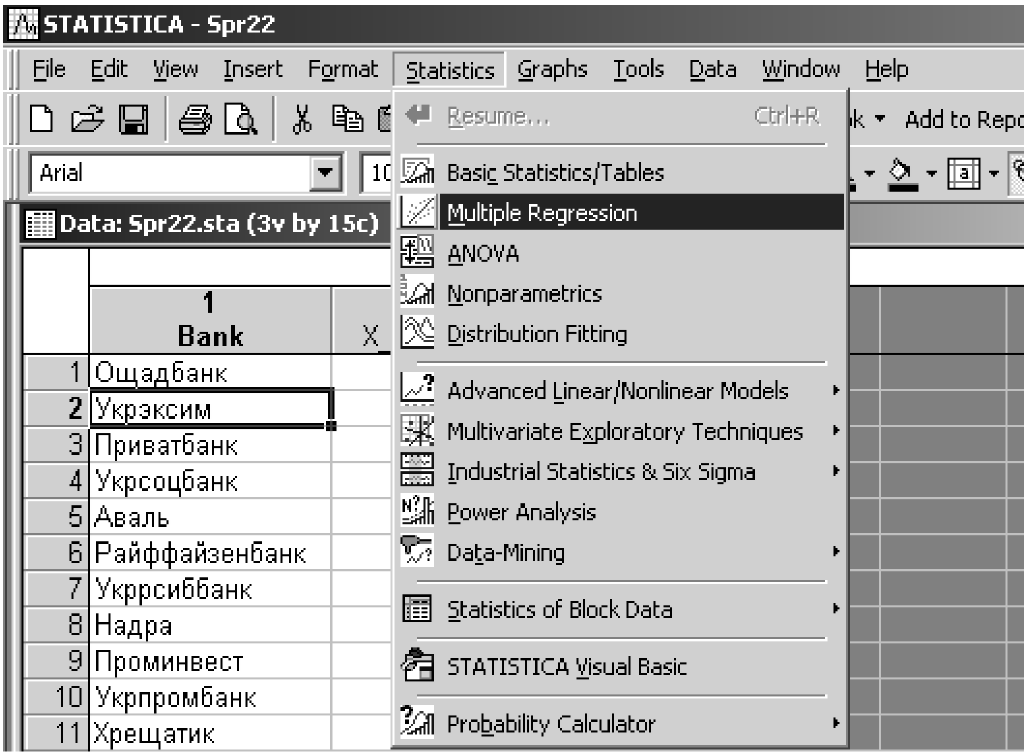


Fig. 1.2. **Module selection**

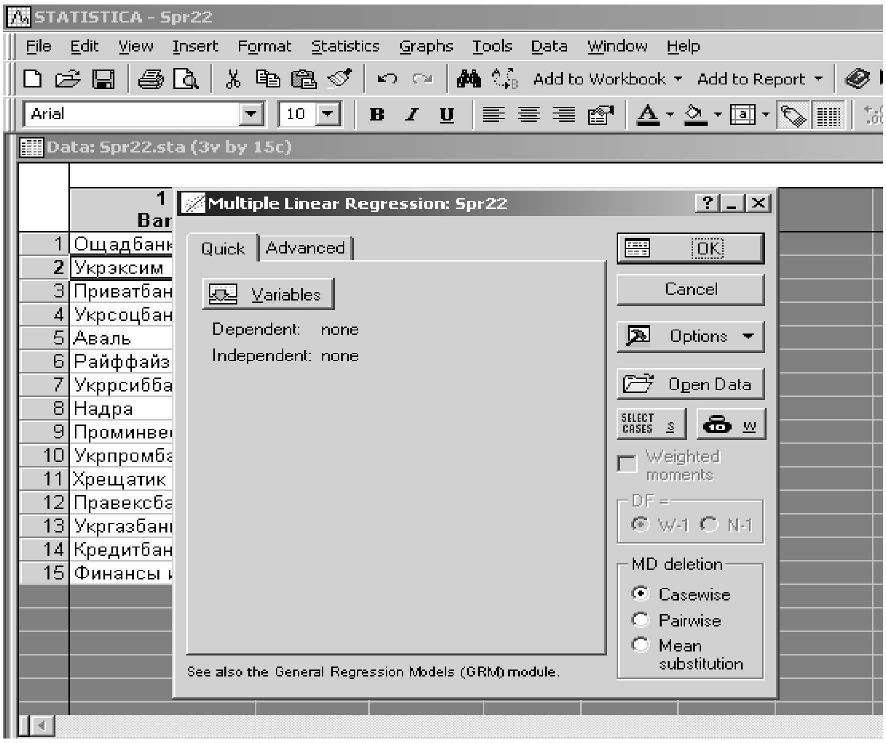


Fig. 1.3. **The module starting panel**

Initiate the button *Variables* and in the window that appears, enter *Dependent* and *Independent* variables to build a simple regression model. The choice of variables is shown in Fig. 1.4.

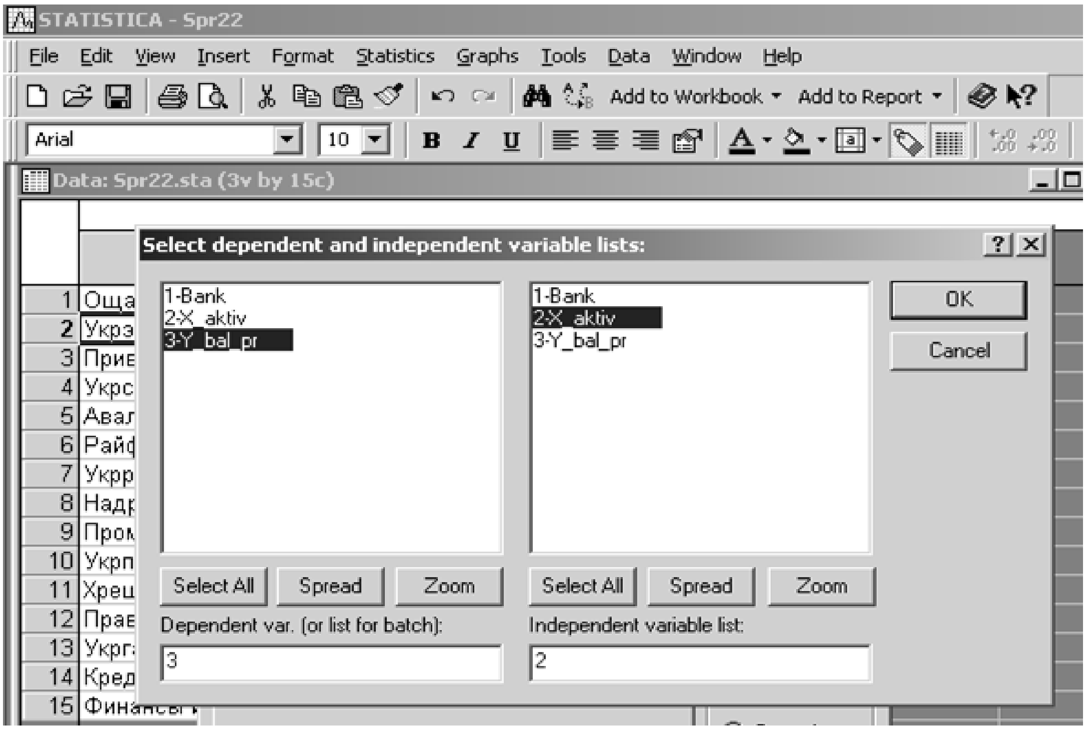


Fig. 1.4. **The choice of variables for analysis**

Confirm your selection by pressing *OK* (Fig. 1.5).

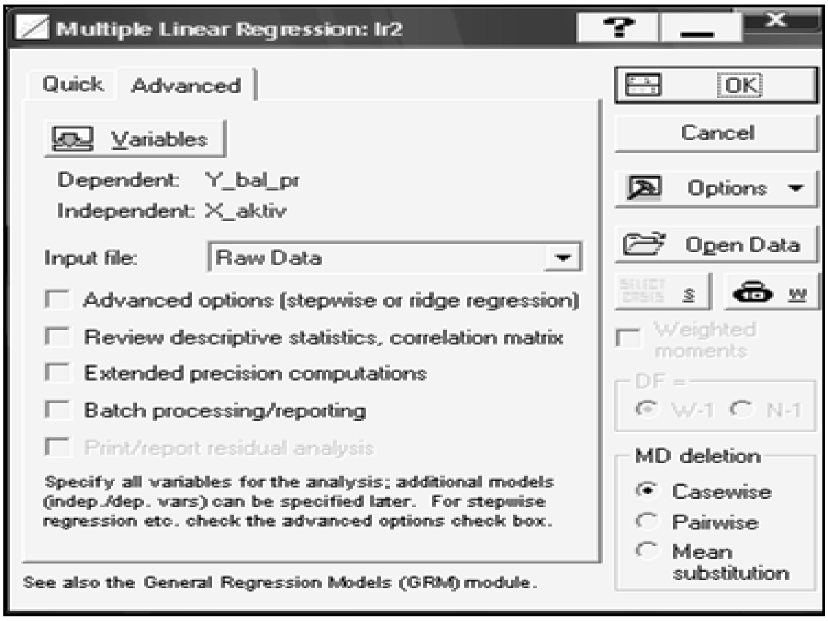


Fig. 1.5. **Confirming the variables**

3. Construction of the model to determine its characteristics, test its adequacy and statistical significance.

Construct a linear econometric model and define all its characteristics. The results of constructing the linear econometric model will be presented in the dialog box (Fig. 1.6). At the top of the window the information about the model is seen. At the bottom there are function buttons that allow you to fully consider the analysis.

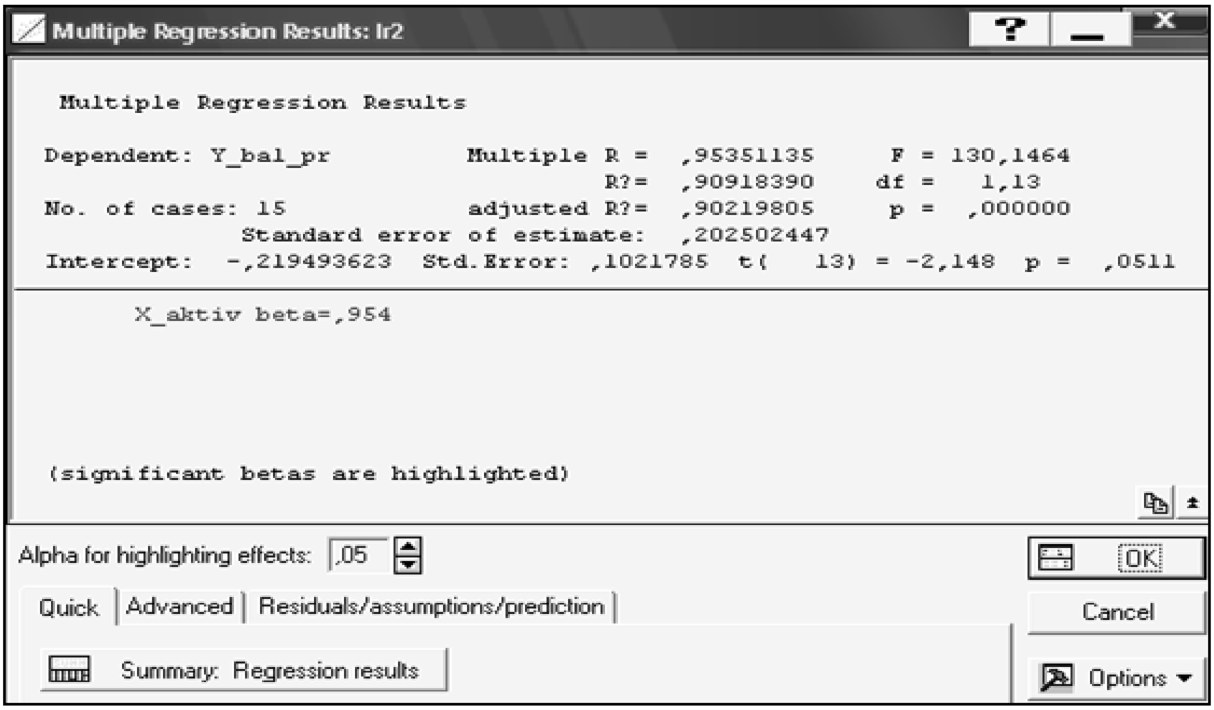


Fig. 1.6. **The window of the regression analysis results**

Initiating the button *Summary: Regression results* (tab *Quick*) (Results of the regression analysis), we define the most important model and the degree of adequacy (Fig. 1.7).

	Regression Summary for Dependent Variable: Y_bal_pr (Spr22) R= ,95351135 R?= ,90918390 Adjusted R?= ,90219805 F(1,13)=130,15 p<,00000 Std.Error of estimate: ,20250					
N=15	Beta	Std.Err. of Beta	B	Std.Err. of B	t(13)	p-level
Intercept			-0,219494	0,102178	-2,14814	0,051130
X_aktiv	0,953511	0,083581	1,279592	0,112164	11,40817	0,000000

Fig. 1.7. **The results of the regression analysis**

Analyze the results of the model:

1) adequacy analysis:

R = 0.9535 is the coefficient of multiple correlation (in the case of a simple linear regression the coefficient is equal to the absolute value of the pair correlation);

R^2 = 0.9091 is the coefficient of the model determination;

Adjusted R^2 = 0.9021 is the adjusted coefficient of determination on the number of observations and the number of model parameters;

F (1,13) = 130.15 is Fisher's criterion of statistical significance model with the number of degrees of freedom and the significance level p;

Std.Error of estimate = 0.2025 is a standard deviation of the model errors; this is statistics that is the measure of screening the values studied relative to the regression line;

2) analysis of parameters and their statistical significance:

Beta (*a*1) = (0.9535) is the standardized meaning of the regression (weight) coefficients which is measured based on the standardized data with a selective mean equal to zero and a standard deviation equal to one;

Std.Error of Beta (*a*1) = 0.08358 is a standard deviation of the model parameters for standardized regression coefficients;

B (*a*1, *a*2) = (-0.2194, 1.2795) are nonstandardized parameters of the model, so the model has the form:

$$Y = -0.2194 + 1.2796 \cdot x;$$

Std.Error of B = (0.10; 0.11) is a standard deviation of the model parameters;

t (13) = (-2.14; 11.4) is the significance of parameters based on Student's test;

p-level = (0.051; 0.000) is the Student's t-test significance level.

4. Constructing a graph of a linear function with the confidence intervals.

To do this, click *Graphs* / *Scatterplots* and specify the variables, the level line and the confidence intervals (Fig. 1.8).

Initiating *OK*, we get the following graph (Fig. 1.9).

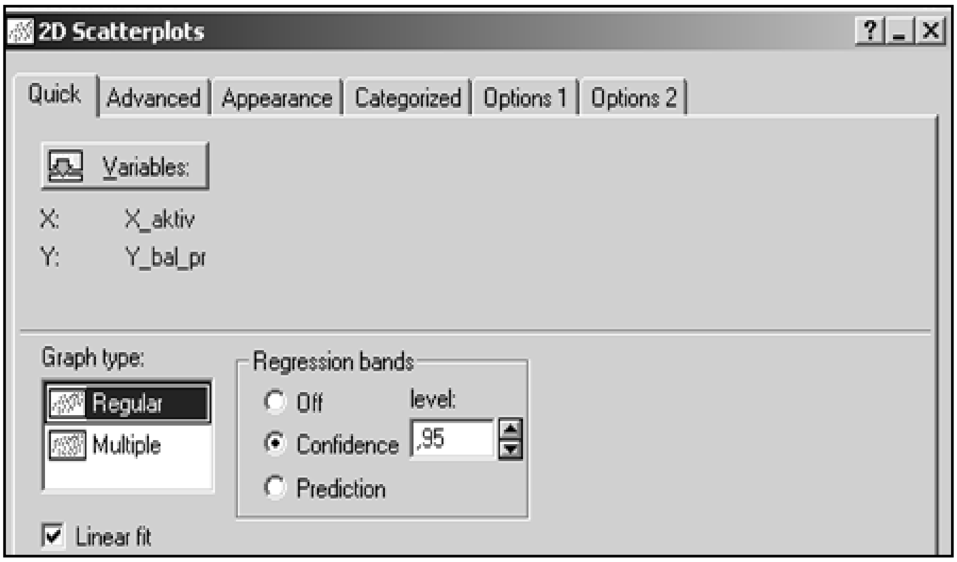


Fig. 1.8. **Indicating the graph options**

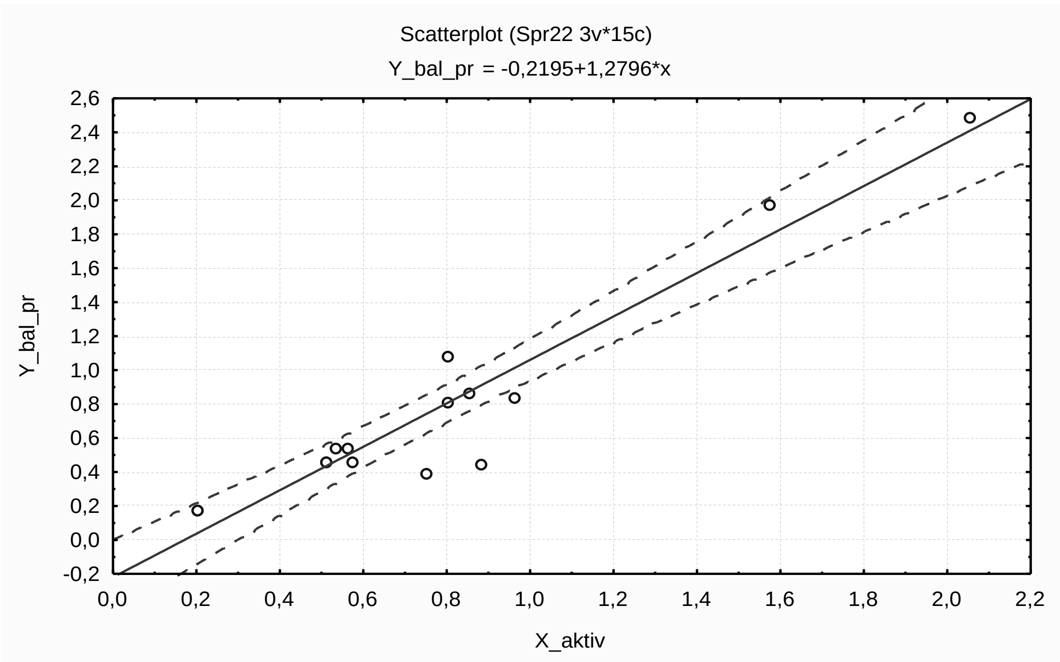


Fig. 1.9. **The graph of the linear relationship**

To test the hypothesis about the significance of the regression model use the analysis of variance. Initiate the button *Advanced* / *ANOVA* at the bottom of the information window (Fig. 1.10).

The results of the analysis of variance for the model under study are shown in Fig. 1.11. This table shows the sum of the squared deviations for regression (*Sums of Squares Regress*), the sum of the squared deviations of errors (*Sums of Squares Residual*), the variance of the errors (*Mean Squares Residual*) and Fisher's criterion.

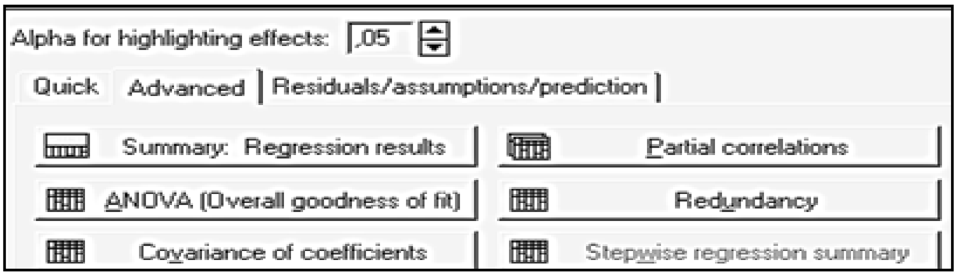


Fig. 1.10. **Choosing the ANOVA window**

Effect	Analysis of Variance; DV: Y_bal_pr (Spreads				
	Sums of Squares	df	Mean Squares	F	p-level
Regress.	5,336946	1	5,336946	130,1464	0,000000
Residual	0,533094	13	0,041007		
Total	5,870040				

Fig. 1.11. **The table of ANOVA**

To calculate the basic characteristics (the mean and standard deviation) of the dependent and independent variables and the matrix of the pair correlation coefficients initiate the option *Means & standard deviations and correlations*, which are found in the menu *Descriptive statistics* (tab *Advanced*) (Fig. 1.12).

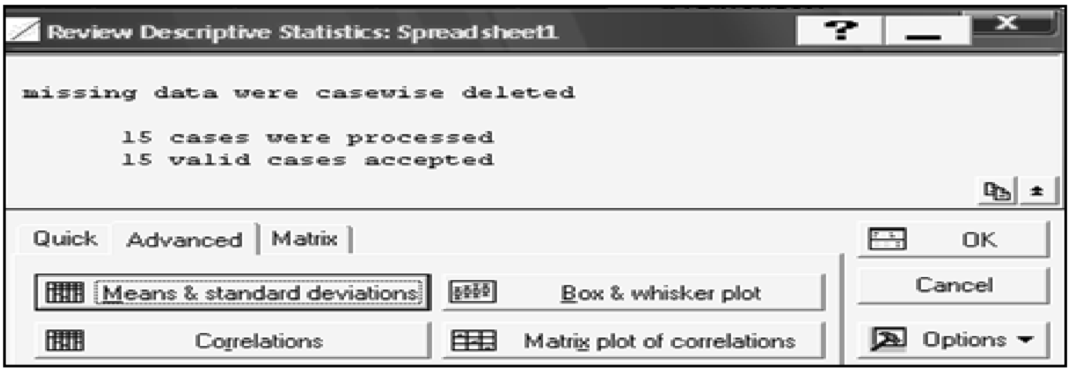


Fig. 1.12. **The window of selection of the basic statistics calculation**

The basic statistics and the pair correlation coefficient matrix are shown in Fig. 1.13.

Variable	Means and Standard Dev		
	Means	Std.Dev.	N
X_aktiv	0,782667	0,482515	15
Y_bal_pr	0,782000	0,647525	15

Variable	Correlations (lr2)	
	X_aktiv	Y_bal_pr
X_aktiv	1,000000	0,953511
Y_bal_pr	0,953511	1,000000

Fig. 1.13. **The basic statistics and the pair correlation coefficient matrix**

4. Analysis of residuals.

Calculate the theoretical values of the dependent variable and the residuals of the model. Construct a histogram and a graph of the distribution of errors on the normal probability paper.

To calculate and analyze the residuals initiate the option *Perform residual analysis* (comprehensive analysis of residuals) at the bottom of the information window showing the results of regression analysis (Fig. 1.14). By initiating this option, we will get a menu for the model error analysis (Fig. 1.15).

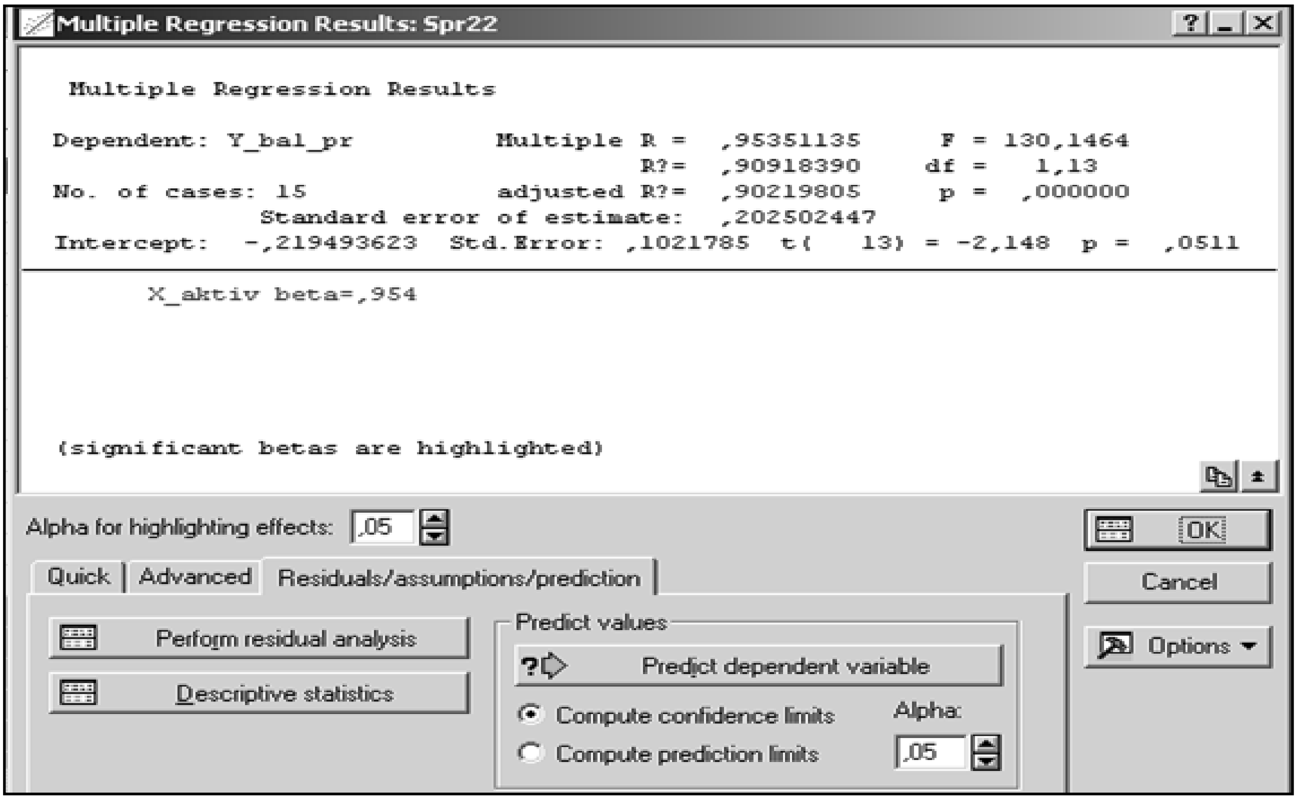


Fig. 1.14. **The window of residual analysis**

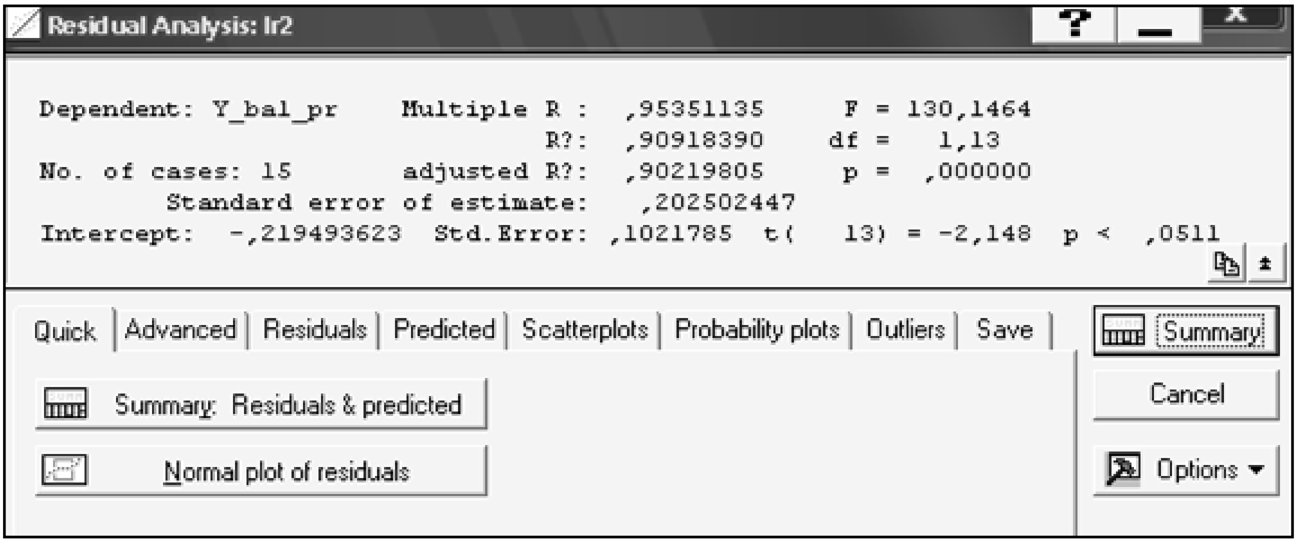


Fig. 1.15. **The residual analysis menu**

The graph of screening the model errors in the range $\pm 3\sigma$ can be obtained by initiating the option *Residuals* / *Casewise plot of residuals* (Fig. 1.16). This graph analyzes sustainability of the error variance.

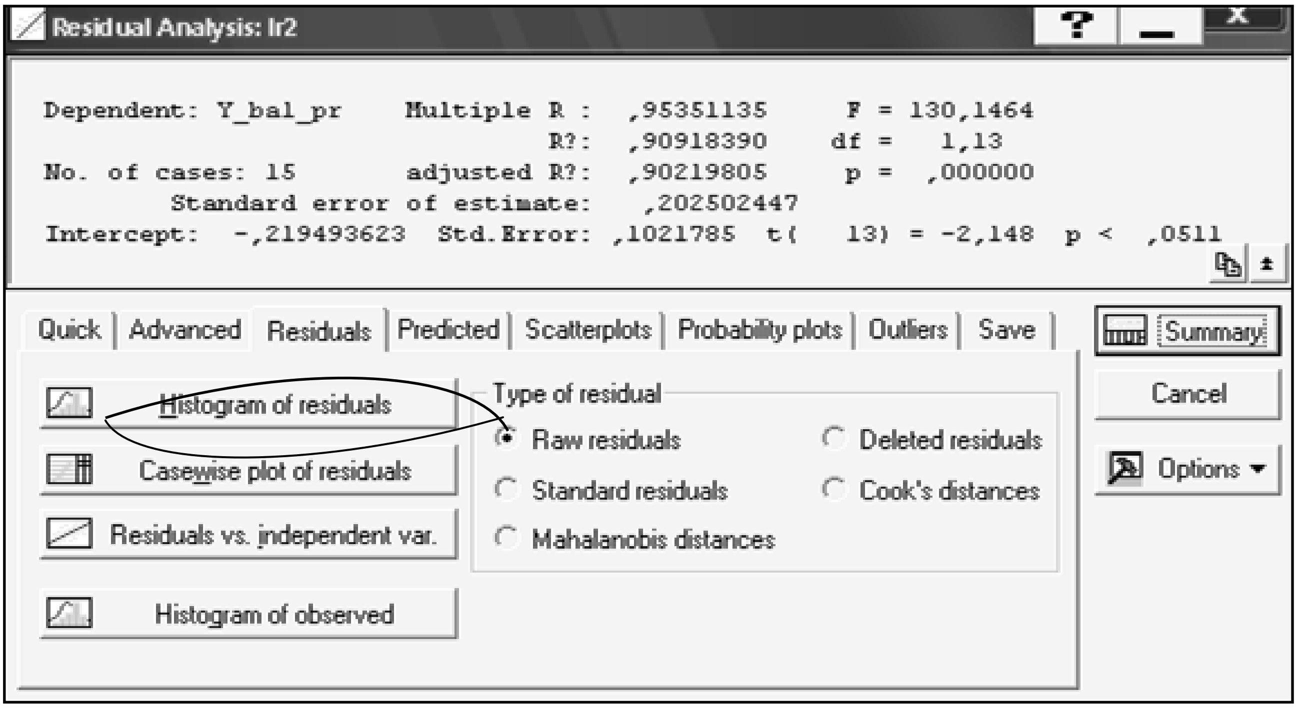


Fig. 1.16. **Analysis of the model residual variance**

The button of the error analysis *Quik* / *Summary: Residuals & Predicted* reflects the observed values of the dependent variable (*Observed value*), the theoretical value of the dependent variable (*Predicted value*) and the model errors (*Residual*) as the difference between the observed and theoretical values (Fig. 1.17).

Case	Raw Residuals -3s . . 0 . . +3s	Raw Residual (Spreadsheet1) Dependent variable: Y_bal_pr Observed Value	Predicted Value	Residual
1	. . . *. . .	0,460000	0,509873	-0,049874
2	. . . .* . .	0,540000	0,497078	0,042922
3	. . . . *. .	0,180000	0,036425	0,143575
4	. . *	0,390000	0,740200	-0,350200
5	. . . . . * .	1,080000	0,804180	0,275820
6	. . . .* . .	0,460000	0,433098	0,026902
7	. . . * . .	0,810000	0,804180	0,005820
8	. *.	0,450000	0,906547	-0,456547
9	. . . . * .	1,980000	1,789465	0,190535
10	. . .* . . .	0,840000	1,008914	-0,168914
11	. . . . * . .	2,490000	2,403669	0,086331
12	. . . * . .	0,870000	0,868159	0,001841
13	. . . .* . .	0,460000	0,433098	0,026902
14	. . . . * . .	0,540000	0,458690	0,081310
15	. . . . *. .	0,180000	0,036425	0,143575
Minimum	. *.	0,180000	0,036425	-0,456547
Maximum	. . . . . * .	2,490000	2,403669	0,275820
Mean	. . . * . .	0,782000	0,782000	0,000000
Median	. . . .* . .	0,540000	0,740200	0,026902

Fig. 1.17. **Analysis of the model residuals**

To test the hypothesis of normality of the error distribution, plot the distribution of errors on the normal probability paper (*Normal plot of residuals*), which is shown in Fig. 1.18.

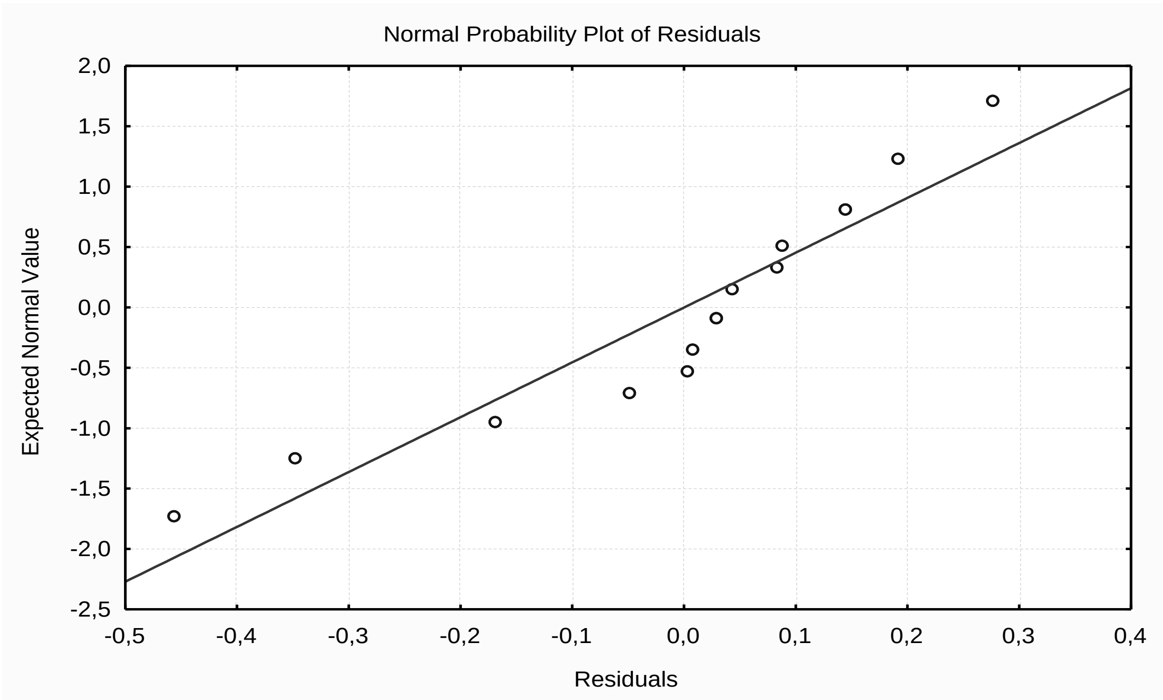


Fig. 1.18. **The graph of residual distribution on the normal probability paper**

Since one of the main hypotheses of the random variable says that errors should be normally distributed according to the normal law, let's present a histogram of error distribution (*Residuals* / *Histogram plot of residuals*) and analyze it (Fig. 1.19). The graph shows that the error distribution curve does not meet the normal law, apparently because of the small number of observations.

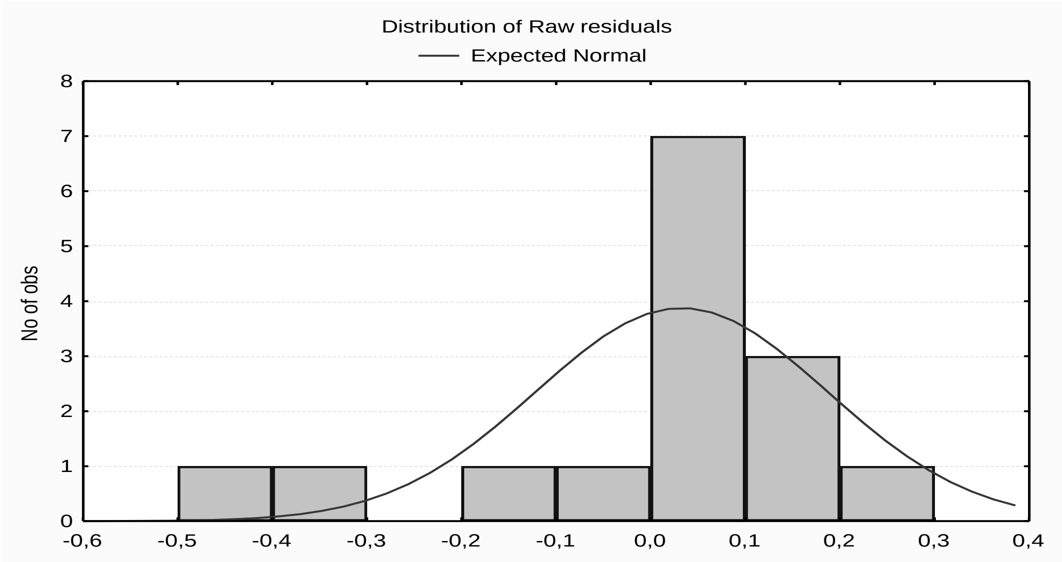


Fig. 1.19. **The histogram of the distribution of residuals**

5. Calculation of the predictive values of the dependent variable and the confidence interval change.

The model is appropriate, its parameters are significant, then the model can be used to make a forecast. To calculate the predicted values of the dependent variable use the option *Predict dependent variable* at the bottom of the results of regression analysis (Fig. 1.14). By initiating this option, you must specify the independent variable for which it is necessary to predict the dependent value (Fig. 1.20).

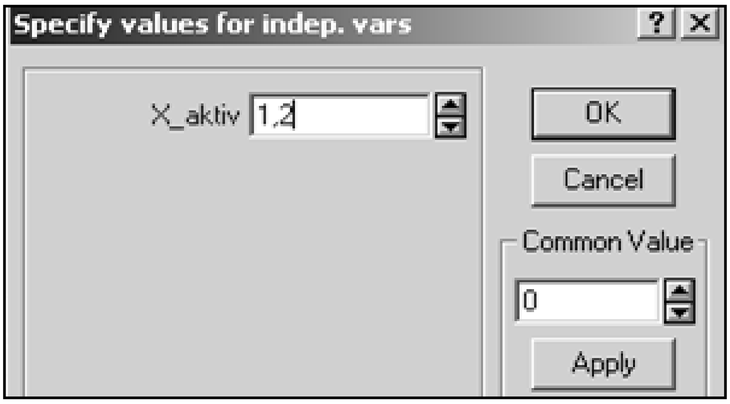


Fig. 1.20. **The value of the independent variable**

The predicted results are presented in the table showing the model coefficient and the calculation procedure (Fig. 1.21).

Variable	Predicting Values for (Spr22) variable: Y_bal_pr		
	B-Weight	Value	B-Weight * Value
X_aktiv	1,279592	1,200000	1,535510
Intercept			-0,219494
Predicted			1,316016
-95,0%CL			1,164405
+95,0%CL			1,467627

Fig. 1.21. **The forecast results**

The predicted value of the dependent variable (*Predicted*) = 1,3160; and the confidence intervals for the predicted value:

$$1,164405 \leq 1,316016 \leq 1,467627.$$

A comprehensive analysis of a one-factor linear econometric model of the balance profit dependence of Ukraine's largest banks on the net asset value of their work has been done.

Lab session 2. The Multiple Linear Regression Model and Multicollinearity

The goal is to consolidate the theoretical and practical material on the theme "Multiple regression", acquire skills in the construction and analysis of multivariate econometric models in the *Multiple Regression* module.

The task is find out if there is a multiple linear relation between the indicators in the *Multiple Regression* module (*Statistica*):

Part 1

1. Construct a linear multifactor econometric model (include all the relevant factors) and define all its characteristics (model parameters, standard deviation of the model parameters, variance and standard deviation of the model errors, coefficients of multiple correlation and determination).

2. Check the statistical significance of the model parameters, the multiple correlation coefficient. Check the adequacy of the model using the Fisher criterion.

3. Provide a table with theoretical values of the dependent parameter and the values of the model errors. Construct a graph of the linear function with confidence intervals. Find the predictive value of the dependent variable and confidence intervals if the data about the future values of independent parameters is known.

Part 2

4. Check the model for multicollinearity. Present the matrix of pair correlations for the factor variables. Evaluate the significance of multicollinearity according to the method Farrar – Glauber.

5. Construct a histogram and a graph of the distribution of errors. Present the grouping of data values of errors, make an economic interpretation.

6. Draw conclusions about the adequacy of the constructed multivariable model; make an economic interpretation of the overall model.

7. Remove from the model the factors that least affect the dependent variable or interdependent variables (using Student's t test results and the pair correlation coefficients). Identify all of the above features of the built models to draw conclusions as to their adequacy.

8. Construct and interpret the models based on the methods of incremental inclusion and incremental exclusion of variables.

9. If there is multicollinearity in the model, the parameters should be assessed by the Ridge regression method. Identify all of the model properties. Provide the graphs of the Ridge model estimated parameter value change depending on the parameter value. Assess the degree of displacement of parameter estimates.

10. Make a comparative analysis of the designed models. Identify the most appropriate and cost-interpreted model.

Guidelines

Part 1

1. Running Statistica and preparing data.

Select the program *Statistica* in the application menu, then select *File / New* to prepare the data. After entering the data, click *OK*. After filling in all the cells of the data field, we will obtain a table similar to that shown in Fig. 2.1, where *X1, X2, X3* are independent variables, *Y* is the resulting dependent variable.

	1 X1	2 X2	3 X3	4 Y
1	0,45	0,83	0,69	0,13
2	0,81	0,8	0,83	0,77
3	0,46	0,66	0,16	0,25
4	1,08	1,11	0,65	0,5
5	0,39	0,67	0,34	0,12
6	0,18	0,19	0,82	0,22
7	0,54	0,35	5,36	0,82
8	0,46	0,23	0,54	0,91
9	0,87	0,81	1,94	0,96
10	0,71	0,9	0,22	0,16
11	0,07	0,13	0,83	0,11
12	0,14	0,24	0,35	0,02
13	0,03	0,05	0,66	0,01
14	0,136	0,176	0,18	0,043
15	0,073	0,064	0,16	0,969
16	0,105	0,066	0,65	0,172
17	0,15	0,182	0,34	0,108
18	0,189	0,359	0,82	0,064
19	9,155	1,234	0,79	4,654
20	4,744	0,339	0,43	0,076
21	63,898	16,332	0,15	25,152
22	7,314	1,004	0,03	1,087
23	0,85	0,453	0,126	0,024
24	0,184	2,635	0,445	0,914
25	0,01	0,031	0,087	0,001

Fig. 2.1. **The output**

2. Construction of multiple linear econometric models.

To start the computational procedures, choose the menu option *Statistics* / *Multiple Regression*. After confirming the selection of the module, you will see the starting panel of the module where you have to set the variables for analysis.

Initiate the button *Variables* and in the window that appears, enter *Dependent* and *Independent* variables to build a multiple regression model. The choice of the variables is presented in Fig. 2.2. After specifying the variables, confirm your selection by pressing the *OK* button.

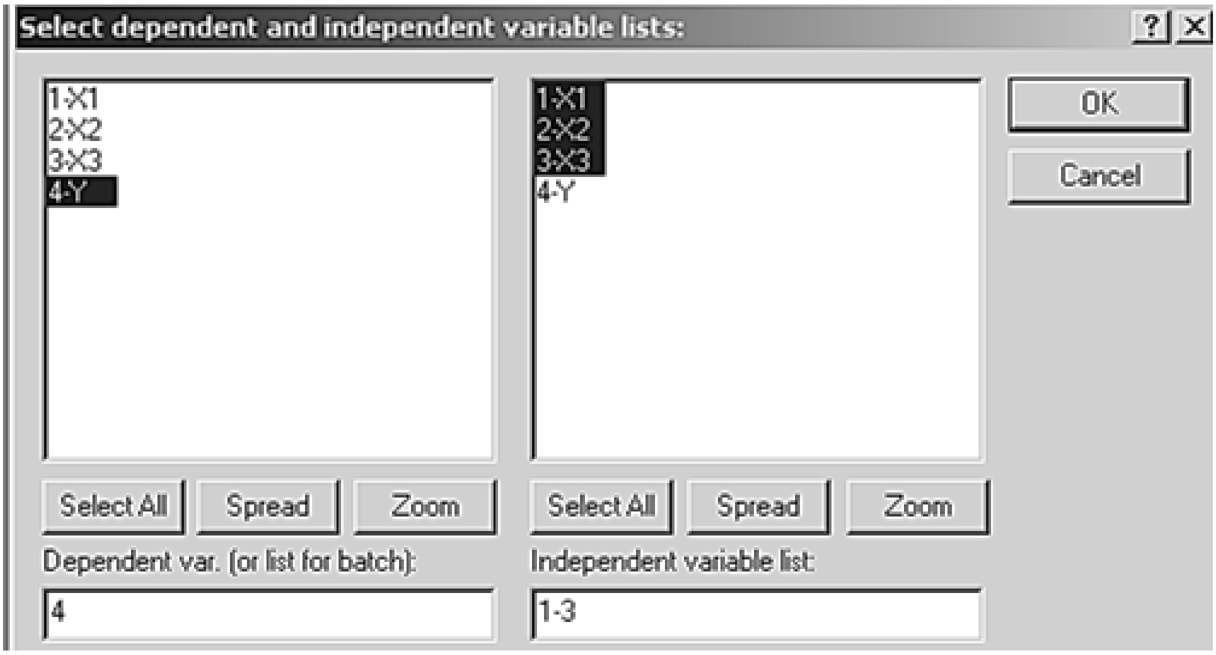


Fig. 2.2. **The choice of variables for analysis**

Construct a linear multifactor econometric model and define all its characteristics. Use the Fisher criterion to verify the statistical significance of the model parameters and the model adequacy. The results of the model construction is shown in Fig. 2.3.

	Regression Summary for Dependent Variable: Y (Spr38) R= ,99329417 R?= ,98663330 Adjusted R?= ,98472378 F(3,21)=516,69 p<0,0000 Std.Error of estimate: ,61913					
N=25	Beta	Std.Err. of Beta	B	Std.Err. of B	t(21)	p-level
Intercept			-0,159728	0,165543	-0,964874	0,345589
X1	0,747946	0,119810	0,293843	0,047069	6,242776	0,000003
X2	0,253338	0,119732	0,396141	0,187223	2,115880	0,046472
X3	0,036745	0,025417	0,175450	0,121359	1,445707	0,163019

Fig. 2.3. **The results of the multivariate regression analysis**

To test the hypothesis about the significance of the regression model, use the analysis of variance. Initiate the button *Advanced*/*ANOVA*. The results of the analysis of variance for the model under study are shown in Fig. 2.4.

This table shows the sum of squared deviations for the regression (*Sums of Squares/Regress.*), the sum of the squared deviations of errors (*Sums of Squares/Residual*), the variance of the error (*Mean Squares/Residual*) and Fisher criterion (*F*).

Effect	Analysis of Variance; DV: Y (Spreadsheet17) Sums of Squares	df	Mean Squares	F	p-level
Regress.	591,6353	1	591,6353	1283,945	0,000000
Residual	10,5983	23	0,4608		
Total	602,2336				

Fig. 2.4. **The table of ANOVA**

Part 2

3. Checking the model for multicollinearity.

The next step in studying the multivariable regression model aims to verify the multicollinearity of the model.

One way to check the model for multicollinearity is to calculate the matrix of pairwise correlations. Initiate the button *Descriptive statistics* / *Correlations* in the menu of the model analysis (Fig. 2.5). The matrix of the pair correlation coefficients is shown in Fig. 2.6.

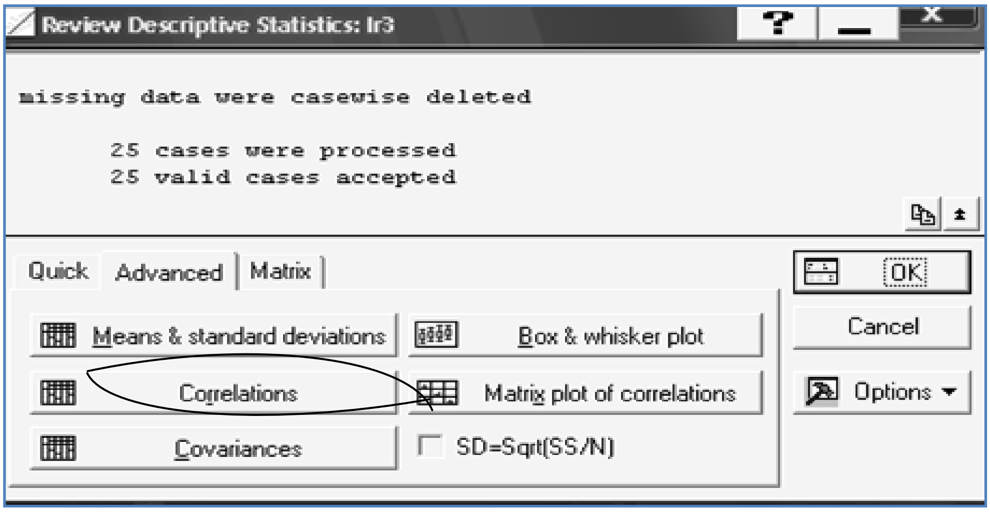


Fig. 2.5. **The window of selection of the correlation dependence assessment**

Variable	Correlations (Spr38) X1	X2	X3	Y
X1	1,000000	0,977544	-0,120649	0,991162
X2	0,977544	1,000000	-0,115205	0,980255
X3	-0,120649	-0,115205	1,000000	-0,082679
Y	0,991162	0,980255	-0,082679	1,000000

Fig. 2.6. **The matrix of the pair correlation coefficients**

Initiating the option *Matrix plot of correlations* (correlation graphs), we obtain histograms and charts of screening of the studied variables in the model (Fig. 2.7).

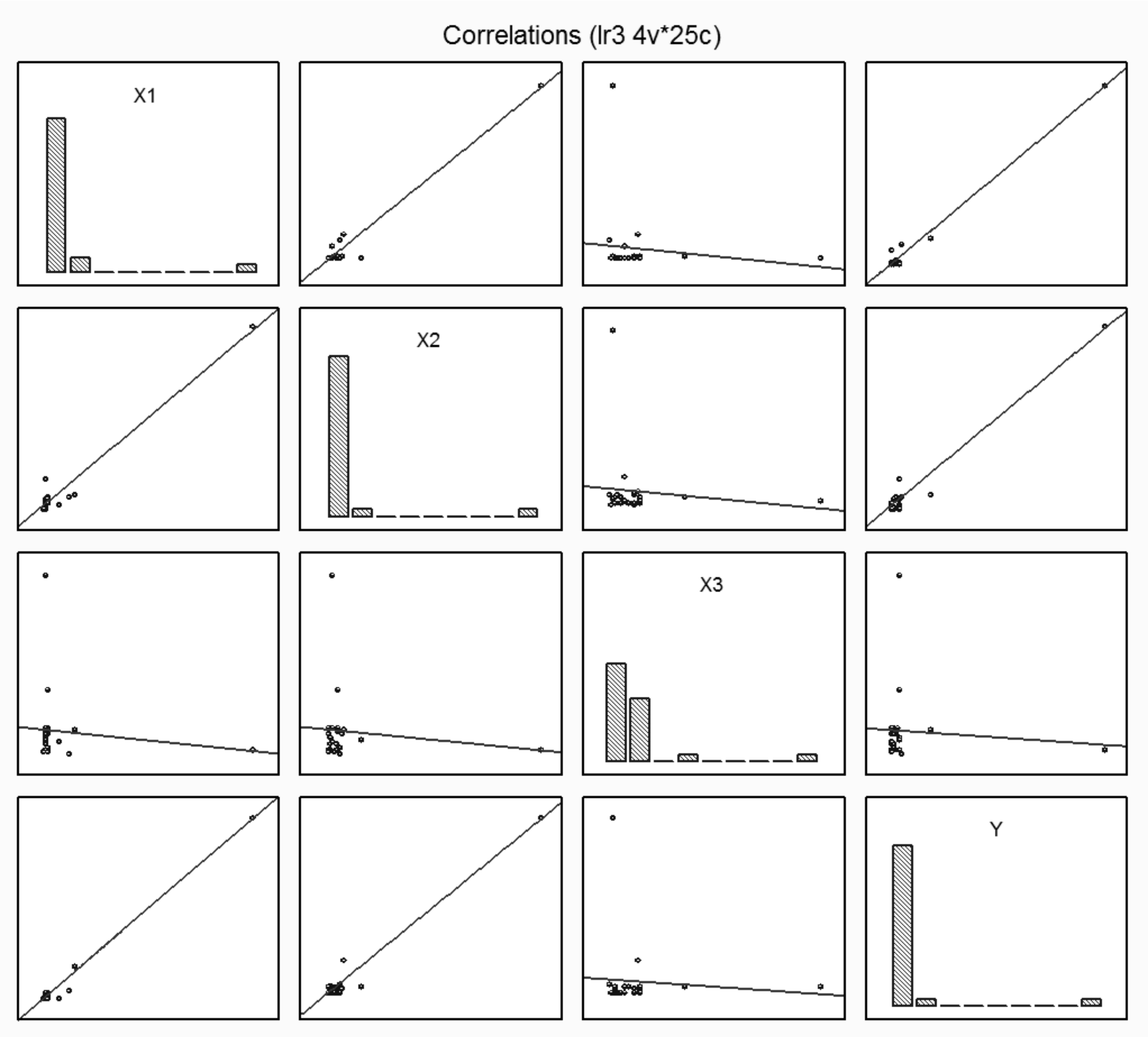


Fig. 2.7. **Bar graphs and charts of screening of the variables studied**

Pair correlation coefficients and charts of screening indicate a strong degree of linear relationship between the pairs of variables studied.

Further analysis involves the calculation of the partial correlation and calculation of the criteria of maximum contingency for each independent variable. Initiating the option *Partial correlations* (Fig. 2.8), we obtain the following table (Fig. 2.9). This table contains the values of standardized regression coefficients (*Beta in*); *Partial correlations* reflecting the impact of each independent variable on the resulting one, provided that other variables do not affect the given relation; semipartial correlations; the coefficient of determination (*R-square*) between this variable and other independent variables included in the regression equation that shows the maximum extent of conjugation; model tolerance calculated as (1 - R square); the value of Student's t-test (t(21)) to test the hypotheses about the significance of partial correlation coefficients with the number of degrees of freedom; the significance level (*p-level*); the probability of rejection of the hypothesis of the importance of partial correlation coefficients.

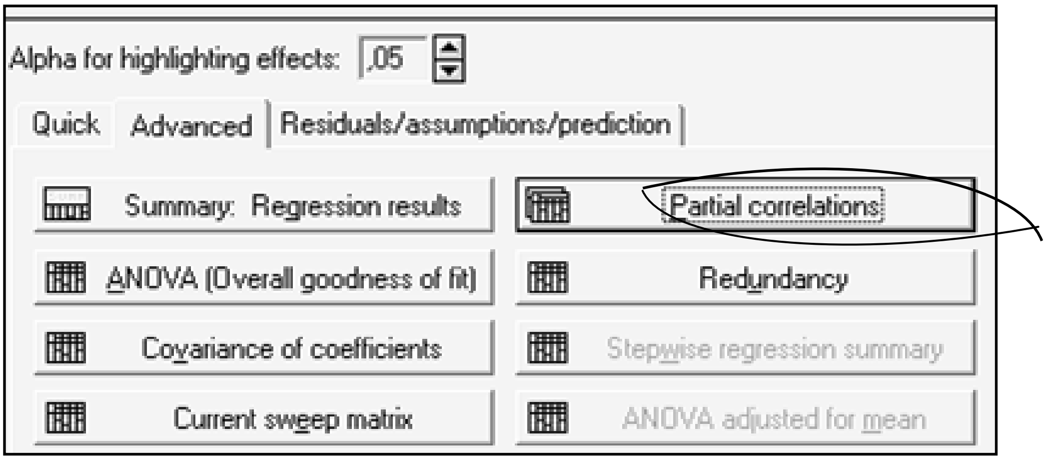


Fig. 2.8. **The selection window for assessing the significance of independent variables**

Variable	Variables currently in the Equation; DV: Y (Spreadsheet17)						
	Beta in	Partial Cor.	Semipart Cor.	Tolerance	R-square	t(21)	p-level
X1	0,747946	0,806125	0,157500	0,044342	0,955658	6,242776	0,000003
X2	0,253338	0,419196	0,053382	0,044400	0,955600	2,115880	0,046472
X3	0,036745	0,300862	0,036474	0,985276	0,014724	1,445707	0,163019

Fig. 2.9. **The significance of variables in the regression equation**

Evaluation of the impact of independent variables on the resulting value may be performed by a keystroke *Redundancy*, the results of which are shown in Fig. 2.10.

Variable	Redundancy of Independent Variables; DV: R-square column contains R-square of resp variable with all other independent variables			
	Toleran.	R-square	Partial Cor.	Semipart Cor.
X1	0,044342	0,955658	0,806125	0,157500
X2	0,044400	0,955600	0,419196	0,053382
X3	0,985276	0,014724	0,300862	0,036474

Fig. 2.10. **Assessment of excessively independent variables**

To assess the degree of multicollinearity based on the algorithm of Farrar – Glauber, partial coefficients of correlation between the factor variables and their statistical significance are used. To calculate them it is necessary to investigate the model without the dependent variable by making dependent any of the factor variables, and identify the data characteristics (Fig. 2.11).

Variable	Variables currently in the Equation; DV: X1(lr3)						
	Beta in	Partial Cor.	Semipart. Cor.	Tolerance	R-square	t(22)	p-level
X2	0.976606	0.977242	0.970104	0.986728	0.013272	21.60825	0.000000
X3	-0.008139	-0.038366	-0.008085	0.986728	0.013272	-0.18008	0.858735

Variable	Variables currently in the Equation; DV: X2(lr3)						
	Beta in	Partial Cor.	Semipart. Cor.	Tolerance	R-square	t(22)	p-level
X2	0.977879	0.977242	0.970736	0.985444	0.014556	21.60825	0.000000
X3	0.002775	0.013073	0.002755	0.985444	0.014556	0.06132	0.951656

Fig. 2.11. **Evaluation of the relation between independent variables**

Thus, the partial correlation coefficient value is:

$$r_{12} \approx 0.977; r_{13} \approx -0.038;\ r_{23} \approx 0.013.$$

The significance of the partial correlation coefficients is calculated using Student's t-test:

$$t_{12} \approx 21.608;\ t_{13} \approx -0.18;\ t_{23} \approx 0.061.$$

The value of the criteria t_{kj} is compared with the tabular values with $(n-m)$ degrees of freedom and significance level α. If $t_{kj} > t_{tab}$, then multicollinearity exists between the independent variables x_k and x_l.

$t_{tab}(22; 0.05) = 2.07$. Therefore $t_{12} > t_{tab}$, so we can conclude that there is a close linear relationship (multicollinearity) between the variables x_1 and x_2.

4. Calculation of errors and research on the model.

For further comprehensive analysis of errors build a histogram and a graph of the distribution of errors on the normal probability paper (Fig. 2.12 and 2.13).

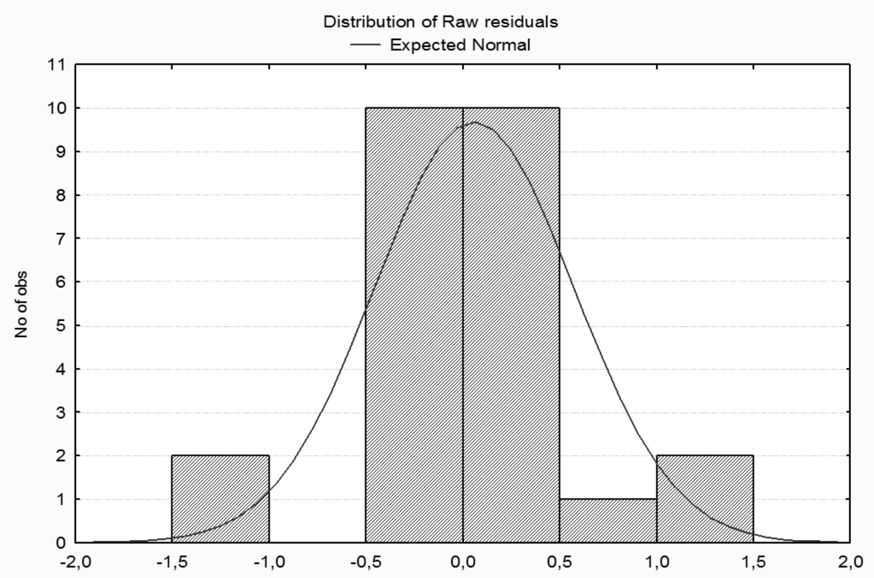


Fig. 2.12. **The histogram of the distribution of errors**

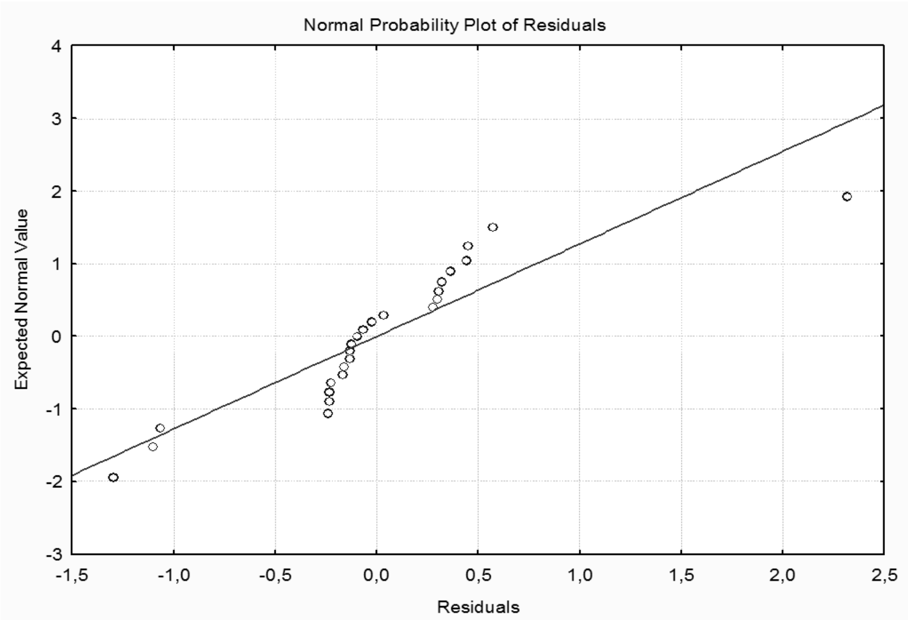


Fig. 2.13. **The graph of the distribution of errors on the normal probability paper**

According to the results it is necessary to group the objects for positive and negative values of errors, and make an economic interpretation of this grouping.

5. Construction of models based on the methods of incremental inclusion and incremental exclusion of independent variables.

In terms of the independent variables multicollinearity, an effective method for estimating the parameters of econometric models is the implementation of incremental regression, which involves the evaluation of the model parameters via the coefficients of correlation.

The *Multiple Regression* module was used to implement the method of incremental inclusion of variables (*Forward stepwise*) and the stepwise exclusion method (*Backward stepwise*). The choice of methods is carried out on the start menu bar by the initiation of *Advanced options* (*stepwise or ridge regression*) (Fig. 2.14).

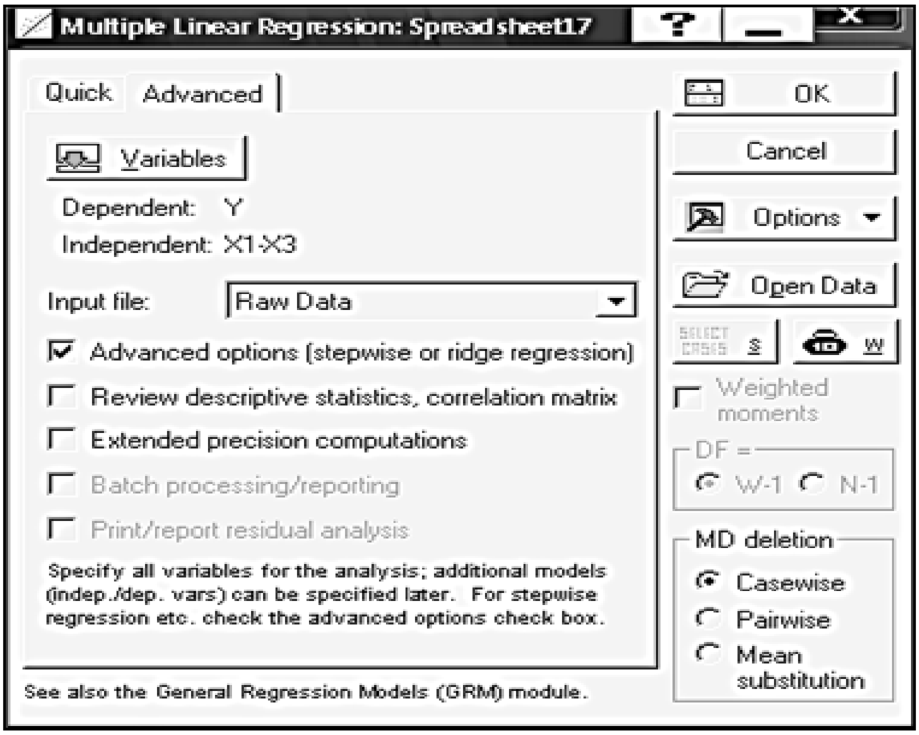


Fig. 2.14. **The choice of the method of stepwise regression**

The choice of the evaluation method, the threshold values of F-inclusion or exclusion criteria, consistency in presenting the results are selected in the tab *Stepwise* (Fig. 2.15).

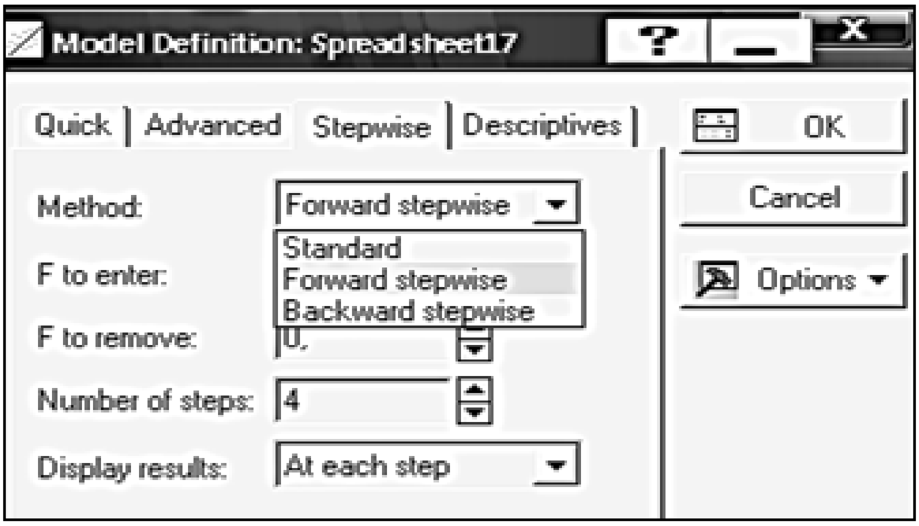


Fig. 2.15. **The choice of the methods for parameter evaluation**

The sequence of steps of the algorithm of stepwise inclusion (*Forward stepwise*) is shown in Fig. 2.16.

```
                Multiple Regression Results (Step 0)
Dependent: Y                  Multiple R = 0,00000000    F = 0,000000
R? = 0,00000000    df = 0,24
  No. of cases: 25        adjusted R? = 0,00000000     p = -0,00000
             Standard error of estimate: 5,009297816
     Step 0:  No variables in the regression equation

                 Multiple Regression Results (Step 1)
  Dependent: Y                Multiple R =  ,99116180      F = 1283,945
R? =  ,98240171    df = 1,23
  No. of cases: 25          adjusted R? =  ,98163656    p = 0,000000
               Standard error of estimate:  ,678818769
Intercept:  ,081243426  Std. Error: ,1416544  t(23) = ,57353  p = ,5719
             X1 beta =,991

                 Multiple Regression Results (Step 2)
  Dependent: Y                 Multiple R =  ,99262428     F = 737,4498
R? = ,98530296    df = 2,22
  No. of cases: 25          adjusted R? =  ,98396686    p = 0,000000
               Standard error of estimate:  ,634287475
Intercept: -,030737796  Std. Error: ,1428533 t(22) = -,2152  p = ,8316
           X1 beta = ,741            X2 beta = ,256

          Multiple Regression Results (Step 3, final solution)
No other F to enter exceeds specified limit
  Dependent: Y                 Multiple R =  ,99329417     F = 516,6896
R? =  ,98663330    df = 3,21
  No. of cases: 25          adjusted R? =  ,98472378    p = 0,000000
               Standard error of estimate:  ,619134289
Intercept: -,159727778 Std. Error: ,1655426  t(21) = -,9649  p = ,3456
     X1 beta = ,748          X2 beta = ,253         X3 beta = ,037
```

Fig. 2.16. **Implementation of the model of incremental inclusion of variables**

As soon as the stepwise regression procedure is over, the option *Stepwise regression summary* (stepwise regression results) at the bottom of the information window becomes active (Fig. 2.17).

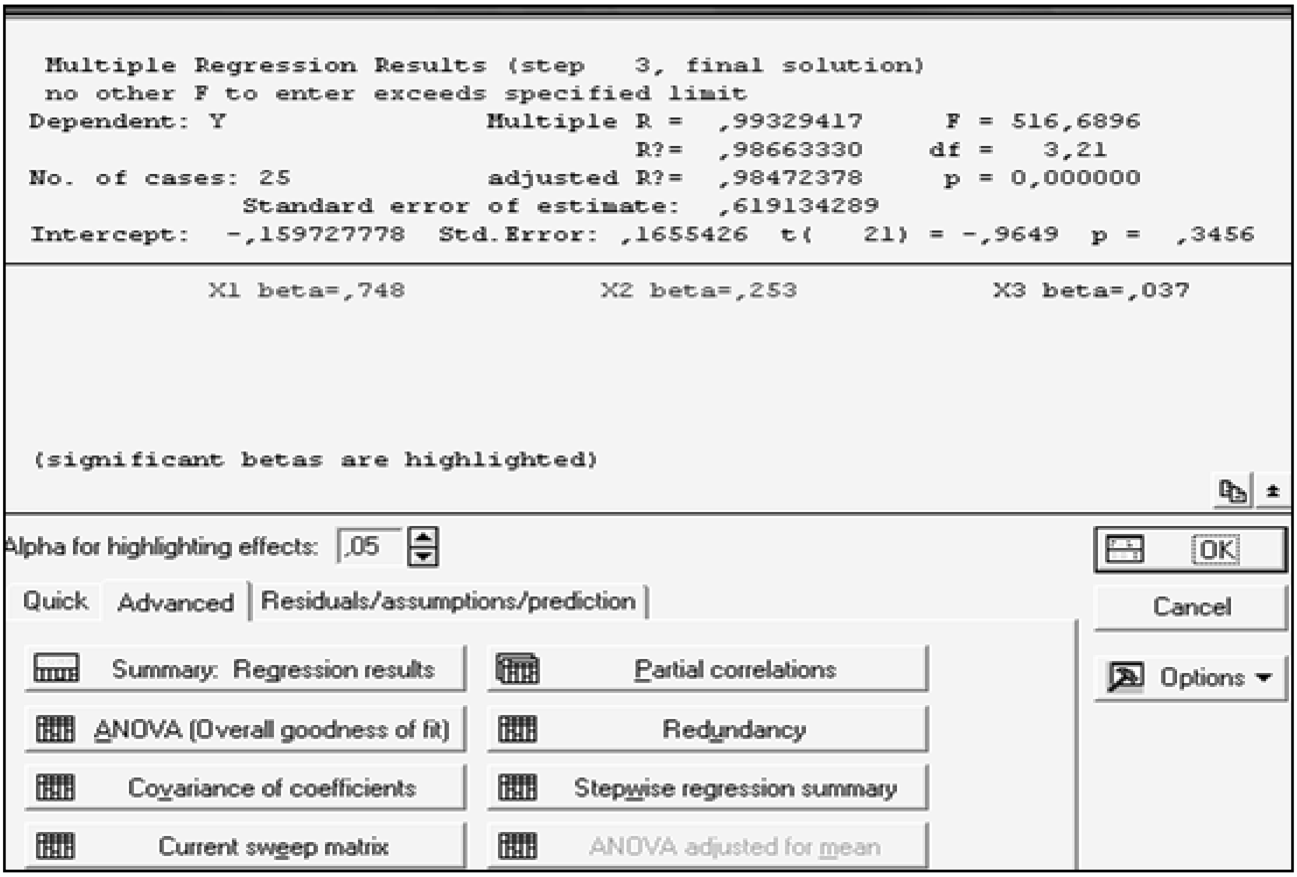


Fig. 2.17. **The window of selection of the stepwise regression results**

Clicking this option will give the table of stepwise regression results, shown in Fig. 2.18, which shows the adequacy of the model at each stage and change in the characteristics of the model for each step.

	Summary of Stepwise Regression; DV: Y (lr3)						
Variable	Step +in/-out	Multiple R	Multiple R-square	R-square change	F - to entr/rem	p-level	Variables included
X1	1	0,991162	0,982402	0,982402	1283,945	0,000000	1
X2	2	0,992624	0,985303	0,002901	4,343	0,048989	2
X3	3	0,993294	0,986633	0,001330	2,090	0,163019	3

Fig. 2.18. **The results of the stepwise inclusion regression**

Analysis of the regression model by the method of stepwise inclusion of the model variables (adequacy and statistical significance) is available with initiation of the option *Summary: Regression results* shown in Fig. 2.19.

N=25	Regression Summary for Dependent Variable: Y (Spreadsheet17) R= ,99329417 R?= ,98663330 Adjusted R?= ,98472378 F(3,21)=516,69 p<0,0000 Std.Error of estimate: ,61913 Beta	Std.Err. of Beta	B	Std.Err. of B	t(21)	p-level
Intercept			-0,15972ℇ	0,16554:	-0,96487⁄	0,34558ꞔ
X1	0,74794€	0,11981C	0,29384:	0,04706ꞔ	6,24277€	0,00000:
X2	0,25333ℇ	0,11973ꞇ	0,396141	0,18722:	2,11588C	0,04647ꞇ
X3	0,03674ꞩ	0,025417	0,17545C	0,12135ꞔ	1,445707	0,16301ꞔ

Fig. 2.19. **The regression model built by stepwise inclusion**

The sequence of the steps of the stepwise exclusion algorithm (*Backward stepwise*) is shown in Fig. 2.20.

```
                  Multiple Regression Results (Step 0)
Dependent: Y                Multiple R =  ,99329417     F = 516,6896
R? =  ,98663330    df = 3,21
  No. of cases: 25           adjusted R? =  ,98472378     p = 0,000000
               Standard error of estimate:  ,619134289
  Intercept:  -,159727778  Std. Error: ,1655426  t(21) = -,9649  p =  ,3456
         X1 beta = ,748         X2 beta = ,253         X3 beta = ,037

                  Multiple Regression Results (Step 1)
  Dependent: Y                Multiple R =  ,99262428     F = 737,4498
R? =  ,98530296    df = 2,22
  No. of cases: 25           adjusted R? =  ,98396686     p = 0,000000
               Standard error of estimate:  ,634287475
  Intercept:  -,030737796  Std. Error: ,1428533  t(22) = -,2152  p =  ,8316
           X1 beta = ,741           X2 beta = ,256

          Multiple Regression Results (Step 2, final solution)
No other F to remove is less than specified limit
  Dependent: Y                Multiple R =  ,99116180     F = 1283,945
R? = ,98240171    df = 1,23
  No. of cases: 25           adjusted R? =  ,98163656     p = 0,000000
               Standard error of estimate:  ,678818769
  Intercept:   ,081243426  Std. Error: ,1416544  t(23) = ,57353  p =  ,5719
           X1 beta = ,991
```

Fig. 2.20. **Implementation of the model of variable stepwise exclusion**

Clicking the option *Stepwise regression summary*, we obtain the results in the table of the stepwise exclusion regression for each stage (Fig. 2.21).

Variable	Summary of Stepwise Regression; DV:Y (lr3)						
	Step +in/-out	Multiple R	Multiple R-square	R-square change	F - to entr/rem	p-level	Variables included
X3	-1	0,992624	0,98530З	-0,00133C	2,09006ᶜ	0,16301ᶜ	2
X2	-2	0,991162	0,982402	-0,002901	4,342880	0,04898ᶜ	1

Fig. 2.21. **The results of the stepwise regression of exclusion**

The regression model built using the stepwise exclusion of variables is shown in Fig. 2.22.

N=25	Regression Summary for Dependent Variable: Y (Spreadsheet1 R= ,99116180 R?= ,98240171 Adjusted R?= ,98163656 F(1,23)=1283,9 p<0,0000 Std.Error of estimate: ,67882					
	Beta	Std.Err. of Beta	B	Std.Err. of B	t(23)	p-level
Intercept			0,081243	0,141654	0,57353	0,571850
X1	0,991162	0,027661	0,389395	0,010867	35,83218	0,000000

Fig. 2.22. **The regression model built using the stepwise exclusion**

Analysis of the constructed models will help select the most appropriate model in terms of description of real economic processes and their mutual relations and problem-solving purposes.

6. Construction of the model on the basis of the Ridge regression method for elimination of multicollinearity.

One method that allows you to adjust the matrix of independent variables with multicollinearity in the model, is the method of *Ridge regression*. The module *Multiple Regression* of this method can be carried out on the start menu bar with initiation of the *Advanced options* (*stepwise or ridge regression*) (see Fig. 2.14).

To choose the method *Ridge regression* and the displacement *parameter λ* in the *Advanced* menu, initiate the option *Ridge regression; lambda* (Fig. 2.23).

Obtain the results of the regression models built by Ridge regression (adequacy and statistical significance) with different values of the displacement parameter λ initiating the option *Summary: Regression results*, which are shown in Fig. 2.24.

The recommended change of the value of the parameter λ is in the range from 0.1 to 0.4.

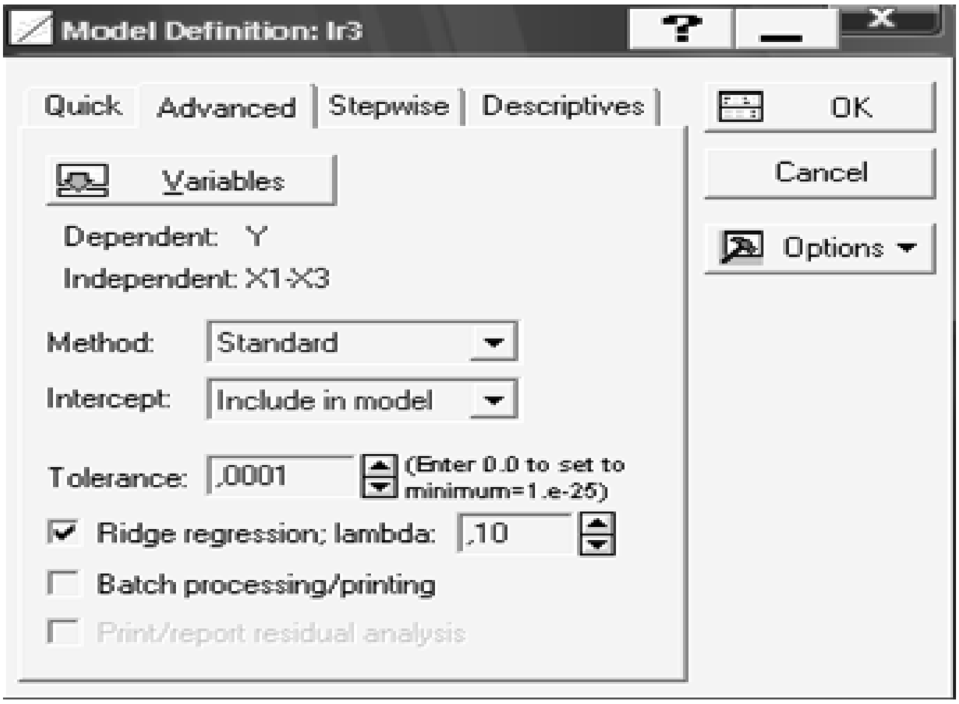


Fig. 2.23. **The choice of the Ridge regression method and the displacement parameter λ**

N=25	Ridge Regression Summary for Dependent Variable: Y (lr3) l=,10000 R= ,96780176 R?= ,93664024 Adjusted R?= ,92758884 F(3,21)=103,48 p<,00000 Std.Error of estimate: 1,3480					
	Beta	Std.Err. of Beta	B	Std.Err. of B	t(21)	p-level
Intercept			-0,127295	0,342878	-0,371254	0,714167
X1	0,521135	0,114292	0,204737	0,044902	4,559671	0,000171
X2	0,430861	0,114231	0,673730	0,178621	3,771846	0,001120
X3	0,027120	0,052697	0,129493	0,251613	0,514651	0,612171

Fig. 2.24. **Ridge regression with λ = 0.1**

7. Forecasting.

As the built models are adequate, their parameters are statistically significant, then the model can be used to make predictions. To calculate the predicted values of the dependent variable, you must use the option *Predict dependent variable* at the bottom of the results of the regression analysis. Initiating this option you must specify the value of the independent variables for which the dependent value is to be predicted (see Lab 1).

The results of forecasting for one-factor and three-factor econometric models are shown in Fig. 2.25.

Variable	Predicting Values for (lr3) variable: Y B-Weight	Value	B-Weight * Value
X1	0,293843	3,720000	1,093096
X2	0,396141	1,200000	0,475369
X3	0,175450	0,700000	0,122815
Intercept			-0,159728
Predicted			1,531552
-95,0%CL			1,274027
+95,0%CL			1,789077

Variable	Predicting Values for (lr3) variable: Y B-Weight	Value	B-Weight * Value
X1	0,389395	3,720000	1,448548
Intercept			0,081243
Predicted			1,529791
-95,0%CL			1,248942
+95,0%CL			1,810640

Fig. 2.25. **Forecast results for the three-factor and one-factor models**

The predicted value of the dependent variable (*Predicted*) for the three-factor models = 1.5315; and the confidence intervals for the forecast value are:

$$1.2740 \leq 1.5315 \leq 1.7891.$$

The predicted value of the dependent variable (*Predicted*) for the one-factor model = 1.5298; and the confidence intervals for the forecast value are:

$$1.2489 \leq 1.5298 \leq 1.8106.$$

Thus, we can conclude that the model predictions are almost the same, but the confidence intervals of change in the dependent variable are wider for the one-factor model.

Lab session 3. Autocorrelation and Heteroscedasticity

The goal is to consolidate the theoretical and practical material on the themes "Autocorrelation" and "Heteroscedasticity", acquire skills in the construction and analysis of multivariate econometric models in the *Multiple Regression* module.

The task is to test the residuals multiple regression that was constructed in Lab 2 for autocorrelation and heteroscedasticity:

1. Check the model for autocorrelation of residuals using the Durbin – Watson statistics and the cyclic autocorrelation coefficient.
2. Check the model for heteroscedasticity of errors using the nonparametric Goldfeld – Quandt test, the μ-test, the parametric Goldfeld – Quandt test.
3. Draw conclusions.

Guidelines

1. Determination of autocorrelation using the Durbin – Watson method and the cyclic autocorrelation coefficient.

This criterion is based on the examination of autocorrelation between the adjacent members of a series of residuals.

To conduct this analysis you must calculate the theoretical value of the dependent variable and model residuals, give the results of the study of the models using the criterion of Durbin – Watson and the cyclic autocorrelation coefficient and draw conclusions about the presence of autocorrelation.

To calculate and analyze the residues, use the option *Perform residual analysis* at the bottom of the results of the regression analysis. By initiating this option, you will get the menu for the error analysis model (see Lab 2).

Initiating the key of the residual analysis *Summary: Residuals & Predicted*, get a table of the observed values of the dependent variable (*Observed value*), the theoretical values of the dependent variable (*Predicted value*), model errors (*Residual*) and the model error screening schedule in the range $\pm 3\sigma$ which analyzes the error variance sustainability (Fig. 3.1).

		Raw Residual (lr3) Dependent variable: Y		
Case	Raw Residuals -3s . . 0 . . +3	Observed Value	Predicted Value	Residual
1	. . . * . . .	0,13000	0,42236	-0,29236
2	. . . . * . .	0,77000	0,54082	0,22918
3	. . . * . .	0,25000	0,26496	-0,01496
4	. . . * . . .	0,50000	0,71138	-0,21138
5	. . . *. . .	0,12000	0,27994	-0,15994
6	. . . .* . .	0,22000	0,11230	0,10770
7	. . . * . . .	0,82000	1,07801	-0,25801
8	. . . . .* .	0,91000	0,16130	0,74870

Fig. 3.1. **The model error analysis. A fragment**

Initiating the key *Advanced / Durbin – Watson statistic* in the menu of error analysis, we obtain the value of autocorrelation of the model errors according to the criterion Durbin – Watson and the value of the cyclic autocorrelation coefficient (Fig. 3.2).

	Durbin-Watson d (Spr3 and serial correlation of	
	Durbin-Watson d	Serial Corr.
Estimate	2,464231	-0,238982

Fig. 3.2. **Autocorrelation error model**

The values of the coefficients are compared with the tabulated values and conclusions are drawn about the presence of autocorrelation of residuals in the model. If $d \to 0$, then there is positive autocorrelation. If $d \to 2$, then there is no autocorrelation. If $d \to 4$, then there is negative autocorrelation.

In the Durbin – Watson statistics there are upper and lower limits. For the model with three independent variables and 25 observations, the lower limit is $d_l = 0.90$, the upper one is $d_u = 1.41$.

Since the calculated value falls within the range of $d_u \le d < 4 - d_u$ $(1.41 \le d < 2.59)$, a conclusion can be drawn that there is no autocorrelation of residuals in the tested model.

The cyclic autocorrelation coefficient (r^0) expresses the degree of relationship of each subsequent value with the previous one. The coefficient r^0 can take the values in the interval (-1, +1). Negative values indicate a negative autocorrelation, the positive ones indicate a positive autocorrelation. The values contained in a certain critical area near zero indicate lack of autocorrelation.

In fact, the calculated value of the cyclic autocorrelation coefficient is compared with the tabulated value for the selected significance level and the number of observations n. If $r^0 > 0$ and $r^0 \ge r_{tab}^{+}$, there is a positive autocorrelation. If $r^0 < 0$ and $|r^0| \ge r_{tab}^{-}$, there is a negative autocorrelation.

In this case, the calculated value of the cyclic autocorrelation coefficient $-0.239 < 0$ and the tabulated value is $r_{tab}^{-}(25; 0.05) = 0.356$. Therefore, the actual value is less than the tabulated one ($-0.239 < 0.356$), then there is no autocorrelation of residuals.

For estimation of parameters with autocorrelating residuals the following methods are used: Aitken's method (GLS), the method of conversion of initial information, Durbin method, Cochrane – Orcutt estimation.

2. The nonparametric Goldfeld – Quandt test for determination of heteroscedasticity of errors.

This test is based on the number of peaks in the values of residuals after ranging the observations in the ascending order of x_{ij}. The regularity of changes of residuals while the variance is homogeneous is called homoscedasticity, which is illustrated in Fig. 3.3, whereas Fig. 3.4 depicts the phenomenon of heteroscedasticity.

This test is not as reliable as the parametric one, but it is quite simple.

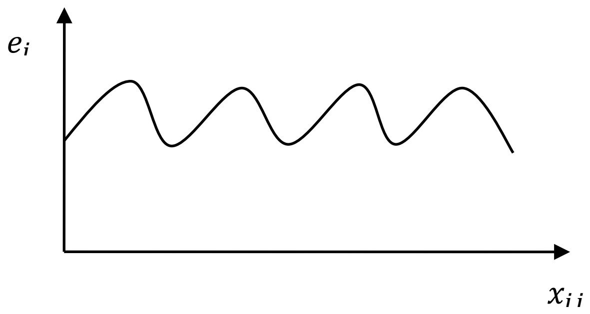


Fig. 3.3. **Homoscedasticity**

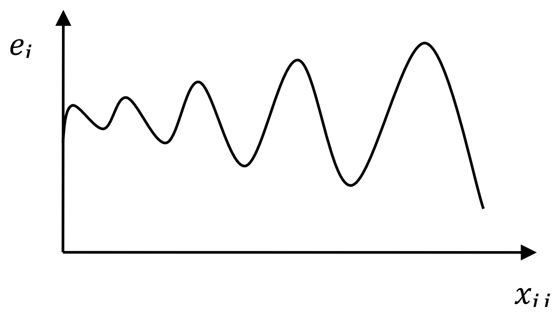


Fig. 3.4. **Heteroscedasticity**

Note that Fig. 3.3 shows how residuals with a constant variance change, and Fig. 3.4 shows residuals which have variable variance for different groups of observations.

For the analysis of this problem, add one column to the table with the input data. We'll call this column "Residuals". In this column we will copy the values of residuals from the table shown in Fig. 3.1.

To construct a graph of the residuals use Graphs / 2D Graphs / Scatterplots and select variables and parameters of the graph (Fig. 3.5) and draw the graph (Fig. 3.6).

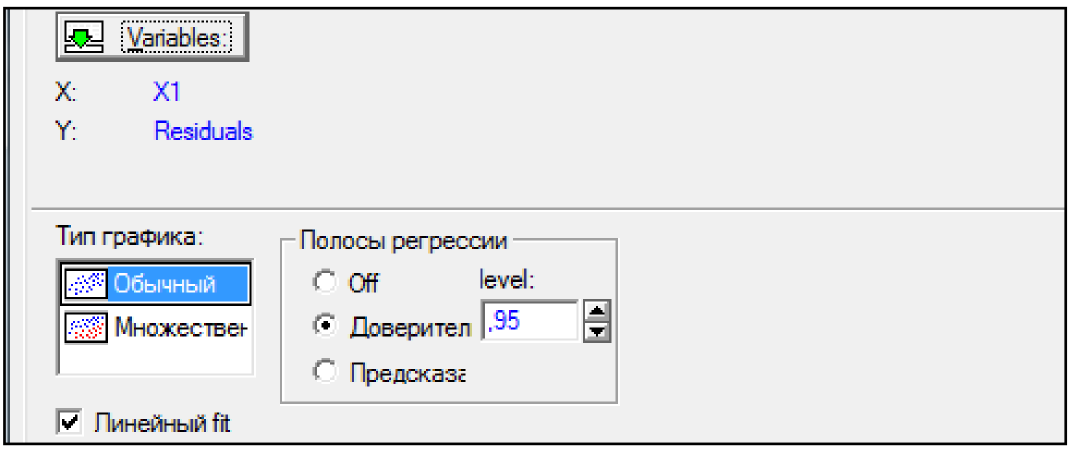


Fig. 3.5. **Parameters of the graph**

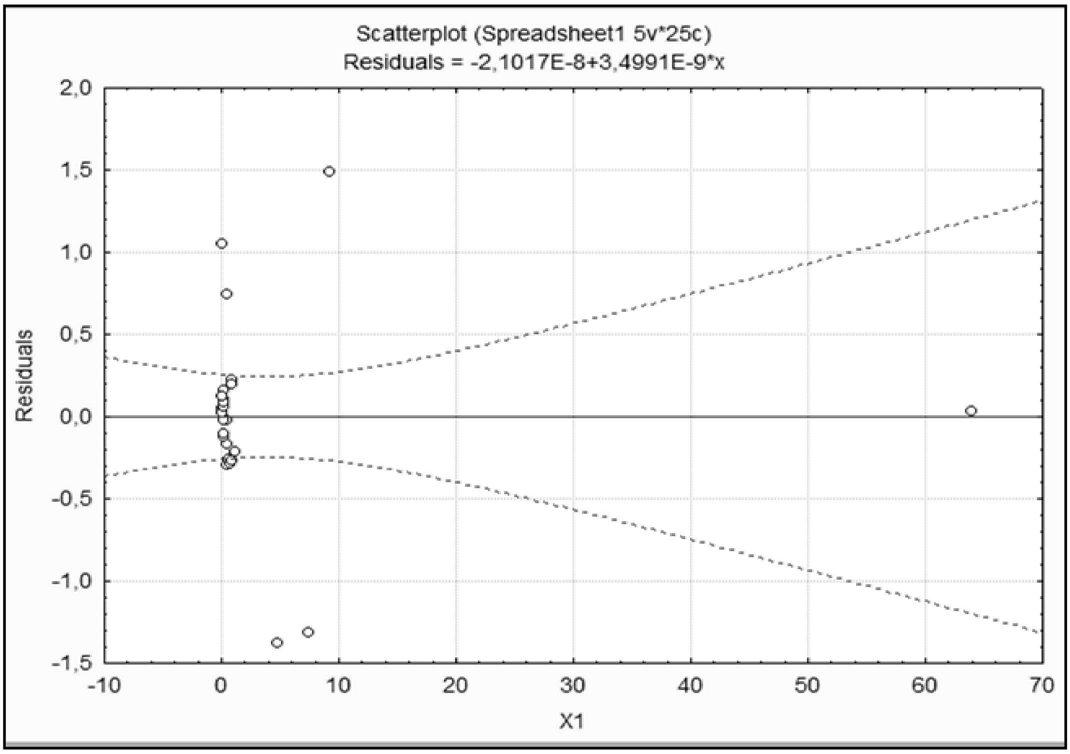


Fig. 3.6. **The graph of the residuals of the variables *X*1**

We must repeat plotting the residuals of the variables *X*2 and *X*3 (Fig. 3.7).

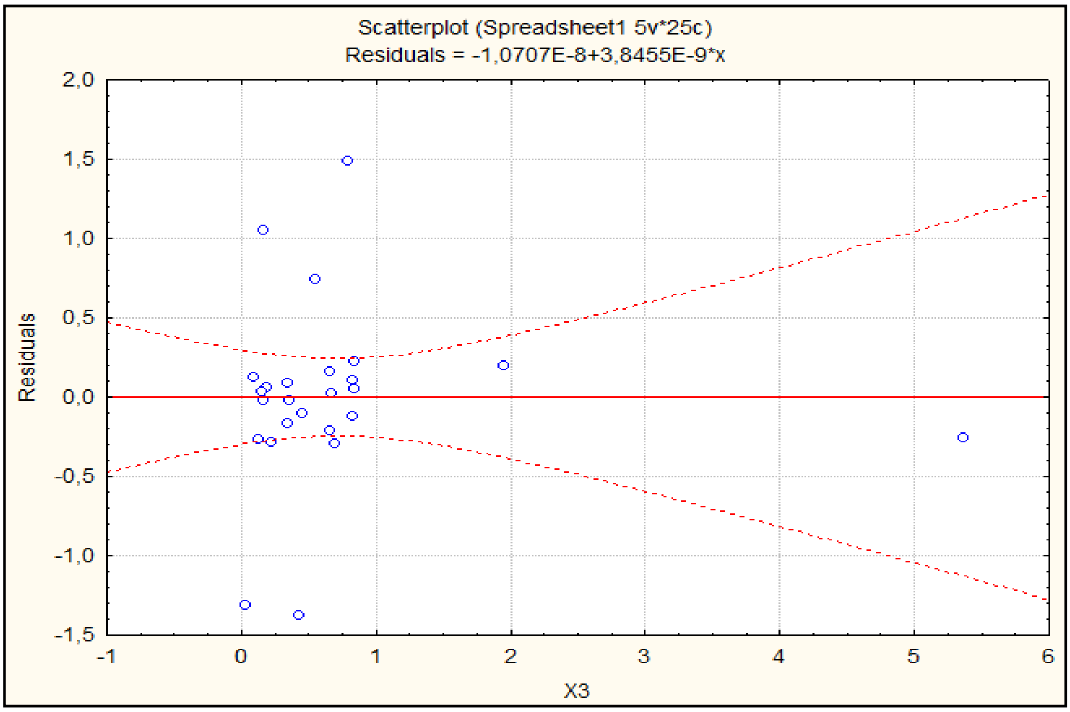


Fig. 3.7. **The graph of the residuals of the variables *X*3**

After analyzing the graphs (Fig. 3.6 and 3.7), we can assume that there is heteroscedasticity residuals.

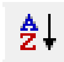
3. Determination of heteroscedasticity using the parametric Goldfeld – Quandt test.

This test can be used for a small number of observations.

In this case, Goldfeld and Quandt offered to consider the case, when the variance of residuals grows in proportion to the square of one of the independent variables of the regression: $Y = XA + u$.

Assume that the source of heteroscedasticity is variable $X1$.

To detect the presence of heteroscedasticity, we will apply the parametric test, for which you must perform the following steps:

Step 1. Sort the observations according to the values of $X1$ vector elements. To do this, select the $X1$ column and click .

Step 2. Discarding c observations from the center of the vector, where $c = 4n/15$, n is the number of elements of the vector $X1$:

$$c = 4 \cdot 25/15 \approx 7.$$

As a result, two sets of observations appear that have the following volume:

$$n_1 = \frac{n-c}{2} = \frac{25-7}{2} = 9, \qquad n_2 = n - c - n_1 = 25 - 7 - 9 = 9.$$

Step 3. Construct two regressions using the *Multiple Regression* module based on two sets of observations of volumes n_1 and n_2 (given that n_1 and n_2 exceed the number of variables m). To do this, click 

in the start panel of the *Multiple Regression* module and select the cases number 1 – 9. Then select the data click OK. The results of the first regression construction are shown in Fig. 3.8.

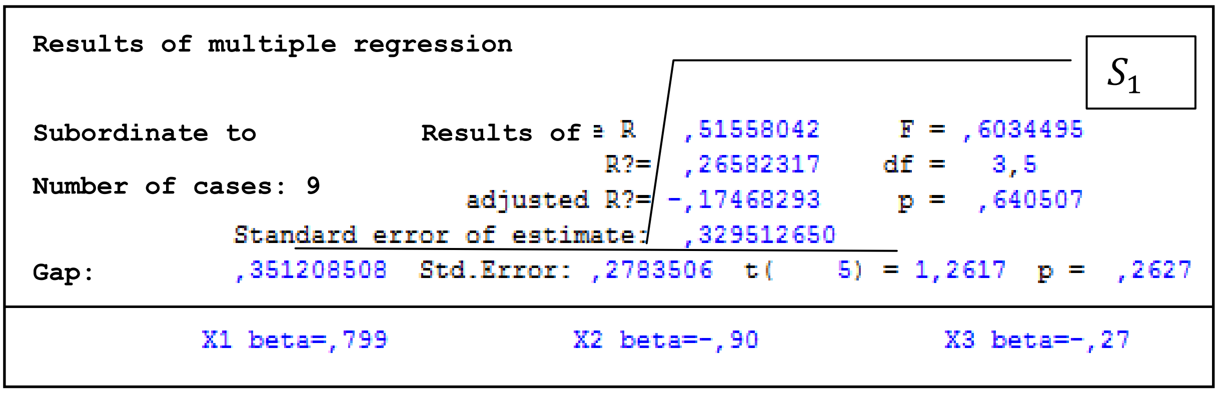
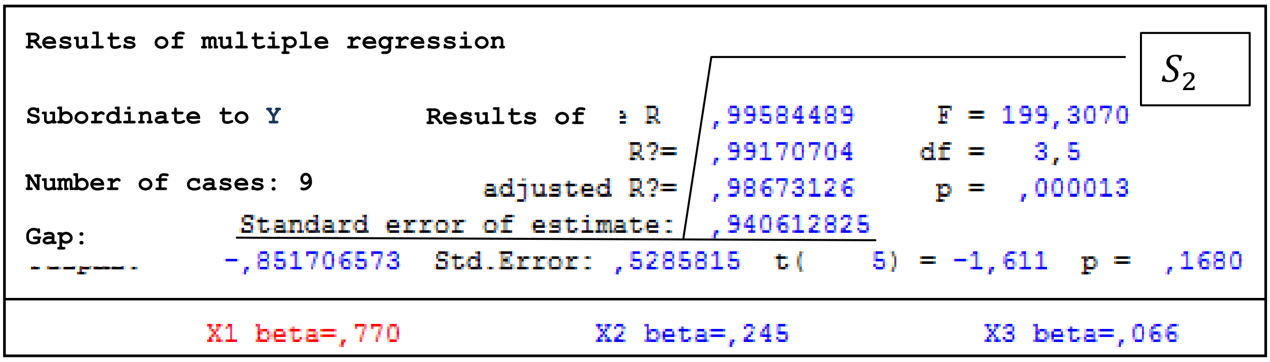


Fig. 3.8. **The results of the first regression. A fragment**

Click  again in the start panel the *Multiple Regression* module and select the cases number 17 – 25 (the last 9 observations). Then select the data click *OK*. The results of the second regression construction are shown in Fig. 3.9.


```
Results of multiple regression                                          S2

Subordinate to Y          Results of  : R    ,99584489     F = 199,3070
                                      R?=    ,99170704    df =    3,5
Number of cases: 9           adjusted R?=    ,98673126     p =  ,000013
Gap:              Standard error of estimate: ,940612825
---------      -,851706573  Std.Error: ,5285815  t(    5) = -1,611  p =  ,1680

            X1 beta=,770              X2 beta=,245             X3 beta=,066
```

Fig. 3.9. **The results of the second regression. A fragment**

Step 4. Find the sums of the squared residuals S_1 and S_2 based on the first and second models. These values are standard error residuals (*Standard errors of estimate*) (Fig. 3.7 and 3.8). $S_1 = 0.3295;\ S_2 = 0.9406$.

Step 5. Calculate the value of the criterion:

$$F^* = {S_1^2}/{S_2^2},\ \text{if } S_1^2 > S_2^2 \quad \text{or} \quad F^* = {S_2^2}/{S_1^2},\ \text{if } S_2^2 > S_1^2,$$

which in the case of execution of the hypothesis of homoscedasticity will meet F-distribution with $k_1 = n_1 - m$ and $k_2 = n_2 - m$ degrees of freedom, where n_1 is the number of observations corresponding to the denominator value; n_2 is the number of observations corresponding to the numerator value. In this case $S_2 > S_1$:

$$F^* = \frac{(0.9406)^2}{(0.3295)^2} = 2.85.$$

Step 5. Compare the calculated value of F^* with the tabulated value of the F-test for k_1 and k_2 degrees of freedom, and the selected confidence level. Compare F^* criterion with the critical value of F-test for $\gamma_1 = 6$ and $\gamma_2 = 6$ degrees of freedom and a confidence level of $\alpha = 0.95$ ($F_{0.05} = 4.28$). Since $F^* \le F_{tab}$, there is no heteroscedasticity. There is a homoscedasticity.

Lab session 4. The Multiple Nonlinear Cobb – Douglas Production Function

The goal is to assimilate the theoretical and practical materials of the theme "Nonlinear Regression", practice building and analysis of non-linear production functions in the module *Nonlinear Estimation*.

The task is to test the linear and nonlinear relationships between the volume of production and the available production resources in the modules *Multiple Regression* and *Nonlinear Estimation* (*Statistica*):

1. Construct a linear multiple regression model. Identify all of its characteristics (find the model parameters using the least square method, standard deviation of the model parameters, variance and standard deviation of the model residuals, the coefficients of multiple correlation and determination).
2. Check the significance of econometric models using the Fisher criterion.
3. Check for residual autocorrelation using the Durbin – Watson criterion and the cyclic autocorrelation coefficient. Provide a histogram and a graph of the distribution of errors. Draw conclusions about the presence of autocorrelation.
4. Draw conclusions about the adequacy of the linear multiple regression model.
5. Check the existence of a nonlinear relation between the volume of production and the value of production resources by constructing a Cobb – Douglas production function.
6. Draw a histogram and a graph of the distribution of errors. Draw conclusions regarding the presence of autocorrelation of residuals.
7. Draw conclusions about the adequacy of the nonlinear econometric models.
8. Identify the characteristics of the production function. Find combinations of productive resources at the fixed levels of production. Draw a graph of isoquants.

Guidelines

For the construction and comprehensive analysis of multiple linear econometric models, the Multiple Regression module (Multiple Regression) is used. For building nonlinear econometric models the module provides *Nonlinear Estimation*.

1. Running *Statistica* and preparing data.

In the applications menu select the program *Statistica*. After launching it choose *File/New* in the menu to prepare the data. After entering the data, click *OK*. After filling in all the cells in the data field, you will get a table similar to that shown in Fig. 4.1, where *L* is the labor force (thousand people), *K* is the capital assets (thousand UAH), *Y* is the volume of output (million UAH).

	1 L	2 K	3 Y
1	0,14	0,16	0,11
2	0,09	0,21	0,22
3	0,3	0,26	0,3
4	0,38	0,18	0,56
5	5,07	3,8	7,71
6	6,14	2,19	3,5
7	1,53	1,13	6,78
8	0,71	1,42	1,44
9	0,7	0,63	0,84
10	0,55	0,45	0,76
11	0,47	0,3	2,03
12	0,27	0,25	0,81
13	0,14	0,16	0,11
14	0,09	0,21	0,22
15	0,3	0,26	0,3
16	0,38	0,18	0,56
17	0,87	1	1
18	3,79	0,65	0,85
19	0,53	0,77	0,56
20	0,05	0,55	0,17
21	0,121	0,165	0,109
22	0,271	0,092	0,116
23	0,189	0,197	0,192
24	0,026	0,108	0,173
25	0,142	0,212	0,028

Fig. 4.1. **The output for the construction of the production function**

Thus, there is a classical production function between the factors of production (the number of workers, the cost of the core business) and the volume of production.

Consider the order of constructing the multiple production functions.

2. Construction of the multiple linear production function.

The construction and analysis of the multiple linear production function should be carried out in the *Multiple Regression* module as it was discussed in the previous laboratory sessions. The characteristics of the linear production function are shown in Fig. 4.2.

	Regression Summary for Dependent Variable: Y (Spreadsheet9) R= ,83173856 R?= ,69178903 Adjusted R?= ,66376985 F(2,22)=24,690 p<,00000 Std.Error of estimate: 1,1488					
N=25	Beta	Std.Err. of Beta	B	Std.Err. of B	t(22)	p-level
Intercept			-0,060536	0,289925	-0,208798	0,836527
L	-0,007511	0,207505	-0,009274	0,256206	-0,036198	0,971451
K	0,837897	0,207505	2,007017	0,497037	4,037966	0,000550

Fig. 4.2. **The model of the linear production function**

Thus the linear production function is represented like $\hat{Y} = -0.06054 - 0.009274 \cdot \mathrm{L} + 2.007 \cdot K$.

To construct the graph of the production factors and the production volume use *Graphs / 2D Graphs / Scaterplots* and select variables and parameters of the graph (Fig. 4.3) and draw the graph (Fig. 4.4).

The comprehensive adequacy analysis of the linear multivariable function should be held similar to that in Lab 1.

Fig. 4.3. **The graph task options**

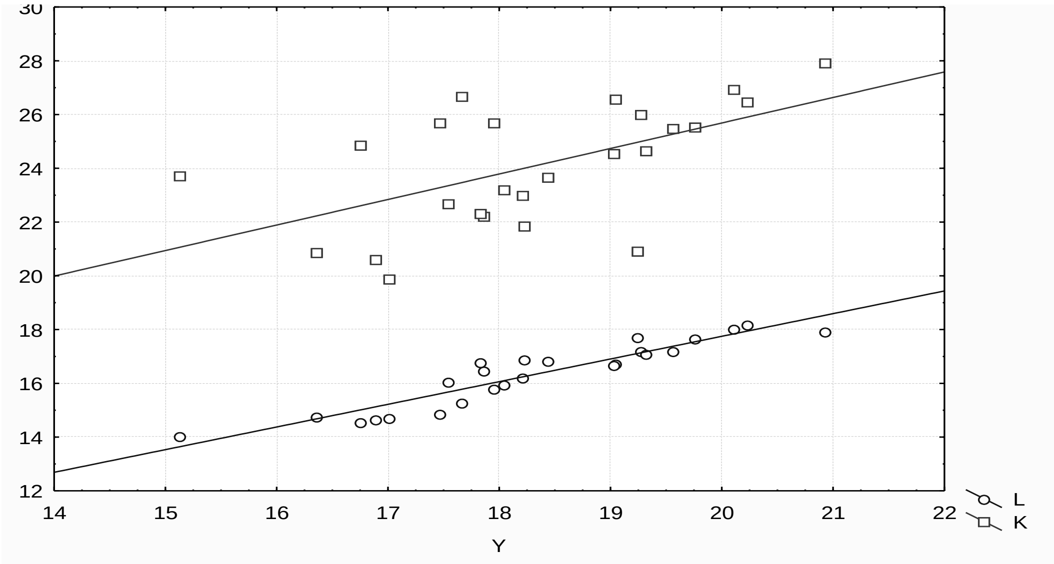


Fig. 4.4. **The graph of the production factors**

3. Construction of the Cobb – Douglas nonlinear production function.

Check for the existence of a nonlinear relation between the volume of production and the value of production resources by building a Cobb – Douglas production function in the *Advanced Linear / Nonlinear Models* module. The selection of the module *Nonlinear Estimation* is shown in Fig. 4.5.

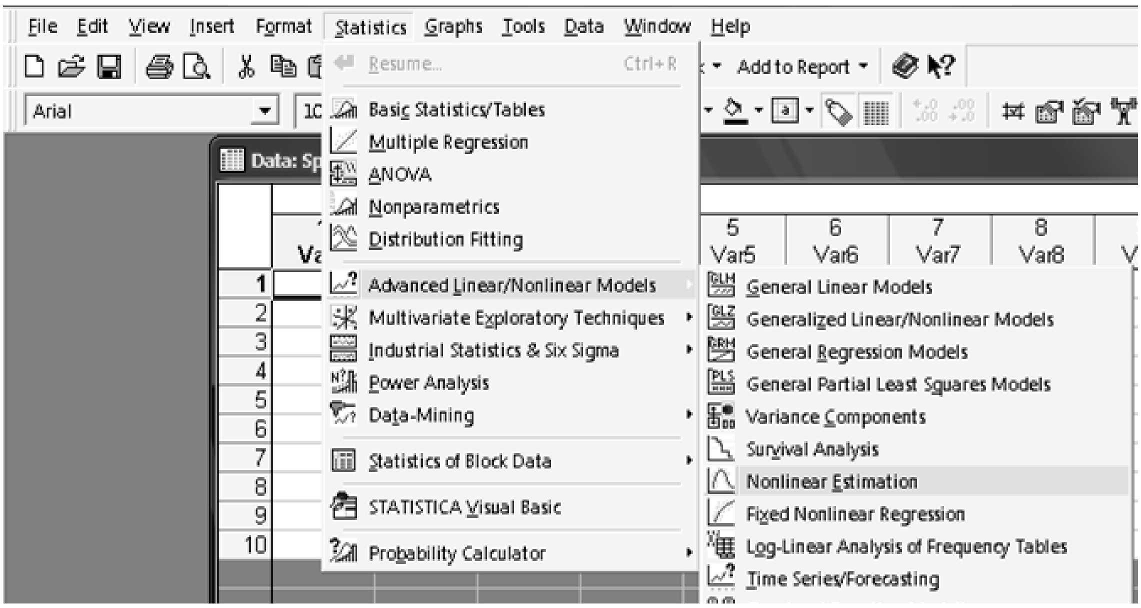


Fig. 4.5. **Selecting the Nonlinear Estimation module**

The type of the starting panel module is shown in Fig. 4.6. There are the following nonlinear estimations: *the user-specified regression, least squares*; *the user-specified regression, the custom loss function; the Quick Logit* *regression; the Quick Probit regression; the exponential growth regression; the piecewise linear regression*.

To solve this task, you must select the *user-specified regression, least squares* (Fig. 4.6).

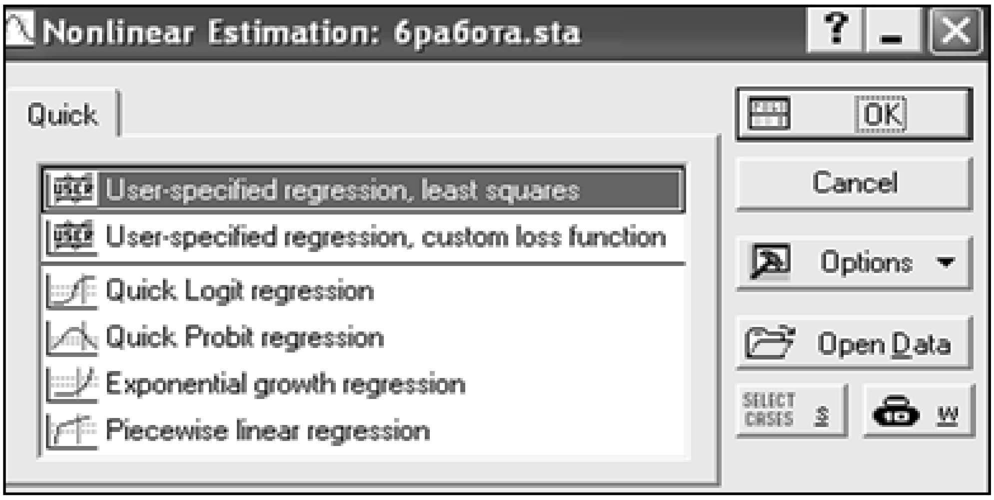


Fig. 4.6. **The starting panel of the module Nonlinear Estimation**

Then, in the window *Function to be estimated* complete the kind of function that fits the mathematical model of the Cobb – Douglas type: $Y = a_0L^{a1}K^{a2}$, as shown in Fig. 4.7.

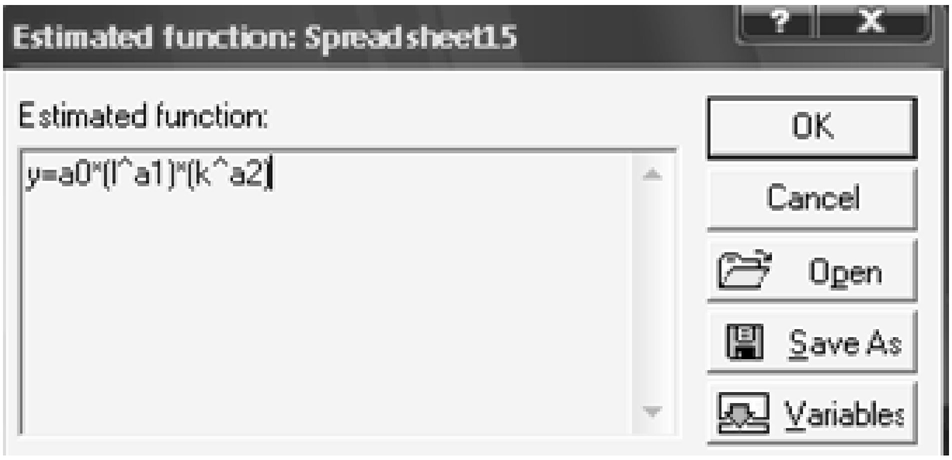


Fig. 4.7. **Determining the type of the function evaluation**

In the following window, specify the method of parameter estimation of the nonlinear functions Levenberg – Marquardt or Gauss – Newton. If necessary, define the parameters of the iterative procedure and the initial values of the estimated parameters on the *Advanced* tab and run the assessment procedure (Fig. 4.8).

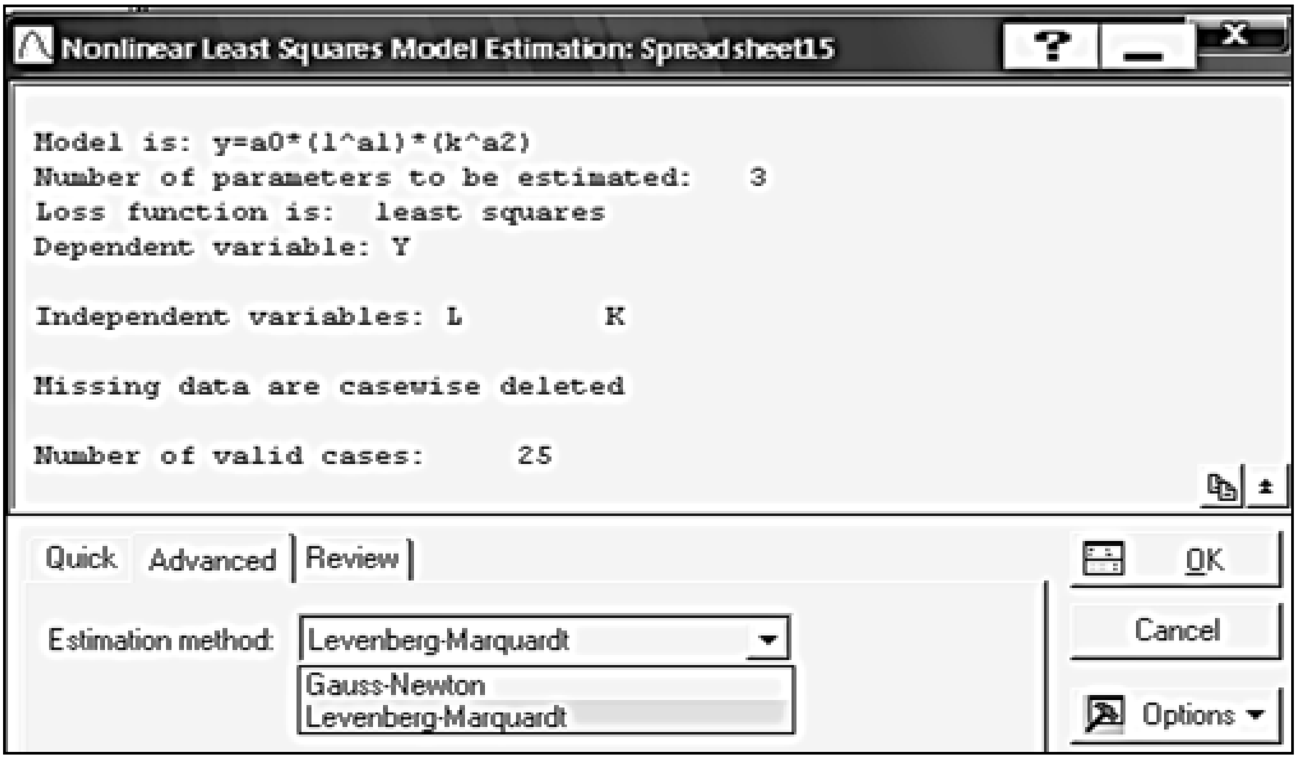


Fig. 4.8. **The choice of the method of the model parameter estimation**

The results of the Cobb – Douglas production function can be analyzed using the window shown in Fig. 4.9.

This box presents the results of the model at the bottom of the information part. The options for comprehensive analysis of the models are presented.

The quality of the resulting model is estimated using the values: *Final value of the Loss function* (the final value of the loss function is the sum of the squares of the error model); *Proportion of variance accounted for* (the percentage of the variance explained); *R* (the multiple correlation coefficient).

The estimates of the model parameters can be obtained by initiating the option Quick / *Summary: Parameter estimates* (*Result: parameters of the model*) (Fig. 4.9). The estimation of parameters is shown in Fig. 4.10.

This window shows the evaluation of the model parameters (Estimates); mean square deviation parameters (*Standard error*); Student statistics (t-value); Student statistical significance level (*p*-level); the lower and upper limits of the interval parameter estimates (*Lo. Conf Limit; Up. Conf Limit*).

Thus, the production function will be as follows: $\hat{Y} = 1.0151 \cdot L^{0.832} \cdot K^{0.1791}$. The interval of the estimation of parameters can vary within the limits:

$$0.6078 \le a_0 \le 1.4224;\ 0.6931 \le a_1 \le 0.9708;\ 0.0703 \le a_2 \le 0.2879.$$

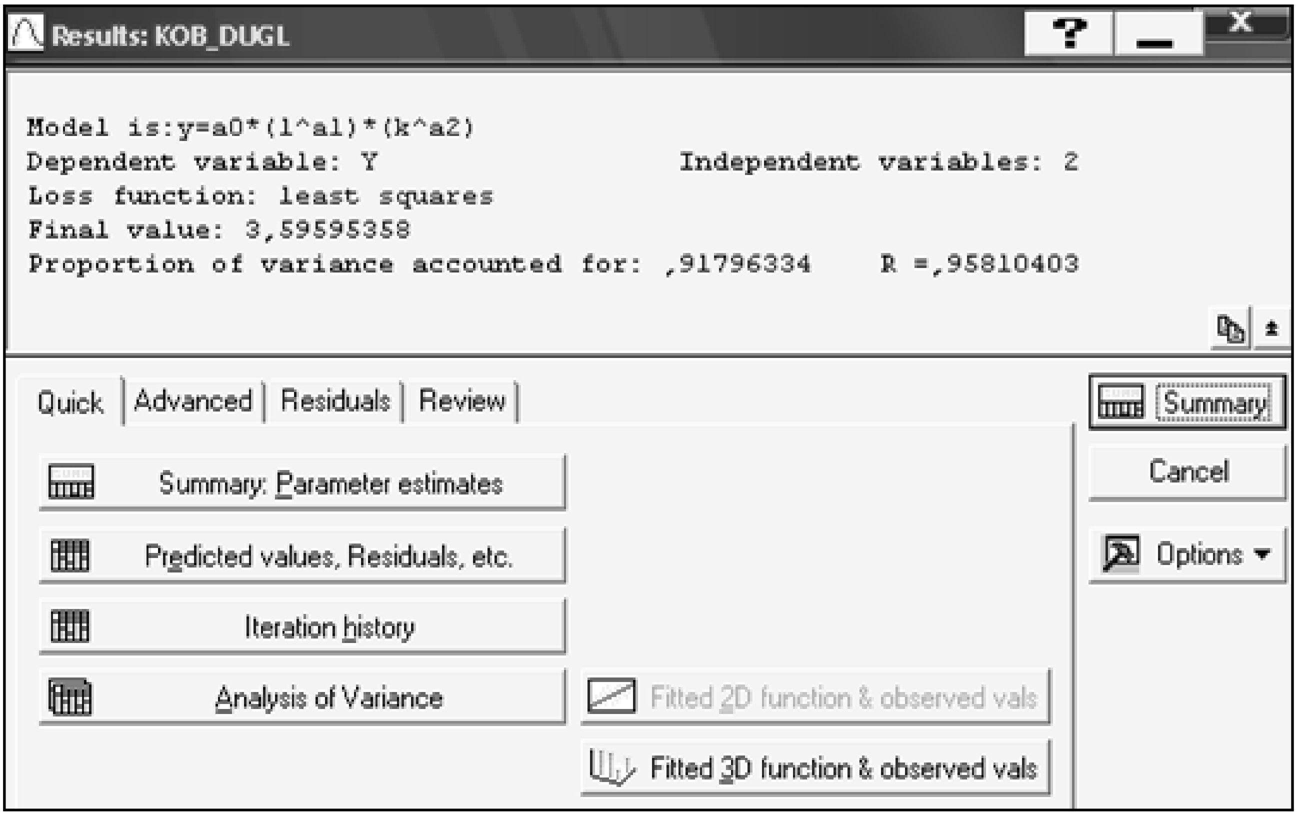


Fig. 4.9. **The results of constructing the nonlinear function**

	Model is: y=a0*(l^a1)*(k^a2) (KOB_DUGL) Dep. Var. : Y Level of confidence: 95.0% (alpha=0.050)					
	Estimate	Standard error	t-value df = 22	p-level	Lo. Conf Limit	Up. Conf Limit
a0	1,015082	0,196381	5,16894	0,000035	0,607813	1,422351
a1	0,831944	0,066939	12,42837	0,000000	0,693121	0,970767
a2	0,179076	0,052450	3,41424	0,002485	0,070302	0,287850

Fig. 4.10. **The estimation of the model parameters**

Initiating the key *Iteration history*, we obtain the results of the incremental reassessment of the procedure parameters. The model parameters that minimize the criterion of the minimum sum of the squared errors (*Loss function – The root of the sum of the squares of errors*) are obtained in 18 iterations (Fig. 4.11).

	Model is: Y=a0*(L^a1)*(K^a2) (KOB_DUGL Dep. Var. : Y			
	Loss Function	a0	a1	a2
1	90,88450	0,100000	0,100000	0,100000
2	86,61525	0,162661	0,324647	0,297077
3	82,21830	0,197060	0,400325	0,363446
4	74,07371	0,238798	0,476075	0,429868
5	57,39187	0,293465	0,557900	0,501593
6	23,82586	0,362687	0,642153	0,575360
7	3,45091	0,467805	0,725975	0,519059
8	3,00184	0,504927	0,741867	0,476206
9	2,77441	0,586740	0,777673	0,394912
10	2,41607	0,676775	0,814787	0,317968
11	2,24122	0,783303	0,843207	0,246994
12	1,95950	0,848843	0,852135	0,216425
13	1,92320	0,874916	0,853537	0,206579
14	1,91323	0,930557	0,847682	0,191915
15	1,90420	0,987293	0,837301	0,182518
16	1,89661	1,014538	0,831927	0,179142
17	1,89630	1,015079	0,831944	0,179077
18	1,89630	1,015082	0,831944	0,179076

Fig. 4.11. **The iterative procedure for determining the parameters**

The analysis of variance of the built model is done by initiating *Analysis of Variance* (Fig. 4.12). The table of the variance analysis contains the sum of the squares of deviations for regression (*Sums of Squares Regress*), the sum of the squares of the model errors (*Sums of Squares Residua*l), the variance of errors (*Mean Squares Residual*) and the *Fisher criterion* (*F-value*).

	Model is: Y=a0*(L^a1)*(K^a2) (KOB_DUGL) Dep. Var. : Y				
Effect	1 Sum of Sqares	2 DF	3 Mean Squares	4 F-value	5 p-value
Regression	8422,006	3,00000	2807,335	17175,24	0,00
Residual	3,596	22,00000	0,163		
Total	8425,602	25,00000			
Corrected Total	43,833	24,00000			
Regression vs.Corrected Total	8422,006	3,00000	2807,335	1537,09	0,00

Fig. 4.12. **The results of the analysis of variance**

For further analysis it is necessary to obtain the theoretical values of the dependent variable and the model errors, which can be done by initiating the key (*Predicted values, Residuals*) on the tab *Quick*. The results are shown in Fig. 4.13.

	Model is: y=a0*(L^a1)*(K^a2) (KOB_ Dep. Var. : Y		
	Observed	Predicted	Residuals
1	18,22000	18,51119	-0,291194
2	17,53000	17,87537	-0,345367
3	16,35000	16,40290	-0,052898
4	17,94000	18,03605	-0,096054
5	19,24000	19,12316	0,116838
6	17,66000	17,67086	-0,010860
7	18,21000	18,05382	0,156184
8	18,44000	18,71646	-0,276463
9	19,76000	19,78501	-0,025006
10	17,85000	18,17116	-0,321163
11	19,27000	19,37352	-0,103524
12	20,93000	20,31371	0,616287
13	19,04000	19,04406	-0,004055
14	16,75000	16,72527	0,024725
15	19,56000	19,32760	0,232404
16	17,82000	18,49325	-0,673249
17	16,88000	16,29501	0,584987
18	17,46000	17,13878	0,321217
19	19,02000	18,71514	0,304857
20	20,11000	20,27532	-0,165316
21	19,32000	19,08862	0,231383
22	18,04000	17,85077	0,189229
23	20,22000	20,39817	-0,178171
24	17,01000	16,24533	0,764668
25	15,13000	16,12579	-0,995790

Fig. 4.13. **The theoretical value of the dependent variable and the model errors**

To compare the empirical (*Observed*) and theoretical *(Predicted*, calculated by the model) values of the dependent variable, construct a graph. Select the options of the graphical analysis of the results and the type of graphics shown in Fig. 4.14. Fig. 4.15. shows a graph comparing the theoretical and empirical values of the dependent variable.

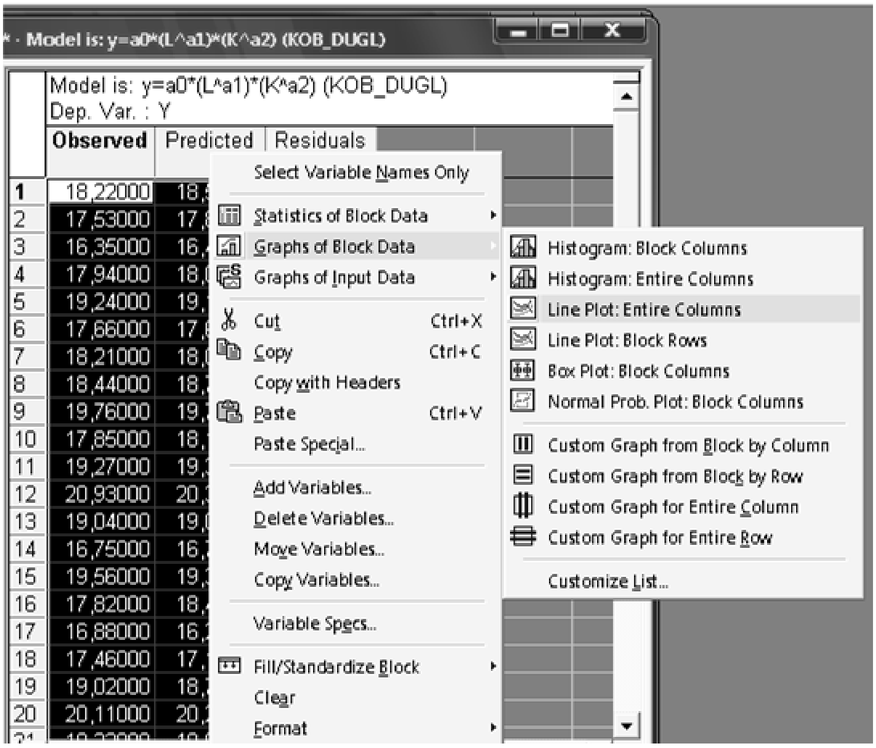


Fig. 4.14. **Selecting the options of the graphical analysis of variables**

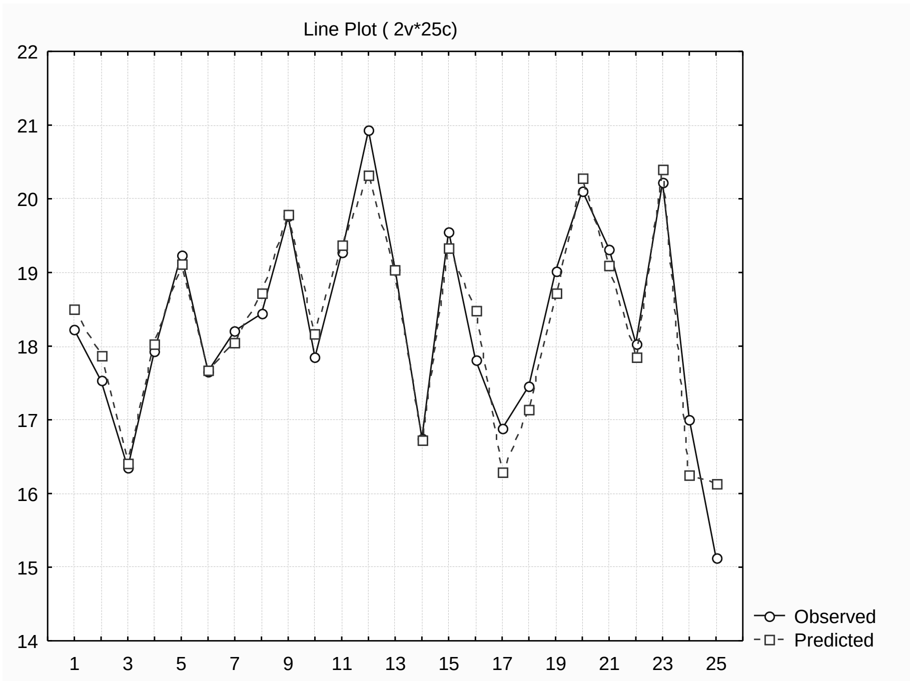


Fig. 4.15. **The graph of the theoretical and empirical values of the dependent variable**

The graph of the production function in three dimensions (factors of production and the dependence of output) will be received by initiating the key *Fitted 3D function & observed vals* (Fig. 4.9). The graph is presented in Fig. 4.16.

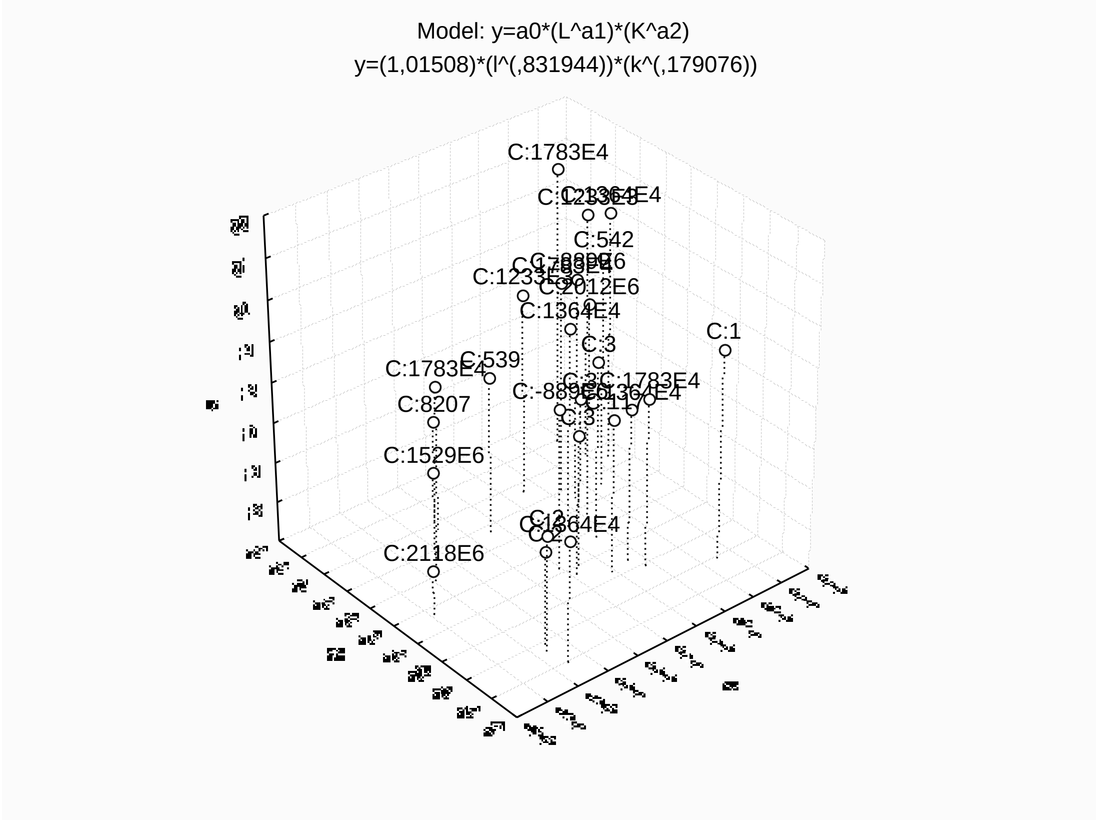


Fig. 4.16. **The graph of the production function in three dimensions**

A comprehensive analysis of the model errors is available in the options *Residuals* (Errors) (Fig. 4.17).

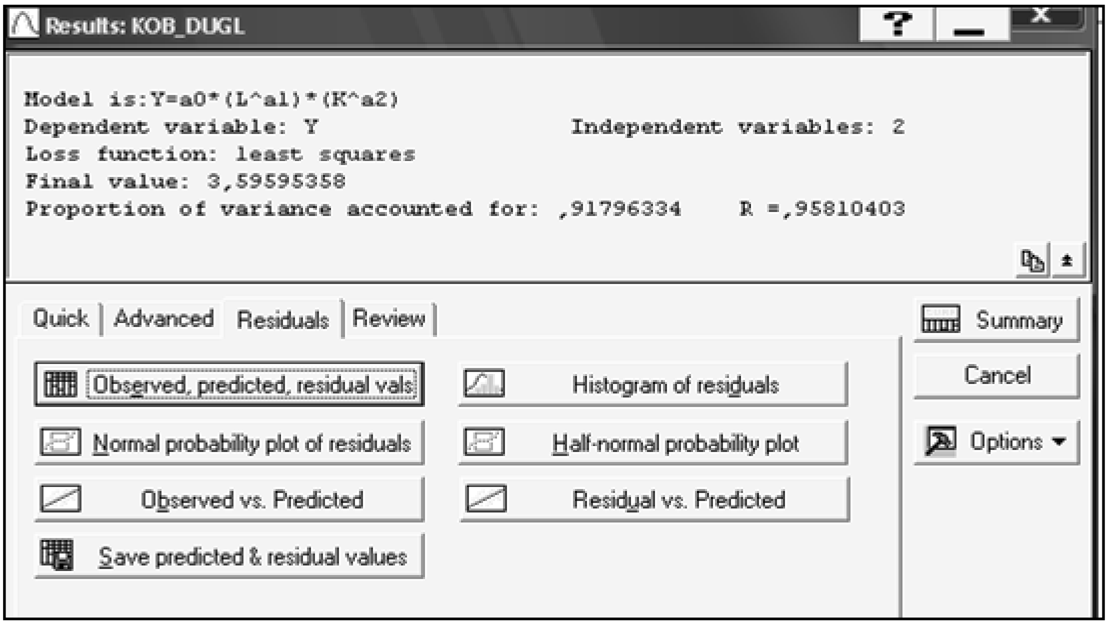


Fig. 4.17. **The menu of the model error analysis**

The histogram of errors and the graph of timing the errors on the normal distribution probability paper are shown in Fig. 4.18 and 4.19.

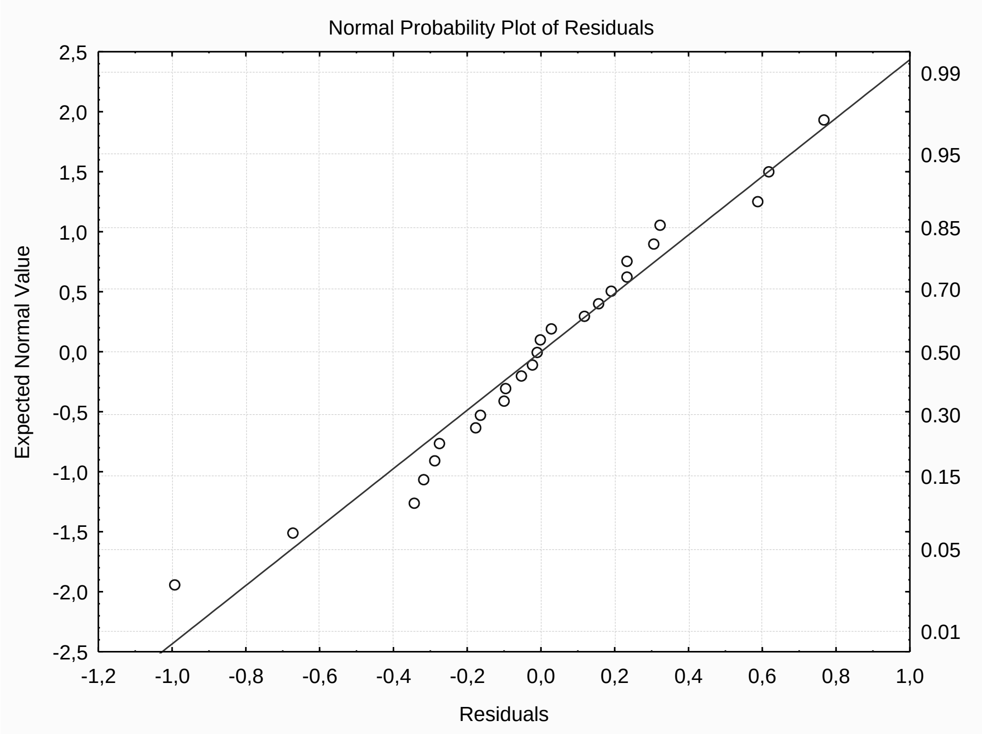


Fig. 4.18. **The graph of the error on the normal probability paper**

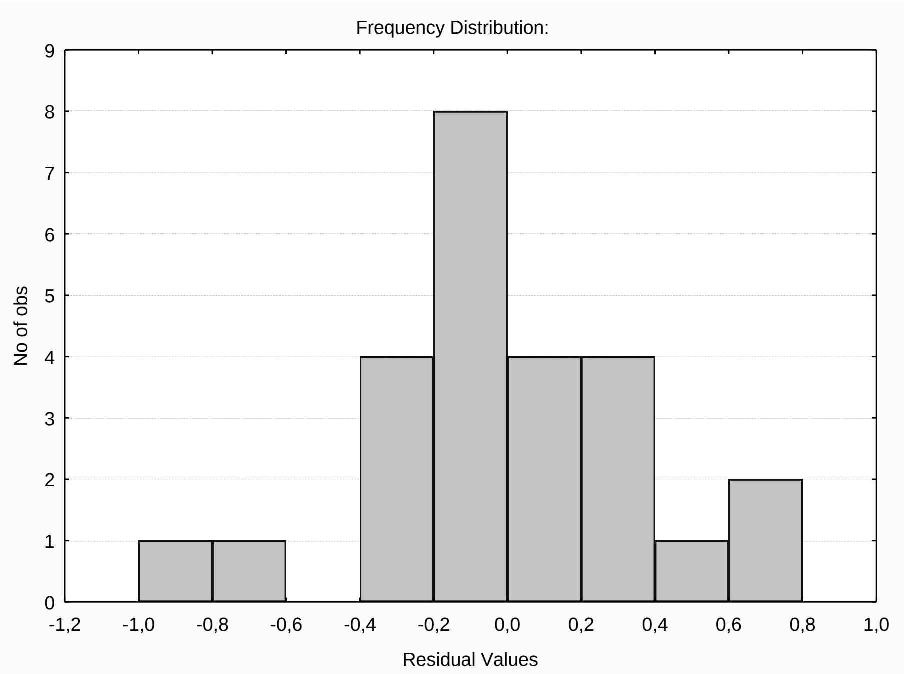


Fig. 4.19. **The histogram of the error distribution**

4. Determination of the main characteristics of the production function.

Calculation of the characteristics of the production function is possible by setting the calculation formula in the specification field of the *variable Long name* (Fig. 4.20).

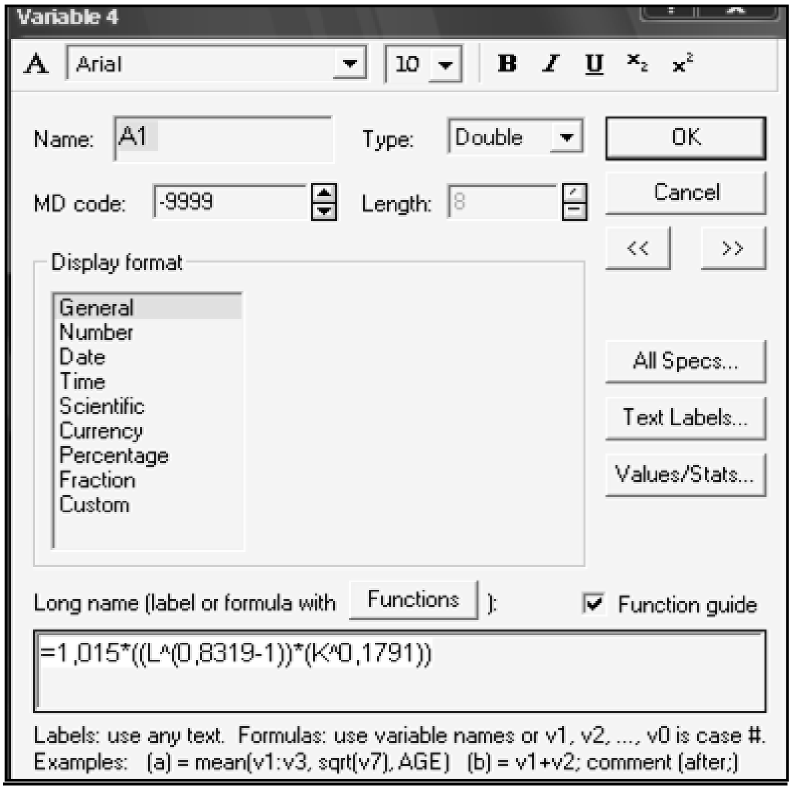


Fig. 4.20. **The histogram of the error distribution**

For further analysis it is necessary to determine the following characteristics:

1) the average productivity: $A_1 = \dfrac{y}{L} = a_0 \cdot L^{a_1 - 1} \cdot K^{a_2}$;

2) the average assets ratio: $A_2 = \dfrac{y}{K} = a_0 \cdot L^{a_1} \cdot K^{a_2 - 1}$;

3) the marginal productivity of labor: $M_1 = \dfrac{\partial y}{\partial L} = a_0 \cdot a_1 \cdot L^{a_1 - 1} \cdot K^{a_2}$;

4) the marginal capital productivity: $M_2 = \dfrac{\partial y}{\partial K} = a_0 \cdot a_2 \cdot L^{a_1} \cdot K^{a_2 - 1}$;

5) the elasticity of output per labor costs: $E_1 = \dfrac{\partial y}{\partial L} \cdot \dfrac{L}{y}$; $\quad E_1 = \dfrac{M_1}{A_1}$;

6) the elasticity of output per cost of assets: $E_2 = \dfrac{\partial y}{\partial K} \cdot \dfrac{K}{y}$; $\quad E_2 = \dfrac{M_2}{A_2}$;

7) the total expenditure elasticity: $E = E_1 + E_2$;

8) the capital-labor ratio of works: $FT = \dfrac{K}{L} = a_0^{-1/a_2} \cdot y_0^{1/a_2} \cdot L^{-1-a_1/a_2}$;

9) the capital-labor ratio of products: $F = \dfrac{K}{Y} = \dfrac{1}{a_0}\left(\dfrac{K}{L}\right)^{a_1}$;

10) the need for labor and capital: $L = \left(\dfrac{\hat{Y}}{a_0 \cdot K^{a_2}} \right)^{\frac{1}{a_1}}$, $K = \left(\dfrac{\hat{Y}}{a_0 \cdot L^{a_1}} \right)^{\frac{1}{a_2}}$;

12) the marginal rate of replacement of the *i*-th resource by the *j*-th resource:

$$R_{ij} = -\frac{\Delta x_j}{\Delta x_i}, \qquad R'_{ij} = -\frac{\partial x_j}{\partial x_i} \qquad \left| R_{12} \right| = \frac{E_1}{E_2} \cdot \frac{K}{L}.$$

The results of the calculation of the basic characteristics are shown in Fig. 4.21.

	1 L	2 K	3 Y	4 A1	5 A2	6 M1	7 M2	8 E1	9 E2	10 E	11 FT	12 F
1	16,88	21,84	18,22	1,096	0,847	0,912	0,152	0,8319	0,1791	1,01	1,294	1,199
2	16,05	22,71	17,53	1,114	0,787	0,926	0,141	0,8319	0,1791	1,01	1,415	1,295
3	14,74	20,87	16,35	1,113	0,786	0,926	0,141	0,8319	0,1791	1,01	1,416	1,276
4	15,8	25,68	17,94	1,141	0,702	0,95	0,126	0,8319	0,1791	1,01	1,625	1,431
5	17,72	20,9	19,24	1,079	0,915	0,898	0,164	0,8319	0,1791	1,01	1,179	1,086
6	15,29	26,68	17,66	1,156	0,662	0,961	0,119	0,8319	0,1791	1,01	1,745	1,511
7	16,2	22,99	18,21	1,114	0,785	0,927	0,141	0,8319	0,1791	1,01	1,419	1,262
8	16,81	23,68	18,44	1,113	0,79	0,926	0,142	0,8319	0,1791	1,01	1,409	1,284
9	17,68	25,54	19,76	1,119	0,775	0,931	0,139	0,8319	0,1791	1,01	1,445	1,293
10	16,45	22,2	17,85	1,104	0,818	0,919	0,147	0,8319	0,1791	1,01	1,35	1,244
11	17,17	26,02	19,27	1,128	0,744	0,939	0,133	0,8319	0,1791	1,01	1,515	1,35
12	17,9	27,94	20,93	1,135	0,727	0,944	0,13	0,8319	0,1791	1,01	1,561	1,335
13	16,74	26,6	19,04	1,137	0,716	0,946	0,128	0,8319	0,1791	1,01	1,589	1,397
14	14,53	24,87	16,75	1,151	0,672	0,957	0,12	0,8319	0,1791	1,01	1,712	1,485
15	17,2	25,47	19,56	1,124	0,759	0,935	0,136	0,8319	0,1791	1,01	1,481	1,302
16	16,78	22,33	17,82	1,102	0,828	0,917	0,148	0,8319	0,1791	1,01	1,331	1,253
17	14,66	20,63	16,88	1,111	0,79	0,925	0,141	0,8319	0,1791	1,01	1,407	1,222
18	14,86	25,68	17,46	1,153	0,667	0,959	0,12	0,8319	0,1791	1,01	1,728	1,471
19	16,68	24,54	19,02	1,122	0,763	0,933	0,137	0,8319	0,1791	1,01	1,471	1,29
20	18	26,94	20,11	1,126	0,753	0,937	0,135	0,8319	0,1791	1,01	1,497	1,34
21	17,06	24,68	19,32	1,119	0,773	0,931	0,139	0,8319	0,1791	1,01	1,447	1,277
22	15,95	23,2	18,04	1,119	0,769	0,931	0,138	0,8319	0,1791	1,01	1,455	1,286
23	18,2	26,47	20,22	1,121	0,771	0,932	0,138	0,8319	0,1791	1,01	1,454	1,309
24	14,72	19,9	17,01	1,103	0,816	0,918	0,146	0,8319	0,1791	1,01	1,352	1,17
25	14,05	23,71	15,13	1,148	0,68	0,955	0,122	0,8319	0,1791	1,01	1,688	1,567

Fig. 4.21. **The results of the calculation of specifications of the production function**

5. Construction of the production function isoquants.

To build the isoquants it is necessary to find combinations of the productive resources at the fixed levels of production. You must calculate the value of the index cost of a resource under certain expenses of the second resource. Construct the isoquants for the following expectations of production: (Y_{pr} = 26.5 mln UAH, Y_{pr} = 20.2 mln UAH, Y_{pr} = 28.3 mln UAH, Y_{pr} = 25.1 mln UAH), leaving the index value cost of the manpower previous formula:

$$K = \left(\frac{\hat{Y}}{a_0 \cdot L^{a_1}} \right)^{\frac{1}{a_2}} .$$

The calculation of the need for capital expenditure and the value of the limits for replacing the *i*-th resource by *j*-th resource is shown in Fig. 4.22.

	1 L	2 K	3 Y	4 K1	5 K2	6 K3	7 K4	8 R_1_2	9 R_2_1
1	16,88	21,84	18,22	162,0095	35,58604	233,8277	119,6537	6,009741	0,166397
2	16,05	22,71	17,53	204,763	44,97703	295,5337	151,2297	6,572304	0,152154
3	14,74	20,87	16,35	304,0943	66,79554	438,8981	224,5917	6,576586	0,152055
4	15,8	25,68	17,94	220,2522	48,37928	317,8891	162,6693	7,549418	0,132461
5	17,72	20,9	19,24	129,2926	28,39965	186,6076	95,49031	5,478455	0,182533
6	15,29	26,68	17,66	256,5122	56,34394	370,2231	189,4495	8,105016	0,12338
7	16,2	22,99	18,21	196,1039	43,07502	283,036	144,8344	6,591731	0,151705
8	16,81	23,68	18,44	165,167	36,2796	238,3849	121,9857	6,54319	0,152831
9	17,68	25,54	19,76	130,657	28,69933	188,5767	96,49795	6,709871	0,149034
10	16,45	22,2	17,85	182,639	40,11741	263,6022	134,8898	6,268485	0,159528
11	17,17	26,02	19,27	149,685	32,87891	216,0398	110,5513	7,039025	0,142065
12	17,9	27,94	20,93	123,3633	27,09724	178,0497	91,11113	7,250182	0,137928
13	16,74	26,6	19,04	168,3996	36,98965	243,0505	124,3731	7,380771	0,135487
14	14,53	24,87	16,75	325,0533	71,39928	469,1482	240,0712	7,95034	0,125781
15	17,2	25,47	19,56	148,4762	32,61339	214,2951	109,6585	6,87822	0,145386
16	16,78	22,33	17,82	166,5431	36,58186	240,371	123,002	6,181193	0,161781
17	14,66	20,63	16,88	311,8793	68,50554	450,1341	230,3414	6,536433	0,152989
18	14,86	25,68	17,46	292,8546	64,3267	422,6759	216,2906	8,026972	0,12458
19	16,68	24,54	19,02	171,2317	37,61175	247,1381	126,4648	6,833671	0,146334
20	18	26,94	20,11	120,212	26,40505	173,5015	88,7837	6,951854	0,143847
21	17,06	24,68	19,32	154,221	33,87526	222,5865	113,9014	6,719573	0,148819
22	15,95	23,2	18,04	210,7946	46,30188	304,239	155,6843	6,756205	0,148012
23	18,2	26,47	20,22	114,1977	25,08399	164,8211	84,34182	6,755509	0,148027
24	14,72	19,9	17,01	306,0182	67,21813	441,6748	226,0126	6,279438	0,15925
25	14,05	23,71	15,13	379,9446	83,45637	548,3725	280,6117	7,83846	0,127576

Fig. 4.22. **The results of the calculation for the construction of the isoquants**

According to the information received, the graphs of the production function isoquants are plotted. To do this, select the sub-menu *Graphs, Scatterplots* (*Point charts*). When specifying the characteristics of the graph select the *Graph type – Multiple* (*Advanced*), remove the evaluation *Linear fit*, as variables reflected in the chart, select the value of human resources (*L*) on the *X* axis and the calculated values of the need for fixed assets (*K*1 – *K*4) along the axis *Y* as shown in Fig. 4.23.

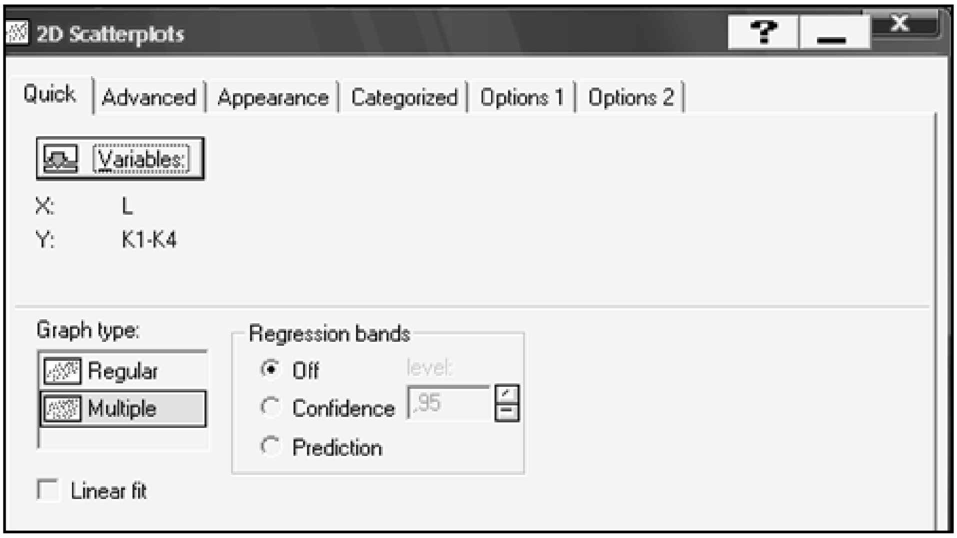


Fig. 4.23. **Determination of the graph**

The graph of the production function isoquants is shown in Fig. 4.24.

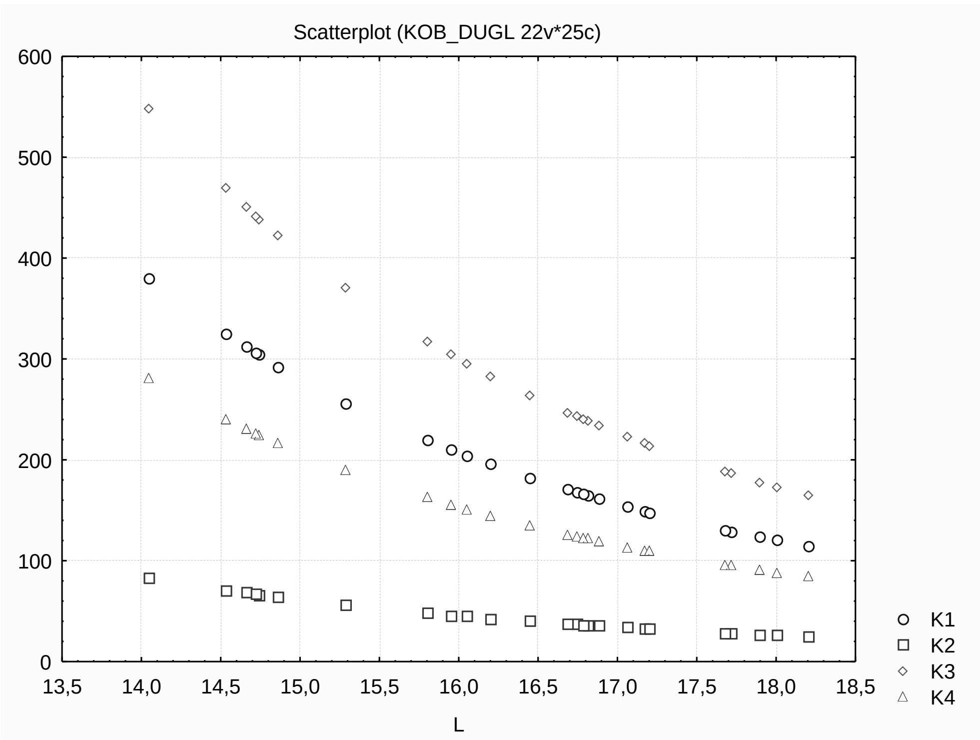


Fig. 4.24. **The graph of the production function isoquants**

The isoquant of the production function is the line of $q = f\ (L, K), (q > 0)$, which is the set of points where VF takes the value equal to q (output). Thus, the isoquant is a different set (value) of resources that provide the same volume of output. In this graph, you can also draw an isoclinal at any point and calculate the marginal rate of substitution of resources.

Lab session 5. Econometric Models of Dynamics

The goal is to consolidate the theoretical and practical material on the theme "Econometric Models of Dynamics", acquire skills in the construction and analysis of dynamic econometric models for the research on real economic processes in the module *Nonlinear Estimation / Advanced Linear / Nonlinear Models*.

The task is to test the module *Nonlinear Estimation / Advanced Linear / Nonlinear Models* for a nonlinear trend.

Table 5.1 shows the data, reflecting the dynamics of the net profits of industrial enterprises in the period of 10 years ($\mathrm{Y_t}$).

Table 5.1

The dynamics of the net profit (mln UAH)

t	1	2	3	4	5	6	7	8	9	10
Y_t	2	2.5	2.2	2.5	3.4	3.7	4	4.2	6	9.7

Do the following tasks:

1. Construct a graph of Y_t vs t.

2. Based on the analysis of the graph make assumptions about the type of the functional dependence between the indicators Y_t and t.

3. Using the module *Nonlinear Estimation / Advanced Linear/Nonlinear Models* (*Statistica*) build 3 most likely dependences of Y_t on the time.

4. Analyze the quality of the constructed models and select the best model (assess the statistical significance of the model parameters of the trend, analyze the correlation coefficients, etc.).

5. Provide quality estimates of the various models of the trend (the mean error, the mean absolute error, the standard deviation of errors, the mean percentage error, the mean absolute percentage error). Make a comparative analysis of the models and determine the most adequate ones among them.

6. Make a forecast for the next 2 years. Provide an economic interpretation of the results.

Guidelines

Consider the process of selection and construction of nonlinear trend model in *Statistica.*

1. Construct a graph of Y_t vs t.

Form a table of the raw data in *Statistica* (Fig. 5.1). For plotting the functional dependence $Y = f(t)$ select the menu item *Graphs* / *Scatterplots.* In the window that appears, choose the type of graph *Regular*, and pressing the *Variables*, choose the variable Y_t on the vertical axis, variable t on the horizontal axis (Fig. 5.1).

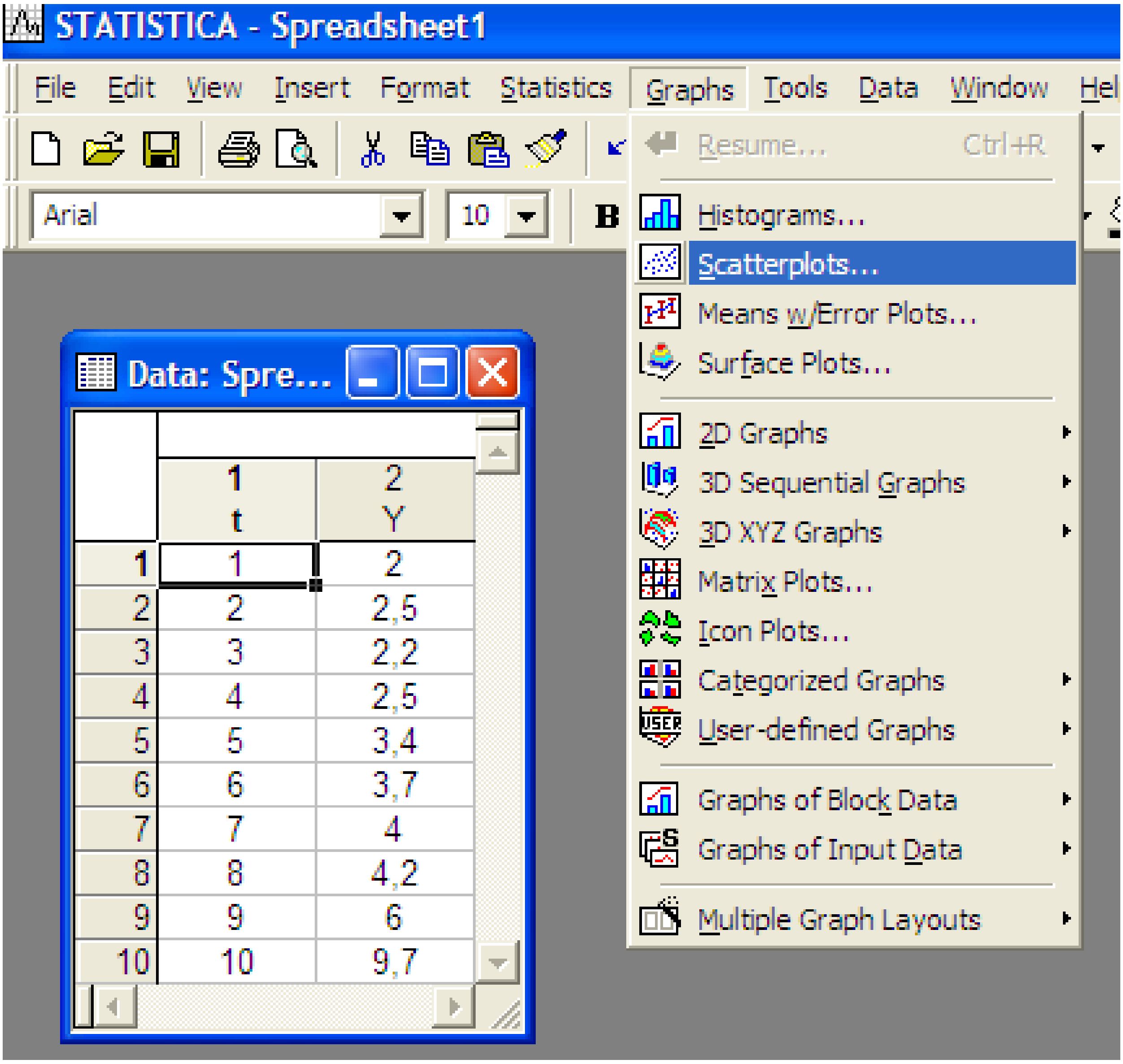


Fig. 5.1. **The output data**

The result is a graph showing the dependence of Y_t on time (Fig. 5.2).

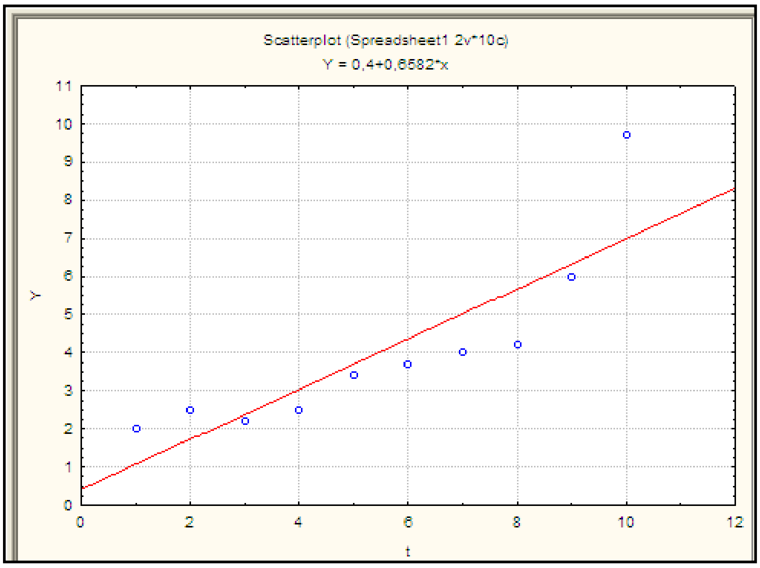


Fig. 5.2. **The graph of the output data**

The analysis of the graph of the output data leads to the conclusion that the following trend growth are curves considered as competing models: the exponential function ($y = a_0 a_1^t$); the power function ($y = a_0 t^{a_1}$); the second-degree polynomial ($y = a_0 + a_1 t + a_2 t^2$).

Find the parameters of the competing trend models and evaluate their predictive properties.

Consider the process of construction of the exponential function ($y = a_0 a_1^t$). To estimate the model parameters, select the menu item *Statistics / Nonlinear Estimation / Advanced Linear/Nonlinear Models* (Fig. 5.3).

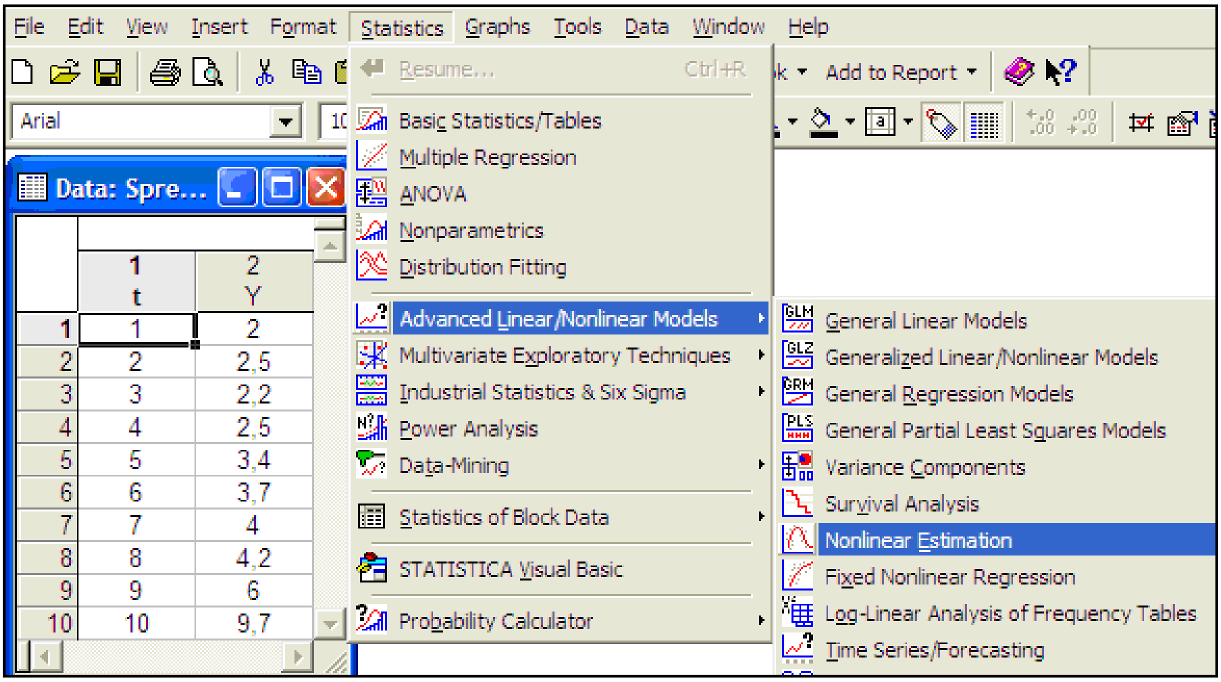


Fig. 5.3. **The non-linear estimation module**

The type of the starting panel module is shown in Fig. 5.4.

To solve this problem, you must select the option *user-specified regression, least squares* (Fig. 5.4).

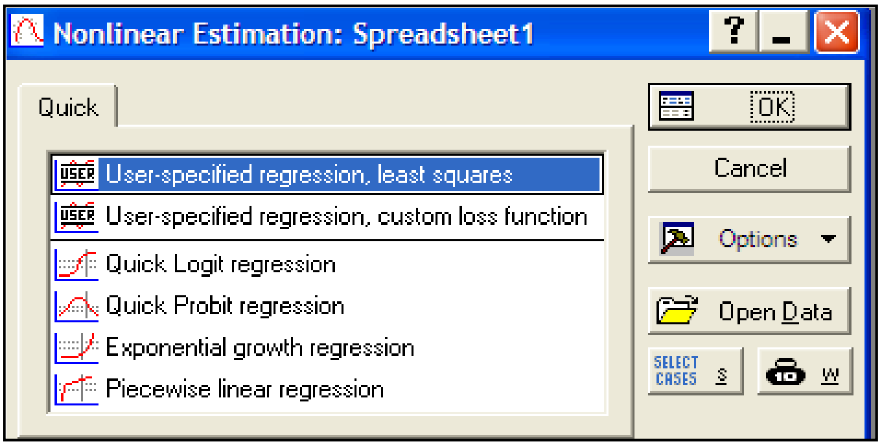


Fig. 5.4. **The starting panel of the module *Nonlinear Estimation***

Then in the window *Function to be estimated,* set the kind of the estimated trend model: $(y = a_0 a_1^t)$ (Fig. 5.5). We should use special mathematical marks that are similar to a set of formulas in *Excel*. For example, the exponential function must be defined as: *Y = a0×a1× t* or *Y = a0×a1^t.*

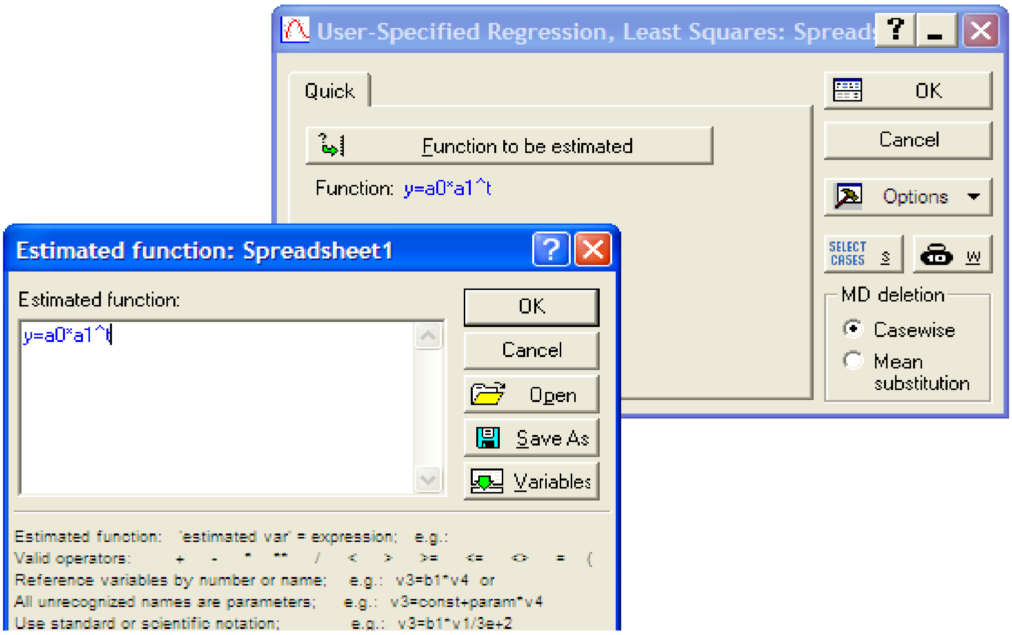


Fig. 5.5. **Determining the type of function**

In the following window specify the method of estimation of parameters of the nonlinear trend model (*Levenberg – Marquardt* or *Gauss – Newton*) if necessary, under *Advanced* settings the iterative procedure and the initial values of the estimated parameters can be defined. Run the assessment procedure (Fig. 5.6).

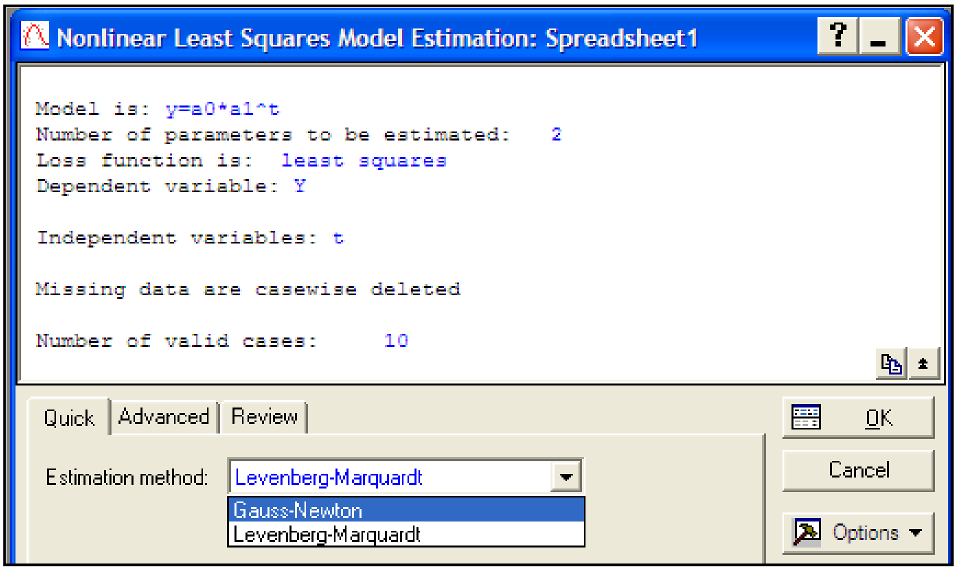


Fig. 5.6. **Selecting the method for estimating the model parameters**

The results of the construction of the exponential trend model can be analyzed with the following window, shown in Fig. 5.7.

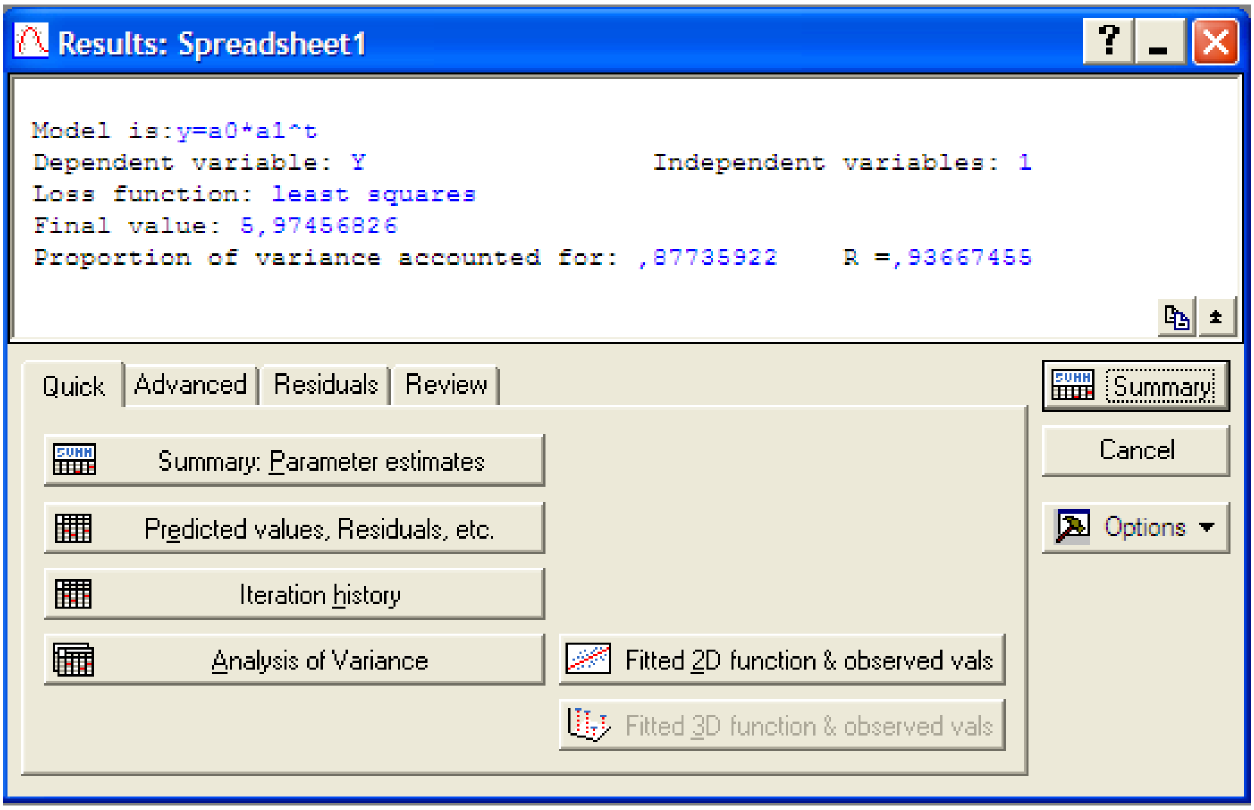


Fig. 5.7. **The window of the results of constructing the exponential trend model**

This window presents the results of the model (the model type, dependent and independent variables, the type of the function evaluation), in the bottom of the information part the options for a comprehensive analysis of the model are given.

The quality of the resulting model is estimated using the values: *Final value of Loss function (the sum of the squares of the error model), Proportion of variance accounted for (the percentage of the explained variance), R (the correlation coefficient).* The estimates of the model parameters can be obtained by initiating the option *Quick / Summary: Parameter estimates* (Fig. 5.7). The result of the evaluation of the parameters is shown in Fig. 5.8.

Model is: y=a0*a1^t (Spreadsheet1)
Dep. Var. : Y
Level of confidence: 95.0% (alpha=0.050)

	Estimate	Standard error	t-value df = 8	p-level	Lo. Conf Limit	Up. Conf Limit
a0	1,101189	0,273196	4,03076	0,003784	0,471197	1,731181
a1	1,224090	0,036037	33,96767	0,000000	1,140989	1,307192

Fig. 5.8. **The result of evaluating the model parameters**

In this window, there are the estimates of the model parameters (*Estimates*), the values of the *Standard error*, the Student's Statistics (*t-value*), the level of Student's statistical significance (*p-level*), the lower and upper limits of the interval parameter estimates (*Lo. Conf Limit; Up. Conf Limit*).

For further analysis it is necessary to obtain the theoretical values of the dependent variable and the model errors (residuals) by initiating the key *Predicted values, Residuals* under *Quick*. The results are shown in Fig. 5.9.

We have to construct a graph to compare the empirical (*Observed*) and the theoretical (*Predicted*, calculated by the model) values of the dependent variable. Select the options of the graphical analysis of the results. These are shown in Fig. 5.10. In Fig. 5.11 there is a graph comparing the theoretical and empirical values of the dependent variable.

	Observed	Predicted	Residuals
1	2,000000	1,347955	0,65205
2	2,500000	1,650018	0,84998
3	2,200000	2,019772	0,18023
4	2,500000	2,472383	0,02762
5	3,400000	3,026420	0,37358
6	3,700000	3,704612	-0,00461
7	4,000000	4,534780	-0,53478
8	4,200000	5,550981	-1,35098
9	6,000000	6,794902	-0,79490
10	9,700000	8,317575	1,38243

Fig. 5.9. **The output, the predicted data and the model errors**

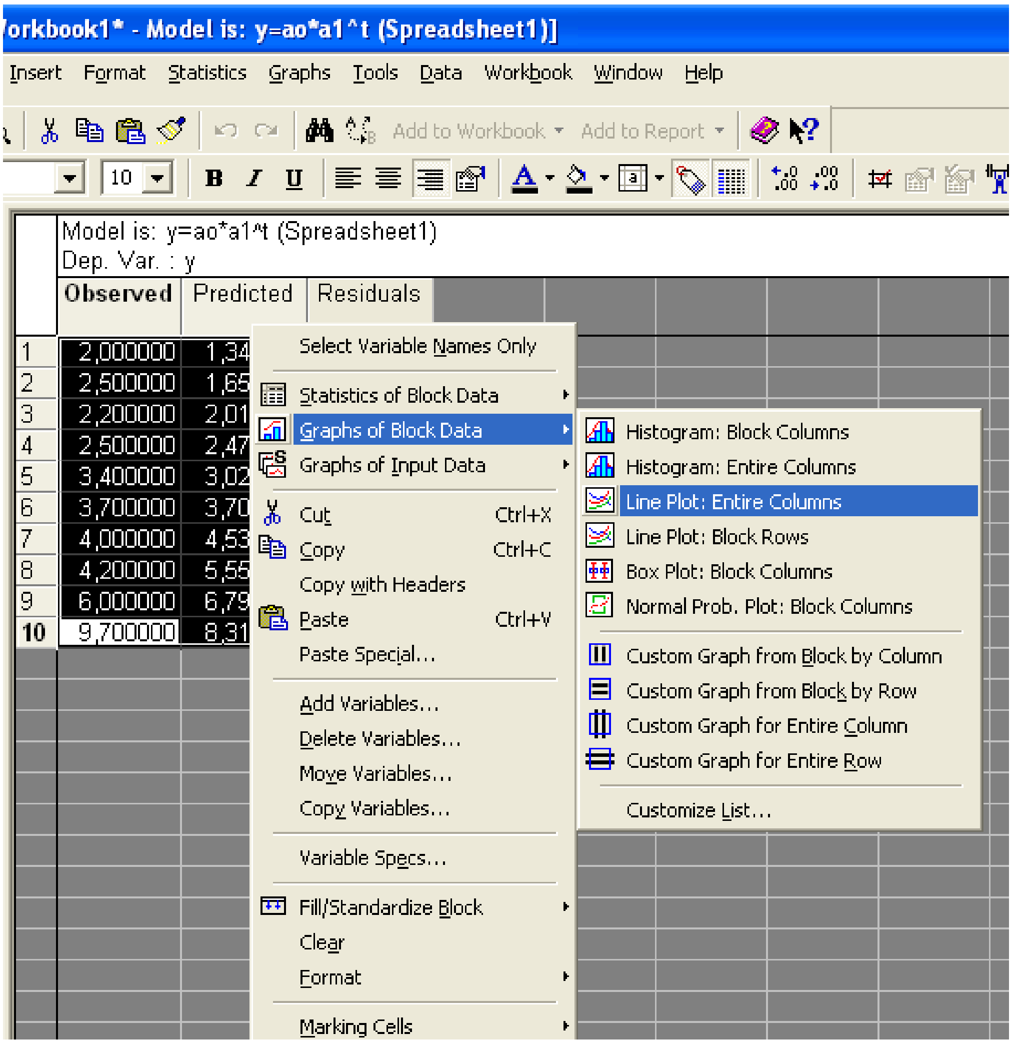


Fig. 5.10. **Selecting the options of the graphical analysis variables**

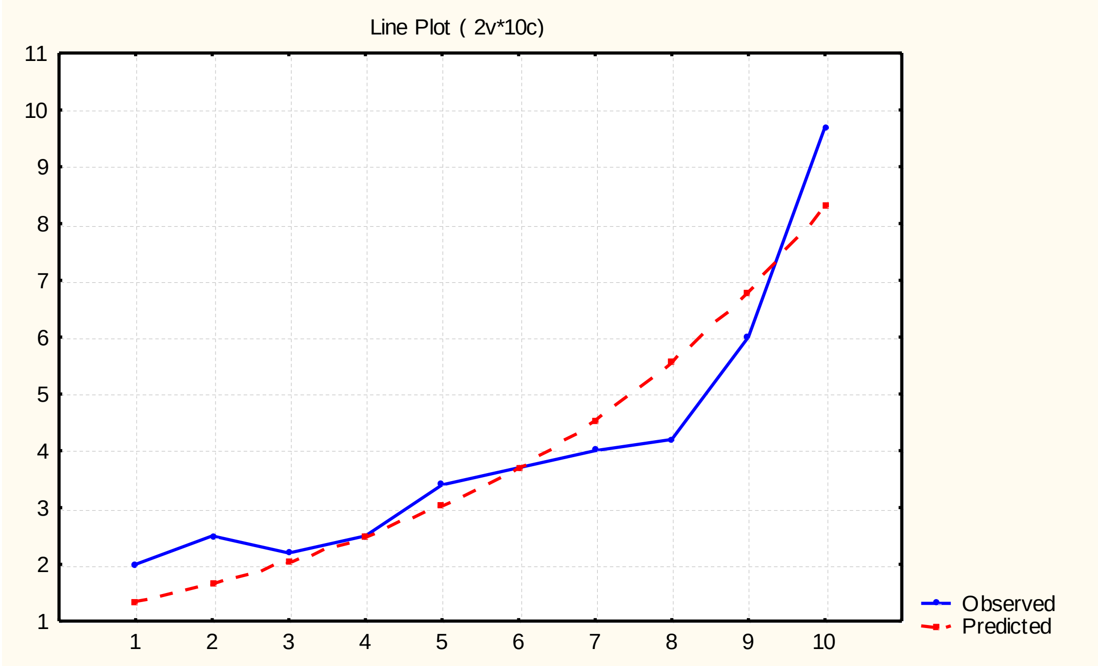


Fig. 5.11. **The graph of the theoretical and empirical values of the dependent variable**

A comprehensive model of the error analysis is available in the options *Residuals* (Fig. 5.12).

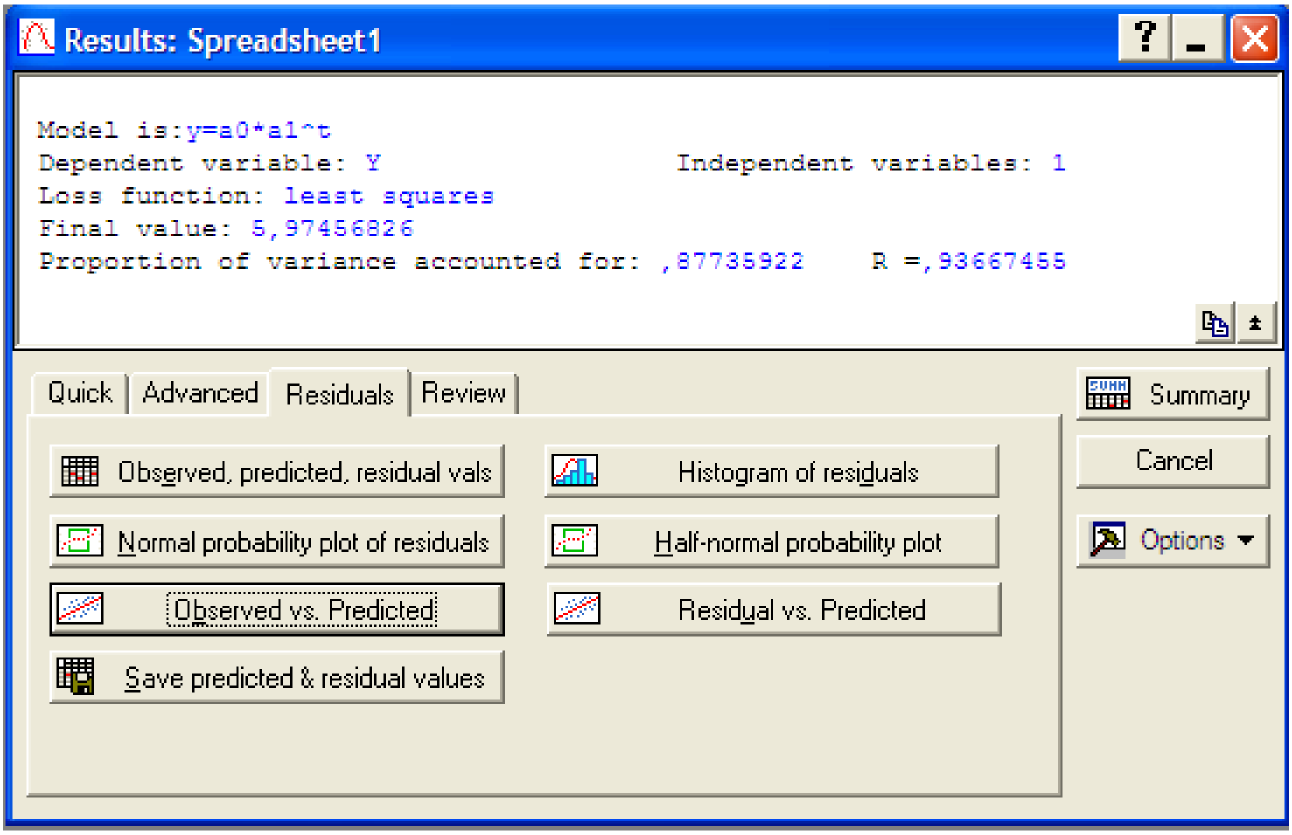


Fig. 5.12. **The menu of the error analysis model**

The histogram of errors and the graph of the distribution of errors on the normal probability paper are shown in Fig. 5.13 and 5.14.

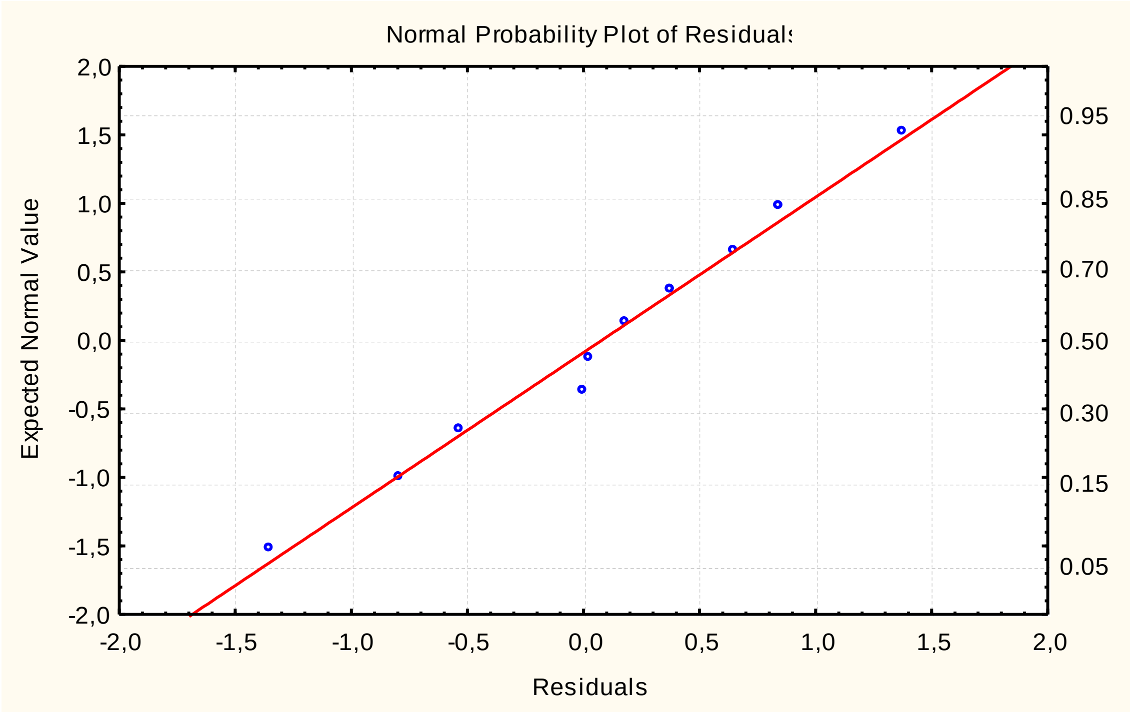


Fig. 5.13. **The graph of the distribution of errors on the normal probability paper**

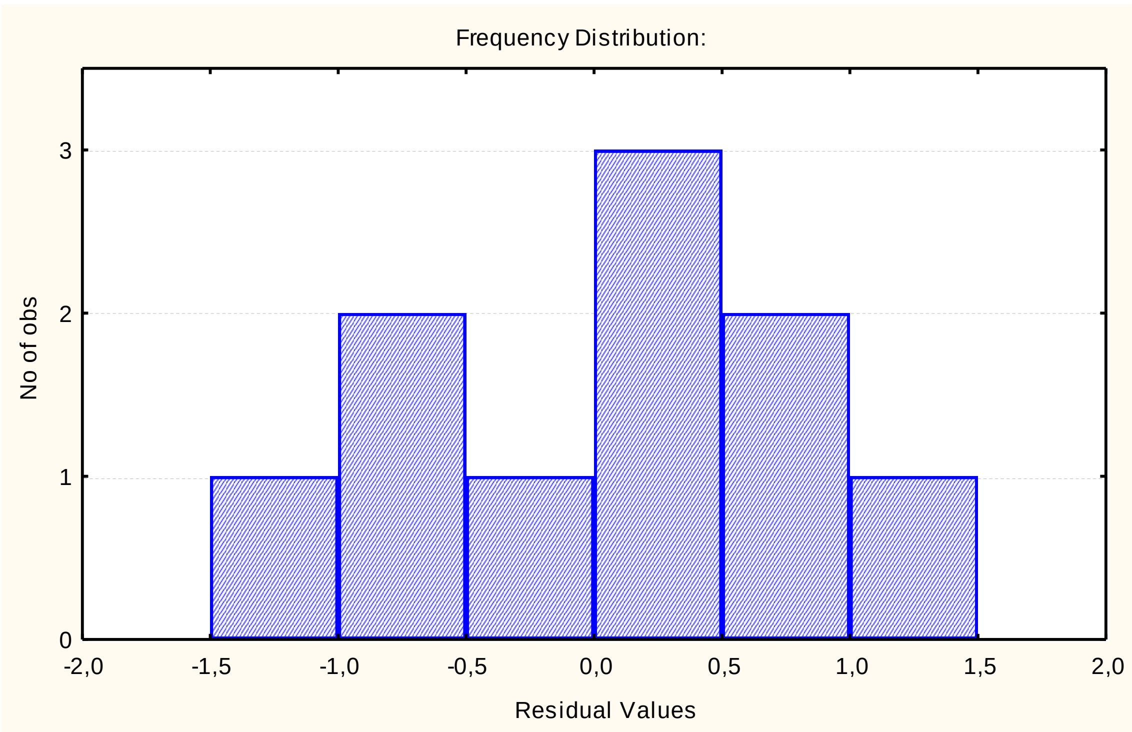


Fig. 5.14. **The histogram of the distribution of errors**

Thus, the exponential trend model of the investigated time series has the form:

$$y = 1.101 \cdot 1.224^t.$$

Based on the fact that the correlation coefficient for the model is $R = 0.93667$, the explained variance percentage is 87.73 %, the model parameters got by Student's t-test are statistically significant, we can conclude that the model can be used for prediction.

Similarly, we define the characteristics of the power trend model and the polynomial trend model. The comparative analysis of the competing trend models is shown in Table 5.2.

Table 5.2

The comparative analysis of the nonlinear trend models

Model	The value of Student's t-test	Coefficient correlation	The percentage of explained variance
$y = 1.101 \cdot 1.224^t$	$t_{a_0} = 4.03$; $t_{a_1} = 33.97$	0.93667	87.73 %
$y = 0.5119 \cdot t^{1.1649}$	$t_{a_0} = 1.59$; $t_{a_1} = 3.83$	0.85614	73.29 %
$y = 3.1417 - 0.7127 \cdot t + 0.1246 \cdot t^2$	$t_{a_0} = 3.23$; $t_{a_1} = 1.76$; $t_{a_2} = 3.47$	0.94970	90.19 %

The analysis of the data presented in Table 5.2 suggests that the polynomial trend model has the highest predictive quality.

Find the predicted net profit value for this model. To do this, add in the primary data table (Fig. 5.1) the number of lines equal to the predicted period (2) and an additional column $\hat{y}$ to show the prediction for the model (Fig. 5.15). The values of the predictive net profit for 2 years can be calculated by setting the variable in the specification Long Name of the constructed trend model (Fig. 5.15). The results of the calculation of the predictive values of the net profit are shown in Fig. 5.15.

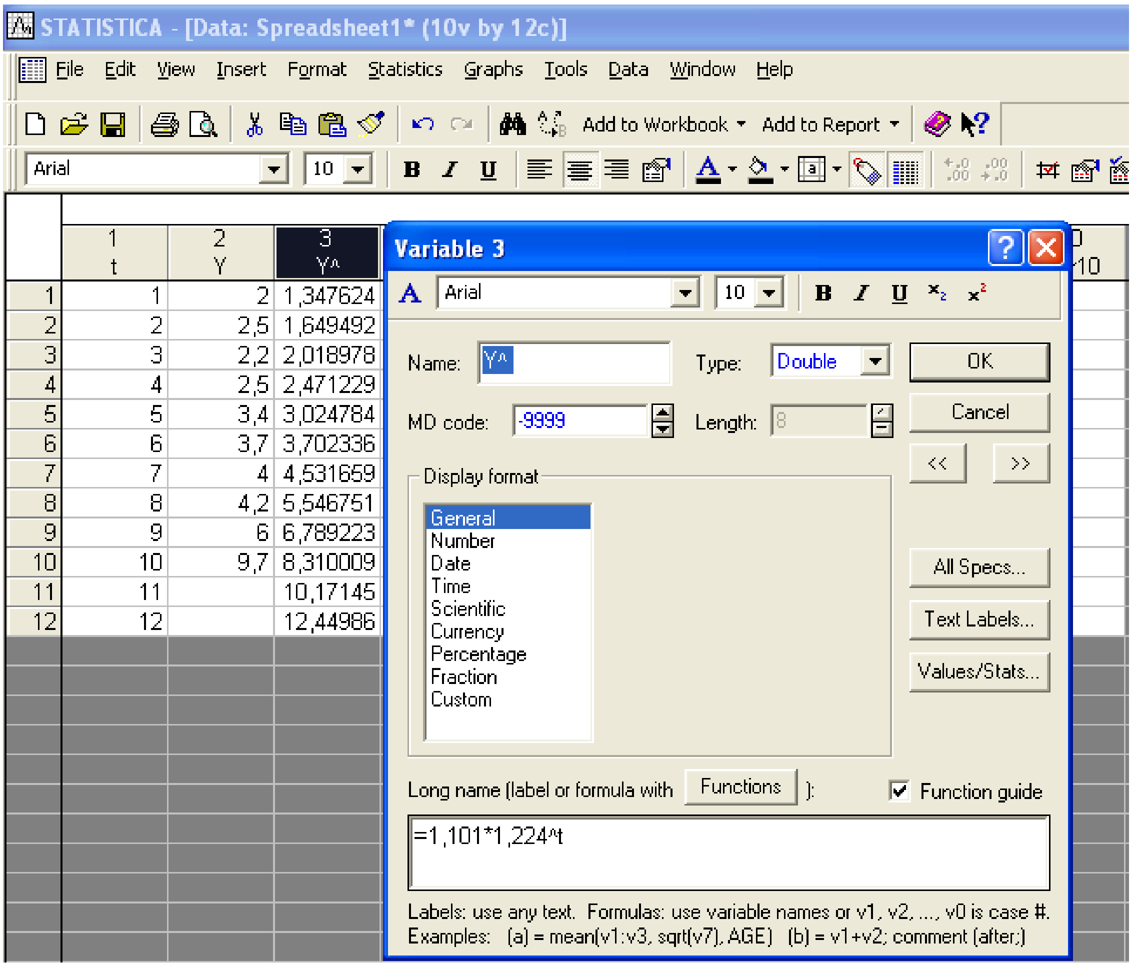


Fig. 5.15. **Building a predictive net profit for 2 coming years**

5. The quality of the estimates of the various trend models.

To explore the quality estimates of the time series model, such parameters as the mean error, the mean absolute error, the standard deviation of errors, the mean percentage error, the mean absolute percentage error, the new *Residuals* (*Model errors*) variable should be entered and calculated by setting the formula *Residuals = Sales Volume – Predict*.

The formulas for calculating the estimates of the model accuracy are shown in Table 5.3.

Table 5.3

Characteristics of the estimation of the model accuracy

Name	Formula
1	2
Mean error	$m.e. = \frac{\sum_{t=1}^{n} e_t}{n}$

1	2
Mean absolute error	$m.a.e. = \frac{\sum_{t=1}^{n} \lvert e_t \rvert}{n}$
Sum of squares of errors	$s.s.e. = \sum_{t=1}^{n} e_t^2$
Mean squared error	$m.s.e. = \sqrt{\frac{\sum_{t=1}^{n} e_t^2}{n}}$
Mean percentage error	$m.p.e. = \frac{1}{n} \sum_{t=1}^{n} \frac{e_t}{y_t} \cdot 100\,\%$
Mean absolute percentage error	$m.a.p.e. = \frac{1}{n} \sum_{t=1}^{n} \frac{\lvert e_t \rvert}{y_t} \cdot 100\,\%$

Draw a comparison of the models for the value of the average absolute percentage error. If the value of the average absolute percentage error is in the range:

0 < m.a.p.e < 10 %, the model provides a high prediction accuracy;

10 % < m.a.p.e < 20 %, the model provides a satisfactory accuracy of the forecast;

m.a.p.e > 20 %, the model is not adequate.

Recommended literature

11.1. Basic

1. Боровиков В. П. Популярное введение в программу STATISTICA / В. П. Боровиков – Москва : Компьютер Пресс, 1998. – 194 с.

2. Доугерти К. Введение в эконометрику / К. Доугерти ; пер. с англ. – Москва : ИНФРА-М, 1997. – 402 с.

3. Економетрика : навчальний посібник / Л. С. Гур'янова, Т. С. Клебанова, О. А. Сергієнко та ін. – Харків : Вид. ХНЕУ ім. С. Кузнеця, 2015. – 389 с.

4. Клебанова Т. С. Эконометрия / Т. С. Клебанова, Н. А. Дубровина, Е. В. Раевнева. – Харків : ИД "ИНЖЭК", 2003. – 128 с.

5. Наконечний С. І. Економетрія / С. І. Наконечний, Т. О. Терещенко, Т. П. Романюк. – Київ : КНЕУ, 1997. – 352 с.

6. Прогнозування соціально-економічних процесів : навчальний посібник / Т. С. Клебанова, В. А. Курзенев, В. М. Наумов та ін. – Харків : Вид. ХНЕУ ім. С. Кузнеця, 2015. – 656 с.

7. Эконометрия на персональном компьютере / Т. С. Клебанова, Н. А. Дубровина, А. В. Милов и др. – Харьков : Изд. ХГЭУ, 2002. – 208 с.

11.2. Additional

8. Боровиков В. П. STATISTICA: искусство анализа данных на компьютере. Для профессионалов / В. П. Боровиков. – Санкт-Петербург : Питер, 2001. – 656 с.

9. Лук'яненко І. Економетрика / І. Лук'яненко, Л. Краснікова. – Київ : Товариство "Знання", КОО, 1998. – 494 с.

10. Магнус Я. Р. Эконометрика. Начальный курс / Я. Р. Магнус, П. К. Катышев, А. А. Пересецкий. – Москва : Дело, 1997. – 248 с.

11. Орлов. А. Н. Эконометрика / А. Н. Орлов. – Москва : Изд. "Экзамен", 2002. – 576 с.

12. Черняк О. І. Динамічна економетрика / О. І. Черняк, А. В. Ставицький. – Киев : КВІЦ, 2000. – 120 с.

13. Hansen Bruce E. Econometrics / Bruce E. Hansen. – Wisconsin : University of Wisconsin, 2016. – 381 p.

14. Hubler O. Modern Econometric Analysis: Surveys on Recent Developments / O. Hubler, J. Frohn. – S. l. : Springer, 2006. – 234 p.

15. Verbeek M. A Guide to Modern Econometrics / M. Verbeek. – 2nd ed. – S. l. : Wiley, 2004. – 446 p.

16. Wooldridge J. M. Intoductory Econometrics : A Modern Approach / J. M. Wooldridge. – 4th edition. – Thompson : South-Western College Publishing, 2012. – 912 p.

11.3. Information resources

17. Економетрика – бібліотека ресурсів [Електронний ресурс]. – Режим доступу : http://efaculty.kiev.ua/ekon.htm.

18. Статистика України : науковий журнал [Електронний ресурс]. – Режим доступу : www.ukrstat.gov.ua.

19. Сайт Национального банка Украины [Електронний ресурс]. – Режим доступа : www.bank.gov.ua.

20. Сайт ПФТС [Электронний ресурс]. – Режим доступа : http://pfts.com.

21. Сайт Агентства по развитию инфраструктуры фондового рынка Украины [Электронный ресурс]. – Режим доступа : http://www.smida.gov.ua/db.

11.4. Methodical support

22. Guryanova L. Guidelines to laboratory sessions of the academic discipline "Mathematical modeling in economics and management: Econometrics" [Electronic resource] / L. Guryanova, O. Sergienko, S. Prokopovych. – Access mode : http://elearn2.ekhneu.org.ua/main/document/document.php?cidReq=MMMECONOMETR&id_session=0&gidReq=0&origin=.

23. Sergienko O. Tests on the academic discipline "Mathematical modeling in economics and management: Econometrics" [Electronic resource] / O. Sergienko, S. Milevskyi, S. Prokopovych. – Access mode : http://elearn2. ekhneu.org.ua/main/exercice.

24. Prokopovych S. Mathematical modeling in economics and management: Econometrics : reference compendium [Electronic resource] / S. Prokopovych, O. Sergienko, S. Milevskyi. – Access mode : http://elearn2.ekhneu.org.ua/main/document/document.php?cidReq=MMMECONOMETR&id_session=0&gidReq=0&origin=&id=53.

НАВЧАЛЬНЕ ВИДАННЯ

Методичні рекомендації до лабораторних робіт з навчальної дисципліни "МАТЕМАТИЧНЕ МОДЕЛЮВАННЯ В ЕКОНОМІЦІ ТА МЕНЕДЖМЕНТІ: ЕКОНОМЕТРІЯ"

для студентів напряму підготовки
6.030601 "Менеджмент"
денної форми навчання
(англ. мовою)

Укладачі: **Гур'янова** Лідія Семенівна
Прокопович Світлана Валеріївна
Сергієнко Олена Андріанівна

Відповідальний за випуск *Т. С. Клебанова*

Редактор *З. В. Зобова*

Коректор *З. В. Зобова*

Розглянуто основні питання аналізу і прогнозування соціально-економічних і фінансових процесів та систем на основі застосування економетричних методів і моделей. Наведено методичні рекомендації до виконання лабораторних робіт з навчальної дисципліни за допомогою програми STATISTICA.

Рекомендовано для студентів економічних спеціальностей.

План 2016 р. Поз. № 89.
Підп. до друку 08.12.2016 р. Формат 60 х 90 1/16. Папір офсетний. Друк цифровий.
Ум. друк. арк. 4,25. Обл.-вид. арк. 5,31. Тираж 50 пр. Зам. № 287.

Видавець і виготовлювач – ХНЕУ ім. С. Кузнеця, 61166, м. Харків, просп. Науки, 9-А

Свідоцтво про внесення суб'єкта видавничої справи до Державного реєстру
ДК № 4853 від 20.02.2015 р.