

MULTIFACTORIAL MODELING OF REGIONAL ECONOMIC DEVELOPMENT IN UKRAINE: ADDRESSING THE MULTICOLLINEARITY ISSUE

SUKHOVYCH K.V., Master's Degree Student
Ivan Franko National University of Lviv, Lviv, Ukraine

The analysis of the dynamics of the Gross Regional Product (GRP) in Ukraine has become particularly relevant in the context of ongoing reforms and decentralization of power. To identify key drivers of economic growth at the regional level, it is advisable to apply econometric and mathematical methods and models. The results of such studies can be used to develop and assess the effectiveness of state policies aimed at stimulating economic activity in regions and reducing regional disparities.

To analyze the regional socio-economic development of Ukraine, scholars use a mathematical approach based on models such as cluster analysis [1], simultaneous models [2], taxonomic analysis [3], ARIMA-model [4; 5], VAR-model [6] and others.

Gross Regional Product (GRP) is traditionally considered a key indicator of regional economic development. To conduct a comprehensive analysis of GRP dynamics across Ukraine's regions, a multifactorial linear regression model was developed. The study allowed for the identification and quantitative assessment of the impact of several determinants, such as:

- x_1 – capital investments;
- x_2 – employed population;
- x_3 – regional exports;
- x_4 – regional imports;
- x_5 – number of active enterprises in the region.

The data for this study on the economic state of Ukraine's regions were collected from the State Statistics Service of Ukraine [7].

When building such a model, it is crucial that the parameter estimates adequately reflect the impact of each factor on the resulting variable, as incorrect parameter estimates can lead to erroneous economic interpretations of the results. The phenomenon where multiple factors are linearly dependent on each other, i.e., there is an interdependence between factors, is called multicollinearity, which can lead to biased parameter estimates.

Therefore, before determining the parameters of the initial model $y = b_0 + b_1x_1 + b_2x_2 + b_3x_3 + b_4x_4 + b_5x_5$, it is necessary to test it for multicollinearity.

The most comprehensive method for testing the model for multicollinearity is the Ferrara-Glober algorithm. This method consists of three main stages:

- Testing the system of independent variables for multicollinearity;
- Testing each factor for multicollinearity;
- Testing pairs of factors for multicollinearity.

For the first stage, it is necessary to normalize the factors, find the correlation matrix, its determinant, and based on this, calculate the empirical value of the χ^2 -criterion using the formula:

$$X_{em}^2 = -\left(n - 1 \cdot \frac{1}{6}(2k + 5)\right) \cdot \ln|R| \quad (5.1)$$

For the analyzed model, x^2 (емп) = -41,6, while x^2 (крит) = 23,2. The absolute critical value exceeds the empirical value, indicating the presence of dependency in the system of factor variables.

Next, each factor was tested for correlation with all other factors using the F-statistic. The formula used for this was:

$$F_j^{em} = \frac{(C_{jj}-1)(n-k)}{k-1} \quad (5.2)$$

For the studied model, the values of the F-statistic were: $F_1 = 56,5$; $F_2 = 75,3$; $F_3 = 22$; $F_4 = 29,1$; $F_5 = 112,1$; $F_{крит} = 14$, with a critical value of $F_{крит} = 14$. Since all empirical values are greater than the critical values, it is concluded that each factor is multicollinear with some others.

The final stage of the Ferrara-Glober algorithm involves testing pairs of factors for multicollinearity using the t-statistic. The critical t-statistic value is 2.89. Empirical values exceed the critical value in two cases, indicating a strong interrelationship between capital investments and exports, as well as between the employed population and the number of active enterprises. Thus, the Ferrara-Glober algorithm confirms the presence of multicollinearity in the model.

The most pronounced relationship is observed between variables x_2 (employment) and x_5 (number of active enterprises), with a correlation coefficient of 0.96. Furthermore, variable x_5 is strongly correlated (greater than 0.8) with two other indicators, while x_2 is only correlated with one. Additionally, when applying the Ferrara-Glober algorithm, the highest values for F-statistics and t-statistics were associated with variable x_5 . Moreover, during the analysis, the Variance Inflation Factor (VIF) was calculated to measure multicollinearity, with the highest value (VIF = 24.6) also corresponding to variable x_5 . Therefore, during the first iteration, the factor variable x_5 (number of active enterprises) was chosen for removal.

This choice is also justified from an economic perspective: employment accounts for not only the production sector but also services and non-commercial sectors, which make significant contributions to the GRP. The number of active enterprises may not fully reflect these sectors, leading to an underestimation of their contribution to the GRP.

After conducting the first iteration, multicollinearity remained present (a strong relationship persisted between capital investments and all other factors). Although capital investments have a relatively strong correlation with GRP, the decision was made to exclude this factor, as it depends on all the other studied factors, causing inadequate parameter estimation in the model. After removing capital investments, multicollinearity was finally eliminated (none of the pairwise correlation coefficients exceeded 0.8, the coefficient of determination was 0.95, and all model parameters were statistically significant).

The regional economic development model, after removing multicollinearity, is given by:

$$y = -24766.7 + 232.8x_2 + 17.4x_3 + 13.2x_4, \quad (5.3)$$

where: y is the GRP,

X_2 is the employed population,

X_3 is regional exports,

X_4 is regional imports.

The Gross Regional Product (GRP) is a fundamental indicator of a region's economic well-being, particularly in the context of global economic challenges and regional disparities. By applying multifactorial linear regression to data from the State Statistics Service of Ukraine for 2021, this study aims to deepen the understanding of how key factors such as capital investments, employment, external trade, and entrepreneurship impact the economic development of Ukraine's regions. The findings of this research are crucial for the development of effective regional policies aimed at stimulating economic growth and fostering balanced development.

References:

1. Зомчак Л. М., Дида А. О. Регіональна диференціація та диспропорції аграрного сектору України: кластерний підхід. Цифрова економіка та економічна безпека. 2024. №6 (15). С. 24-30.
2. Zomchak L., Miskiv D. Structural model of Ukrainian economic performance: interactions between GDP and industrial output. *Smart Economy, Entrepreneurship and Security*. 2024. №2(2). P. 7-16
3. Zomchak L., Vdovyn M., Deresh O. Regional Economic Development Indicators Analysis and Forecasting: Panel Data Evidence from Ukraine. In *The International Conference on Artificial Intelligence and Logistics Engineering* (pp.). Cham: Springer Nature Switzerland. 2023. P. 217-228.
4. Zomchak L., Kukhotska T. Wheat market price dynamics in Ukraine: quantitative exploration and forecasting. *European Journal of Economics and Management*. 2023. №4(9). С. 14-22.
5. Zomchak L., Umrysh H. Modeling and forecasting of meat and eggs producing in Ukraine with seasonal ARIMA-model. *Agricultural and Resource Economics: International Scientific E-Journal*. 2017 №3(3). P.16-27.
6. Зомчак Л. М., Рудницька Я. І. Економетричне моделювання залежностей між макроекономічним станом України та обсягами залучених депозитів. Причорноморські економічні студії. 2020. С. 196-200.
7. State Statistics Service of Ukraine. URL: <https://www.ukrstat.gov.ua> (access date: 12.08.2024).