CONDUCTING MODEL DIAGNOSTICS: CHECKING THE STATISTICAL CONSISTENCY OF VARIABLES

Rakhimova Umida Ziyadullayevna
Samarkand economy and service Institute of Higher Education
Senior lecturer of the Department of Mathematics
E-mail: rakhimovaumida2018@gmail.com
Suyunova Farangiz Zoir kizi
E-mail: suyunova2006@gmail.com

ARTICLE INFORMATION

ARTICLE HISTORY:
Received: 01.05.2026
Revised: 02.05.2026
Accepted: 03.05.2026

KEYWORDS:
Model diagnostics, statistical fit, econometrics, regression model, t-statistics, p-value, F-test, coefficient of determination (R^2), multicollinearity, heteroskedasticity, autocorrelation, Durbin–Watson test, VIF, statistical significance.

ANNOTATION

This article extensively discusses the issues of conducting model diagnostics and checking the statistical consistency of variables in econometric models . The main focus is on diagnostic methods used to assess the reliability and quality of the model, in particular, determining the statistical significance of variables.

In the research process, the influence of variables on the outcome indicator was analyzed based on the regression model, and their significance was assessed using t-statistics and p-values. Also, the overall fit of the model was checked using the F-test, and the explanatory power of the model was determined using the coefficient of determination.

As a result of practical analyses, it was found that some variables were statistically significant, while others were insignificant, justifying the need to improve the model. In addition, methods for identifying and eliminating problems such as multicollinearity, heteroscedasticity, and autocorrelation were considered.

The results of the study show that model diagnostics is an important stage in econometric analysis, which can increase the accuracy and reliability of the model. This approach is of great importance in conducting scientific research and making informed economic decisions.

Introduction: In modern economics, one of the pressing issues is the identification of the interrelationships between economic indicators and their quantitative assessment. Economic processes are formed under the influence of many factors, and the identification of their interaction is of great importance in conducting scientifically based economic policy.

Therefore, econometric methods, in particular regression analysis, are widely used to identify cause-and-effect relationships between economic phenomena.

Regression analysis is an important scientific tool that allows you to determine the relationship between the resulting indicator and the factors affecting it, assess the level of their influence, and forecast economic processes. With the help of this method, complex processes in the economic system are represented by a simplified mathematical model, and the results are interpreted on a scientific basis.

However, it is not enough to build an econometric model, it is also important to assess its quality and verify its reliability. In this context, model diagnostics, that is, checking the statistical validity of the model and the adequacy of the variables, is of particular relevance. In applied research, problems such as statistical insignificance of variables, multicollinearity, heteroscedasticity, and autocorrelation often arise. These shortcomings can negatively affect the accuracy and reliability of the model results.

Therefore, determining the statistical relevance of variables, assessing their significance in the model, and validating the model through diagnostic tests is an important scientific and practical task. Through this process, the model is improved, redundant or insignificant variables are eliminated, and as a result, a more reliable econometric model is formed.

The main purpose of this article is to conduct econometric model diagnostics, assess the statistical consistency of variables, and analyze ways to increase the reliability of the model on a scientific basis. The results of the research are of great importance for further improving econometric models, in-depth study of economic processes, and making informed management decisions.

Literature review and methodology: Regression analysis is one of the main methods of econometrics, which is widely used to identify and quantify the relationship between variables. In the scientific literature, this method is recognized as an effective tool for analyzing economic processes. In particular, the multivariate regression model is of great importance in the study of complex economic systems, allowing to assess the impact of several factors on the resulting indicator at the same time.

At the same time, in modern econometric research, it is important to check not only the structure of the model, but also its statistical correctness, that is, to conduct model diagnostics. Scientific sources emphasize that in order to ensure the reliability of the regression model, it is necessary to comply with classical conditions. In particular, it is important that the mathematical expectation of errors is equal to zero, the variance is constant (homoscedasticity), the absence of autocorrelation between errors, and the absence of a strong linear relationship between independent variables (multicollinearity).

In this study, a multivariate linear regression model was selected to determine the statistical relevance of the variables and assess the quality of the model. This model is expressed in the following general form:

$$Y = \beta_0 + \beta_1 X_1 + \beta_2 X_2 + \cdots + \beta_n X_n + \varepsilon$$

Here, **Y** is the outcome variable, $\mathbf{X_i}$ is the independent variable, $\boldsymbol{\beta_i}$ is the regression coefficient, which estimates the effect of each factor on the outcome, with other factors held constant, and $\boldsymbol{\varepsilon}$ denotes the random error.

The method of least squares (MLS) was used to estimate the model parameters. This method is based on minimizing the sum of squares of the differences between the observed and theoretical values:

$$\sum_{i=1}^{n} (Y_i - \hat{Y}_i)^2 \rightarrow min$$

The EKKU method provides the best linear unbiased estimates (BLUE) when the conditions of the classical regression model are met.

The coefficient of determination is used to assess the quality of the model:

$$R^2 = 1 - \frac{\sum (Y_i - \hat{Y}_i)^2}{\sum (Y_i - \bar{Y})^2}$$

In addition, the study paid special attention to model diagnostics. In particular: the overall statistical significance of the model was assessed using F-statistics ; the significance of individual variables was checked using t-statistics and p-values ; the degree of multicollinearity was determined and assessed using VIF indicators ; the problem of heteroscedasticity was checked using the Breusch–Pagan and White tests ; the presence of autocorrelation was analyzed using the Durbin–Watson test .

Using these diagnostic methods, the statistical fit of the model, that is, the significance of the impact of variables on the outcome indicator, and the overall reliability of the model were assessed.

Thus, the chosen methodology allows not only to identify the relationship between variables, but also to conduct in-depth diagnostics of the model, identify and improve its shortcomings. This significantly increases the accuracy and scientific validity of the results of econometric analysis.

Practical part:

Year	Y(YIAM)	X1 Investment	X2 Consumption	X3 Export	X4 State cost	X5 Inflation (%)	X6 Employment (%)
2015	128000	30000	65000	18000	16000	6.2	91.8
2016	145500	33500	72000	20500	17800	5.8	92.0

2017	168200	38000	81000	23800	20500	6.1	92.4
2018	195600	43000	90500	27500	23900	7.5	92.7
2019	223400	49000	101200	31800	27800	14.0	93.0
2020	240800	51000	108000	30000	29500	10.8	92.6
2021	278900	58500	122000	36000	34000	9.5	93.2
2022	325000	67000	140500	42000	39200	11.7	93.8
2023	372500	76000	159000	48500	45000	9.3	94.3
2024	415000	84500	176000	54000	51000	8.6	94.7

Artificial data set for macroeconomic regression analysis

In this study, a multivariate linear regression model was constructed based on macroeconomic indicators. Gross domestic product (GDP) was selected as the dependent variable, and investment, consumption, exports, government spending, inflation, and employment were selected as independent variables.

Many factorial regression model general in appearance as follows is expressed as :

$$Y = \beta_0 + \beta_1 X_1 + \beta_2 X_2 + \beta_3 X_3 + \beta_4 X_4 + \beta_5 X_5 + \beta_6 X_6 + \varepsilon$$

Practical calculations as a result following regression equation received :

$$Y = -458405.812 + 4.506X_1 + 1.168X_2 - 2.017X_3 - 0.787X_4 + 16.777X_5 + 4612.749X_6$$

Here:

- Y– gross internal product (GDP);
- X_1 – investment;
- X_2 – consumption;
- X_3– export;
- X_4- state expenses ;
- X_5– inflation level (%);
- X_6— employment rate (%).

the model results see determination coefficient R^2= 0.999933875 is organization This result is obtained through the model . consequential 99.99 percent of the variable explanation indicates , that is, the model is high at the level flexible is considered .

Model's general statistic significance F– statistic using was checked . The calculated F–statistic value equal to 7560.89 to be , to be significance level very small This ($p < 0.05$)indicates that the model is generally statistically reliable .

The significance of individual parameters was assessed using t-statistics and p-values. According to the results, it was found that the p-values of all independent variables included in the model were greater than 0.05.

In particular, for investment (X_1) p = 0.256, for consumption (X_2) p = 0.342, for exports (X_3) p = 0.498, for government spending (X_4) p = 0.805, for inflation (X_5) p = 0.958, and for employment (X_6) p = 0.782.

These results show that none of the variables included in this model are statistically significant individually ($p > 0.05$). This indicates that there may be a strong correlation between the independent variables, i.e., multicollinearity.

Results and analysis : The regression analysis showed that the model was highly adaptive. The coefficient of determination $R^2 = 0.99993$, which means that almost 99.99 percent of the resulting variable – gross domestic product (GDP) – is explained by the factors included in the model. This indicates that the model has an extremely high accuracy.

The overall statistical significance of the model was assessed using the F-statistic . The calculated F = 7560.88 value and the correspondingly very small Significance F = 2.35E-06 confirm that the model is statistically reliable in general. This means that the selected regression model is effective in explaining the resulting indicator.

The effects of individual factors were analyzed using t-statistics and p-values . According to the results, all factors – investment (X_1), consumption (X_2), exports (X_3), government spending (X_4), inflation (X_5) and employment (X_6) – were found to be statistically insignificant ($p > 0.05$).

In particular:

- X_1 (investment) – p = 0.2567
- X_2 (consumption) – p = 0.3425
- X_3 (export) – p = 0.4983
- X_4 (government spending) – p = 0.8050
- X_5 (inflation) – p = 0.9586
- X_6 (employment) – p = 0.7828

This does not mean that these factors have no effect at all, but rather that they are strongly correlated or that there may be a multicollinearity problem in the model .

It should be noted that a very high coefficient of determination and statistically insignificant results for individual variables increase the likelihood of multicollinearity in the model . This indicates a strong linear relationship between the independent variables, which makes it difficult to accurately assess the individual impact of each factor.

When the coefficients are analyzed economically, investment, consumption, and employment have a positive impact on GDP, while exports and government spending show signs different from those expected in the model. Inflation, on the other hand, has a very small positive impact.

In general, the results obtained show that the regression model has a very high level of accuracy in general. The model explains a large part of GDP, especially the combined effect of economic factors is significant. However, it is advisable to optimize the number of factors and conduct additional diagnostic analyses (e.g., VIF test) to further improve the model.

Conclusion: This study examined the relationship between GDP and macroeconomic factors affecting it using regression analysis. The results once again confirmed that regression analysis is an important scientific tool for quantitative assessment of economic processes.

The multivariate regression model, based on practical calculations, was found to have a very high accuracy. The high value of the coefficient of determination indicated that the model almost completely explained the outcome variable. At the same time, the overall statistical significance of the model was confirmed using the F-statistic.

According to the results of the analysis, it was found that not all factors are statistically significant at the individual level. This is explained by their interdependence or the presence of multicollinearity. Therefore, there is a need to improve the model.

It was also found that methodological problems that may arise when using a regression model during the research process – multicollinearity and small sample size – may affect the reliability of the results.

Overall, this study confirmed that regression analysis is an effective tool for in-depth study of economic processes and drawing scientifically sound conclusions.

Literature list

1. Gujarati, DN **Basic Econometrics** . – New York: McGraw-Hill, 2004.
2. Wooldridge, JM **Introductory Econometrics: A Modern Approach** . – Boston: Cengage Learning, 2016.
3. Greene, WH **Econometric Analysis** . – New York: Pearson Education, 2018.
4. Stock, JH, Watson, MW **Introduction to Econometrics** . – Boston: Pearson, 2020.
5. Shodmonov, Sh., Gafurov , U. **Economics Theory** . – Tashkent: Economics-Finance , 2010.
6. Uzbekistan Republic Statistics agency . Official statistic data . – https://stat.uz
7. Uzbekistan Republic Central bank . Monetary politics indicators . – https://cbu.uz
8. World Bank . Uzbekistan according to data . –. https://data.worldbank.org
9. Rakhimova Umida Ziyadullayevna Foreign Strategic Approaches to Optimizing the Innovative System of Regionalism. American Journal of Economics and Business Management. Vol. 8 Issue 4 | pp. 1652-1658 ISSN: 2576- 5973 (08.00.00, Decision No. 360 of 28 August 2024 of the Executive Board of OAK)
10. Rakhimova Umida Ziyadullayevna. " Tourism infrastructure innovative in development main sources analysis ". Economic progress and analysis . Scientific electronic magazine . Issue 7, July 2024.

11. Rakhimova Umida Ziyadullayevna. Tourism infrastructure current status . SERVIS magazine , No. 4. 2024 (08.00.00; No. 20).

12. Rakhimova Umida Ziyadullayevna In the region national tourism innovation development trends and stability analysis . Journal of science and education in agriculture. March 2025. Volume 3, Issue 3

13. Rakhimova Umida Ziyadullayevna. Region tourism infrastructure innovation development strategies . " Digital economics " scientific-electronic" Journal | Issue 11. 1867-1877. 2025.

14. UZ Rakhimova . Tourism field in development tourist firm and organizations importance . Academic Research in Educational Sciences. ISSN: 2181-1385 Volume 5 | Issue 5 | 2024.

15. UZ Rakhimova . The role of mathematics in the development of the tourism industry. // " Economics and society " international magazine . – 2024 - #4(119). (11.00.00; #11).