UDC: 330.43, 330.46, 519.237
DOI: 10.37661/1816-0301-2025-22-4-24-35

Original Article
Оригинальная статья

Latent factor analysis of regional growth in Belarus using structural equation modeling

Yuliya A. Osipova

The Economy Research Institute of the Ministry
of Economy of the Republic of Belarus,
st. Slavinskogo, 1/1, Minsk, 200086, Belarus
E-mail: osipova@niei.by

Abstract

Objectives. The study provides a quantitative assessment of the impact of latent variables on regional economic activity and examines the interactions among various economic sectors and their contributions to regional growth. The analysis focuses on identifying latent factors of socio-economic development across Belarusian regions during the period 2016-2024, using factor analysis techniques and structural equation modeling (SEM).

Methods. To reduce data dimensionality and identify latent factors, exploratory factor analysis (EFA) was applied using the Principal Component Analysis (PCA) extraction method. A structural equation model was constructed using the SEMOPY library in Python to estimate relationships among the identified factors. To assess the quality of the model, standard fit indices were used: CFI – comparative fit index; TLI – Tucker-Lewis index; RMSEA – root mean square error of approximation. The values of these indices allow evaluating the degree of consistency between the model and the empirical data. The model is based on a system of 23 socio-economic indicators across 128 administrative districts and cities of regional subordination.

Results. The resulting SEM demonstrates high internal consistency and statistical reliability (CFI = 0,98; TLI = 0,97; RMSEA = 0,04), revealing significant causal linkages between latent factors. It was established that the financial sector is a key driver of investment activity, while growth in the housing stock directly stimulates consumer demand. Negative relationships were identified between agricultural potential and financial stability, as well as between industrial development and financial sustainability.

Conclusion. The developed model is an effective analytical tool for formulating evidence-based regional policy, optimizing resource allocation, and strategic planning. Promising directions for future research include incorporating time lags, adding indicators of innovation and human potential, and applying spatial econometrics methods.

Keywords: structural equation modeling, factor analysis, PCA, latent variables, Varimax, DWLS-estimation

For citation. Osipova Yu. A. *Latent factor analysis of regional growth in Belarus using structural equation modeling.* Informatika [*Informatics*], 2025, vol. 22, no. 4, pp. 24‒35 (In Russ.). DOI: 10.37661/1816-0301-2025-22-4-24-35.

Conflict of interest. The author declares of no conflict of interest.

Received | Поступила в редакцию 09.09.2025
Accepted | Подписана в печать 23.10.2025
Published | Опубликована 30.12.2025

© Osipova Yu. A., 2025

Анализ латентных факторов регионального роста в Беларуси с использованием структурного моделирования

Ю. А. Осипова

Научно-исследовательский экономический институт
Министерства экономики Республики Беларусь,
ул. Славинского, 1/1, Минск, 200086, Беларусь
E-mail: osipova@niei.by

Аннотация
Цели. Целью исследования являются количественная оценка влияния скрытых переменных на экономическую активность регионов Беларуси и анализ взаимодействия между различными секторами экономики путем выявления латентных факторов социально-экономического развития регионов за 2016–2024 гг. с использованием методов факторного анализа и моделирования структурными уравнениями (structural equation modeling, SEM).
Методы. Для сокращения размерности данных и выявления латентных факторов был применен метод объясняющего факторного анализа (EFA) с использованием процедуры извлечения факторов методом главных компонент (PCA). Для оценки связей между выявленными факторами была построена модель структурных уравнений с использованием библиотеки SEMOPY в Python. Для проверки качества модели использовались стандартные показатели соответствия: CFI – сравнительный индекс соответствия; TLI – индекс Такера-Льюиса; RMSEA – среднеквадратичная ошибка аппроксимации. Значения этих индексов позволяют оценить степень согласованности модели с эмпирическими данными. Модель основана на системе из 23 социально-экономических показателей по 128 административным районам и городам областного подчинения за девять лет.
Результаты. Построенная модель структурных уравнений демонстрирует высокую внутреннюю согласованность и статистическую надежность (CFI = 0,98; TLI = 0,97; RMSEA = 0,045) и выявляет значимые причинно-следственные связи между факторами. Финансовый сектор выступает ключевым драйвером инвестиционной активности, а рост жилищного фонда прямо усиливает потребительский спрос. Одновременно обнаружены отрицательные взаимосвязи между аграрным потенциалом и финансовой стабильностью, а также между промышленным развитием и финансовой устойчивостью.
Заключение. Представленная модель – это эффективный аналитический инструмент для разработки научно обоснованной региональной политики, оптимизации распределения ресурсов и стратегического планирования. Перспективные направления дальнейших исследований включают учет временных лагов, добавление показателей инновационного и человеческого потенциала, применение методов пространственной эконометрики.

Ключевые слова: моделирование структурными уравнениями, факторный анализ, PCA, латентные переменные, Varimax-ротация, DWLS-оценка

Для цитирования: Осипова, Ю. А. Анализ латентных факторов регионального роста в Беларуси с использованием структурного моделирования / Ю. А. Осипова // Информатика. – 2025. − Т. 22, № 4. – С. 24–35. – DOI: 10.37661/1816-0301-2025-22-4-24-35.

Конфликт интересов: Автор заявляет об отсутствии конфликта интересов.

Introduction. The socio-economic development of Belarusian regions is shaped by a complex interaction of factors, including agricultural potential, investment activity, and infrastructure provision. This paper serves as a continuation of prior research, in which the author applied machine learning techniques to regional datasets – specifically, factor and cluster analysis – to identify key indicators reflecting common development patterns across Belarus.

In the present study, structural equation modeling (SEM) is employed to quantitatively assess the influence of latent factors on regional socio-economic activity.

The purpose of this work is to uncover latent interdependencies among socio-economic indicators for the period 2016-2024 through factor analysis, and to identify their defining characteristics by constructing a structural equation model.

To examine the spatial dynamics of regional development, the author addresses the following research tasks:

1. Identify the patterns of socio-economic development based on indicators that reflect spatial-temporal dimensions.
2. Assess the quantitative impact of generalized latent patterns on regional development.
3. Construct a structural equation model that incorporates linear relationships between observed and latent variables.

The database was assembled using statistical indicators published by Belstat for 2016-2024. 23 indicators characterizing the socio-economic development of Belarusian administrative-territorial entities (ATEs) were selected for analysis, and their values were calculated for each district. Because some of these indicators reflect agricultural activity, the sample comprises 118 districts and 10 cities of regional subordination, with the agricultural-related indicators set to zero for those 10 cities. The full data matrix is not included in the article owing to its high dimensionality.

The research algorithm includes four stages.

I. Dataset Overview and Indicator Selection. A portion of the indicators was calculated per 1000 residents or per capita. Each region over time is represented as a vector of observations: $X_s(t) = \left(x_{s1}(t), x_{s2}(t), ..., x_{sp}(t)\right)$, where $s = 1, 2, \ldots, m$, and in our case $m = 128$. Given the nature of the data, the temporal parameter is modeled as a discrete sequence of evenly spaced intervals, defined as $t = 1, 2, \ldots, T$. In statistical yearbooks, the duration of one interval typically corresponds to one year. Accordingly, our analysis spans the period 2016-2024, with $T = 9$[1].

The resulting dataset includes 23 indicators reflecting socio-economic characteristics for each of ATEs, yielding $n = 128 \times 9 = 1152$ complete records across nine years.

II. Data Standardization. Standardization was performed using the `StandardScaler` method from the `scikit-learn` library[2]. This technique adjusts feature values so that each has a mean of zero and a standard deviation of one. The purpose of this preprocessing is to enhance algorithmic convergence and prevent features with large magnitudes from dominating the model. In this study, scaling parameters were computed and applied directly to the raw values of each indicator.

III. Factor Analysis: Principal Components. To extract latent patterns, we employed exploratory factor analysis (EFA). The mathematical model is written as

$$x_i = \lambda_{i1}F_1 + \lambda_{i2}F_2 + \ldots + \lambda_{im}F_m + \varepsilon_i,$$

where x_i is an observed variable, F_j is a latent factor, λ_{ij} is a factor loading, and ε_i is a unique error. The objective of exploratory factor analysis (EFA) is to estimate the loading matrix Λ and to determine the number of factors m that explain a significant portion of the total variance.

Factor extraction method: factors were extracted using principal component analysis (PCA) based on common variances, employed to reduce data dimensionality while preserving the maximum possible variance. The model for principal components is represented as a linear combination of the original variables:

$$z_j = a_{j1}x_1 + a_{j2}x_2 + \ldots + a_{jp}x_p,$$

where z_j is the j-th principal component, a_{jk} are the coefficients (elements of the eigenvectors), and x_k are the original variables. The eigenvalues λ_j of the covariance matrix determine the share of variance explained by each component. According to Kaiser's criterion, components with $\lambda_j \geq 1$ are retained.

[1]National Statistical Committee of the Republic of Belarus. Available at: https://www.belstat.gov.by/ (accessed 03.09.2025).

[2]scikit-learn developers. Preprocessing data – scikit-learn 1.7.0 documentation. Available at: https://scikit-learn.org/stable/modules/preprocessing.html (accessed 03.09.2025).

The algorithm comprises several steps:
– calculation of pairwise correlation coefficients among the original indicators;
– construction and rotation of the factor loading matrix;
– extraction of principal components;
– selection of factors based on eigenvalue criteria λ_j, $j = 1, 2, \ldots, S$; $S < q$, for $\lambda_j \geq 1$, where $S < n$ and $q = 23$.

To ensure interpretability, components are retained such that their cumulative explained variance exceeds 70 %. Full details of the factorization procedure are described in [1].

IV. Transition to Structural Equation Modeling. Following factor extraction, SEM is applied to model the relationships among latent and observed variables[3].

To identify latent factors in the development of Belarusian ATEs, the following 23 socio-economic indicators for 2016-2024 were selected and are shown in table 1.

Table 1
Statistical indicator designation

Variable	The variable names in the program	Statistical indicator
x_1	agriculturalland	Agricultural land per 1000 inhabitants, ha
x_2	cattlestockpercapita	Number of cattle per capita, thousands of head
x_3	grainpercapita	Gross grain and legume production per capita, thousand tons
x_4	housingprovisionpercap	Housing provision per capita, m^2
x_5	milk	Milk production per capita, tons
x_6	overduereceivables	Overdue receivable, % of total receivables
x_7	overduecredible	Overdue credible, % of total payables
x_8	overdueloans	Overdue loans, % of total payables
x_9	profitability	Profitability of sales, %
x_{10}	contractedworks	Growth rate of contracted works, %
x_{11}	investment	Growth rate of capital investment, %
x_{12}	constructionwork	Growth rate of construction and installation work, %
x_{13}	residential area	Growth rate of commissioning of residential area, %
x_{14}	housingstock	Growth rate of housing stock, %
x_{15}	smallentreprises	Number of small and medium-sized enterprises per 1000 inhabitants
x_{16}	industrialpercapita	Industrial output per capita, thousand BYN
x_{17}	revenuepercapita	Revenue per capita, thousand BYN
x_{18}	ratehireddismissed	Ratio of hired to dismissed employees
x_{19}	employed	Growth rate of employment, %,
x_{20}	density	Population density, residents per km^2
x_{21}	indexagri	Growth rate of agricultural output, %
x_{22}	publictrade	Growth rate of public trade turnover, %
x_{23}	retail	Retail turnover per capita, rubles

[3]Contreras, C. Analysis and visualisation of SEM using Python SEMOPY (Structural Equation Modeling). *Medium*. Available at: https://cr-c.medium.com/analysis-and-visualisation-of-sem-using-python-semopy-structural-equation-modeling-bd6c3f463e33 (accessed 03.09.2025).

Factorization procedure and dimensional reduction. The factorization procedure was performed using the Python programming language in the Colab environment. Factor component analysis (FCA) was applied to determine latent factors. To validate the computations, additional calculations were conducted using SPSS Statistics.

The dataset includes 128 administrative districts across the years 2016-2024. For further processing, the data were compiled into a matrix comprising 128*9=1152 complete observations over a 9-year period covering all 128 administrative districts of Belarus across 23 variables. Eigenvalues of the correlation coefficient matrix – representing the proportion of total variance in the original indicator system explained by each principal component following axis rotation – are provided in table 2.

As a result of the factor analysis, the initial 23 indicators were transformed into 7 principal components, which together preserve 70,4 % of the system's total variance. Based on data from 2016-2024, the first two components explain approximately 17,3 % of the original variance. In this case, the first four components retain 47,8 % of the variance, while the cumulative variance retained by the first six components reaches approximately 63,2 %. Therefore, the structural relationships among indicators were analyzed using 7 principal components (fig. 1).

Determination of the number of components based on eigenvalue distribution. Model adequacy was assessed using the Kaiser–Meyer–Olkin (KMO) measure of sampling suitability. The KMO value, which reflects the degree to which the data are appropriate for factor analysis, was approximately 0,74, indicating a satisfactory level of adequacy (table 3). An additional measure of inter-variable association is Bartlett's test of sphericity. This test evaluates the null hypothesis that the variables in the correlation matrix are mutually uncorrelated. The observed significance level was 0.00, which is sufficiently low to reject the null hypothesis. Consequently, the relationships among variables are statistically significant, and the extraction of latent factors is justified.

Table 2
Total explained variance

Factor	Extraction (sum of squared loadings)			Rotation (sum of squared loadings)		
	Eigenvalue	Variance, %	Cumulative variance, %	Eigenvalue	Variance, %	Cumulative variance, %
F1	5,74	24,94	24,94	3,98	17,31	17,31
F2	3,02	13,14	38,08	2,73	11,88	29,19
F3	2,00	8,68	46,76	2,43	10,58	39,77
F4	1,79	7,79	54,54	1,84	7,99	47,76
F5	1,30	5,66	60,21	1,78	7,74	55,50
F6	1,22	5,29	65,50	1,77	7,69	63,19
F7	1,12	4,86	70,36	1,65	7,17	70,36

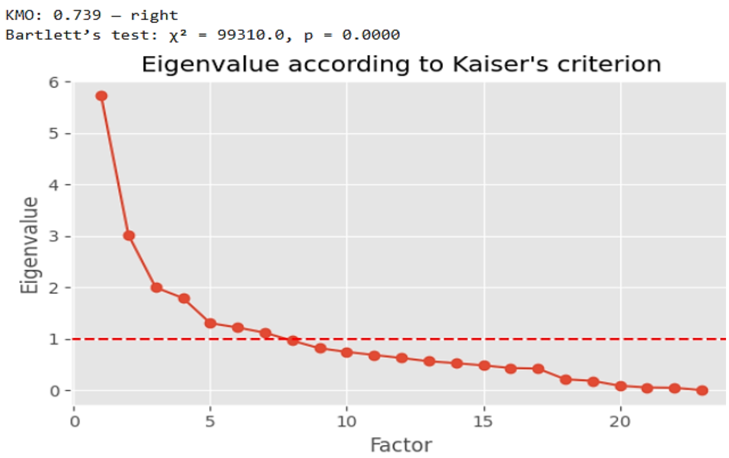

Fig. 1. KMO criteria

Table 3
Kaiser–Meyer–Olkin measure of sampling adequacy and Bartlett's test of sphericity

Statistic	Value
Kaiser–Meyer–Olkin measure	0,739
Bartlett's test of sphericity: approximate $\chi 2$; degrees of freedom; significance ($p < 0{,}001$)	 99310 136 0,000

At the next stage, each of the seven extracted components is interpreted based on the variables that load onto it. For further analysis, exactly seven principal components were retained. The factor analysis was performed on a standardized 1152×23 data matrix, and after Varimax rotation with Kaiser normalization we obtained a 23×7 loading matrix. The rotated loading matrix is visualized in fig. 2 as a heatmap generated in the Colab environment.

In the figure, «warm» colors correspond to loadings exceeding 0,6–0,7, while «cool» colors represent coefficients below this threshold, yet still meaningful in magnitude. Each of the six components incorporates a distinct subset of the 20 selected indicators.

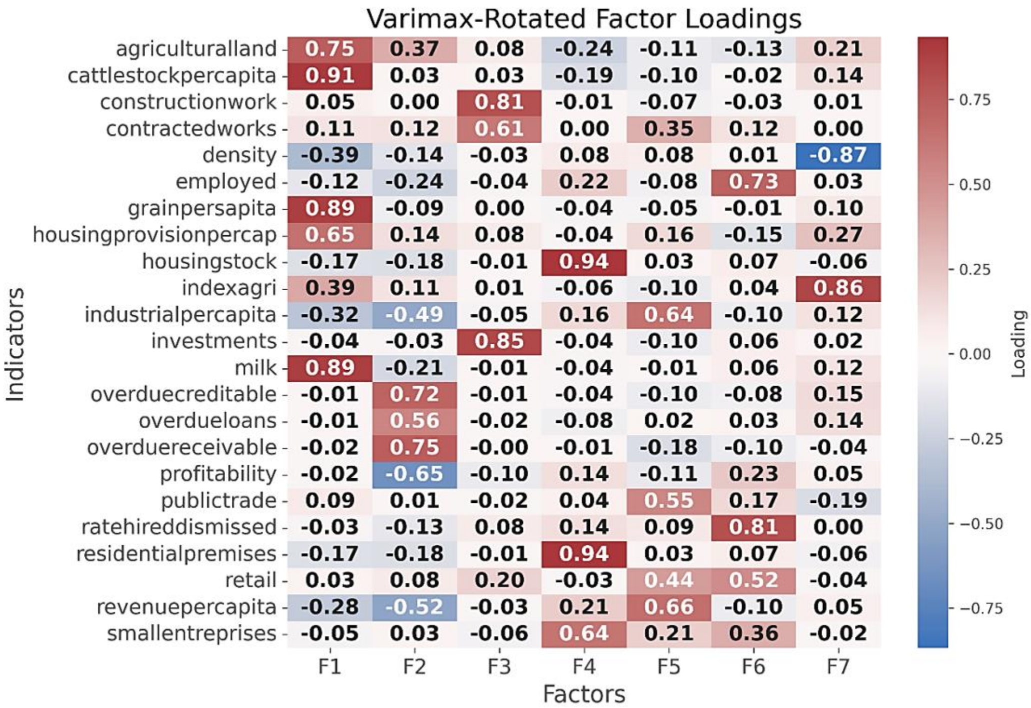

Fig. 2. Heatmap visualization of the rotated factor loading matrix

The rotated factor loading matrix reveals the following generalized structure: components F1-F7 are formed under the influence of specific observed indicators. Table 4 presents each principal component, its economic interpretation, and the variables associated with each of the 7 factors. Based on table 4, and to make the factor labels in the measurement and structural models as clear as possible, we introduce the following labels for factors in the Python program. These labels map factors F1-F7 to names that will be used in the measurement and structural models, tables, figures, and model outputs. In the code and the measurement and structural models F1 is displayed as «Agriculture», F2 – as «Finance», F3 – as «Investments», F4 – as «HousingFund», F5 – as «Industry», F6 – as «Employment», F7 – as «DemographicAgriculturalBalance».

Table 4
Economic interpretation of the factors

Factor	Economic interpretation	Variables forming the factor
F1	Level of agricultural development of the region	x_1 – agricultural land area per l000 inhabitants, ha; x_2 – number of cattle per capita, thousands of heads; x_3 – grain per capita, tons; x_4 – housing provision per capita, m^2; x_5 – milk production per capita, tons
F2	Regional financial stability	x_6 – overdue receivable, % of total receivables; x_7 – overdue credible, % of total payables; x_9 – profitability of sales, %
F3	Investment potential level	x_{10} – growth rate of contracted works, %; x_{11} – growth rate of capital investment, %; x_{12} – growth rate of construction and installation work, %
F4	Level of business and construction activity	x_{13} – growth rate of commissioning of residential area, %; x_{14} – growth rate of housing stock, %; x_{15} – number of small and medium-sized enterprises per 1000 inhabitants
F5	Industrial potential	x_{16} – industrial output per employee, thousand BYN; x_{17} – revenue per capita, thousand BYN
F6	Employment level	x_{18} – ratio of hired to dismissed employees; x_{19} – growth rate of employment, %
F7	Demographic-agrarian balance	x_{20} – population density, residents per km^2; x_{21} – growth rate of agricultural output, % Note: negative loading on agricultural output index reflects urbanization and workforce concentration in developed urban territories

The set of variables influencing the first principal component reflects the level of agricultural development and specialization of the territories. Component F1 is economically interpreted as the «*Level of agricultural development of the region*». High loadings on agricultural land area, gross grain harvest, cattle population, and milk production per capita indicate a pronounced agrarian profile. This factor is important for assessing the resilience of the agricultural sector and the region's food-security potential, explaining approximately 17,3 % of the variance attributed to the first component.

Component F2, labeled «*Regional financial stability*», explains 11,9 % of the variance and captures key indicators of organizational risk and solvency. It loads positively on overdue receivables and payables and negatively on sales profitability, meaning that higher F2 scores correspond to regions where mounting debt burdens coincide with declining profit margins – signaling weak liquidity and elevated financial risk.

Component F3 «*Investment potential level*» reflects the region's capacity to attract and deploy investments with high returns. Two key variables exhibit strong loadings, and this component explains about 10,6 % of the total variance.

Component F4 clusters indicators describing housing infrastructure and entrepreneurship: total housing stock, number of dwelling units, and count of small enterprises. It accounts for roughly 8,0 % of the variance and is interpreted as the «*Level of business and construction activity*».

Component F5 «*Industrial potential*» characterizes regions with developed industrial sectors, high revenues. This factor, indicating macroeconomic resilience and an industrial orientation, explains 7,7 % of the variance.

Component F6 «*Employment level*» (7,7 % of the variance) captures labor market conditions, consumer activity, and social stability.

Component F7 «*Demographic-agrarian balance*» links population density with physical agricultural output. The negative loading on agricultural growth rates suggests urbanization trends: as economies mature, workforce and capital shift from agriculture to industrial and service sectors. This factor explains 7,2 % of the variance.

After preliminary factor analysis using Varimax rotation, the structural equations were formulated and depicted in a comprehensive path diagram (fig. 3). SEM belongs to second-generation multivariate data-analysis methods, overcoming three key limitations of first-generation approaches (e.g., multiple and logistic regression)[4].

First-generation methods assume simple, single-level relationships among observed variables. SEM allows the analysis of multilevel dependencies between latent (unobserved) and observed variables.

Classical methods require error-free data, which is impractical. SEM integrates measurement errors into the model, improving estimation accuracy. SEM also supports testing theoretical constructs (e.g., economic activity) measured indirectly through sets of observed indicators.

The method is also known as causal modeling, covariance structure analysis, or confirmatory factor analysis.

In this paper, SEM is employed to test hypotheses regarding the influence of latent factors on economic activity and to evaluate the complex interrelations among investments, infrastructure, and productivity, thereby explaining the observed raw correlations among directly measured variables.

A comprehensive SEM toolkit was developed in Python using the SEMOPY library[5], offering a robust alternative for both practical and academic applications [2].

The parameters of the structural equation model were estimated using the Diagonally Weighted Least Squares (DWLS) method, implemented in the SEMOPY library in Python. In the code, this was specified by setting the argument obj="DWLS" in the model.fit() function.

SEM Workflow includes the following stages.

1. *Model conceptualization:*
– formulate hypotheses regarding inter-variable relationships;
– specify theoretical models with latent and observed variables.

2. *Data collection and preparation:*
– select appropriate data sets;
– check data quality, address missing values, assess normality.

3. *Measurement model development:*
– identify latent variables and their indicators.

4. *Structural model construction:*
– define dependent and independent variables;
– specify directional paths and causal links among factors.

5. *Model estimation and evaluation:*
– estimate path coefficients and factor loadings;
– assess model fit using indices such as RMSEA, CFI, and TLI.

6. *Analysis and interpretation of results:*
– assessment of the statistical significance of the coefficients;
– analysis of the structure of relationships among variables.

7. *Model modification and final adjustment:*
– adjustment of the model if fit indices are unsatisfactory;
– examination of alternative models and refinement of parameters.

In our case, following exploratory factor analysis, the latent variables listed in table 4 were identified, upon which the measurement model was constructed:

Measurement model:

Agriculture =~ agriculturalland + cattlestockpercapita + grainpersapita + housingprovisionpercap +milk
Finance =~ overduecreditable + overduereceivable + profitability
Investments =~ constructionwork + investments+contractedworks
HousingFund =~ housingstock + residentialpremises + smallentreprises
Industry =~ industrialpercapita + revenuepercapita
Employment =~ ratehireddismissed + employed
DemographicAgriculturalBalance =~ density + indexagri

[4]DataCamp.Structural Equation Modeling: What It Is and When to Use It. Available at: https://www.datacamp.com/tutorial/structural-equation-modeling (accessed 03.09.2025).

[5]Semopy. Structural Equation Modeling in Python. Available at: https://semopy.com/ (accessed 03.09.2025).

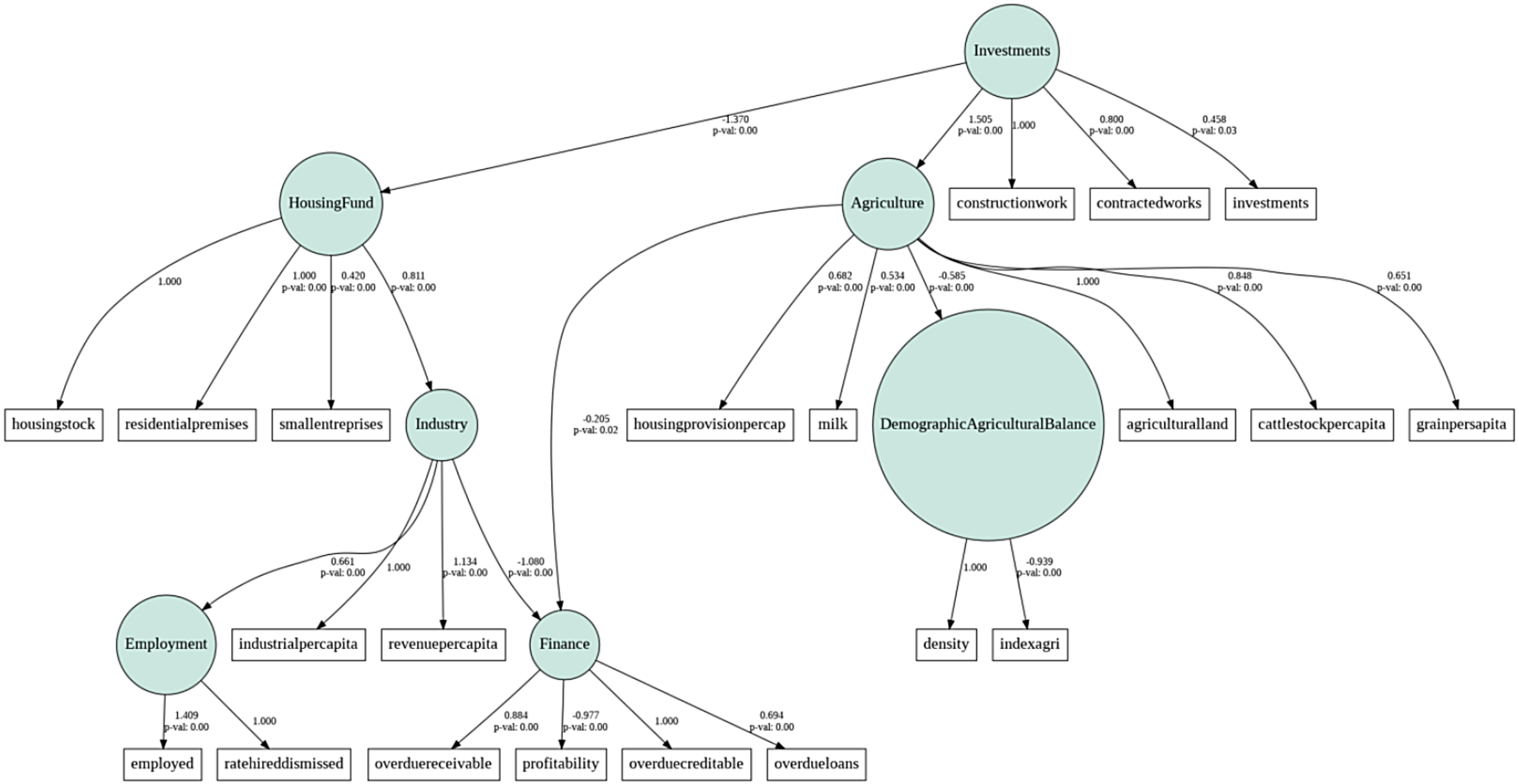

Fig. 3. Structural equation model of the socio-economic development of Belarusian regions for 2016-2024

Each of these measurement linkages is confirmed by high z-statistic values and extremely low p-values (p < 0,001 or thereabouts), indicating the statistical significance of the selected indicators in describing the latent variables (fig. 3). The path diagram in fig. 3 illustrates how latent (unobserved) variables and their observed indicators interrelate and influence one another. Latent variables are depicted as circles and observed indicators as rectangles. Arrows indicate the direction of influence, alongside standardized coefficients and p-values that attest to the statistical significance of each path. Path Diagram of relationships among 7 factors illustrated on fig. 4. On the fig. 4 * denotes (p < 0,05) – the probability of obtaining this or a more extreme result under the null hypothesis is less than 5 %, ** denotes (p < 0,01) – the probability of obtaining this effect by chance is less than 1 %.

These results underscore the reliability of the measurement model and the objectivity of the obtained estimates, allowing them to be used for further interpretation of structural relationships in the regional development model.

Structural model:

DemographicAgriculturalBalance ~ Agriculture
Agriculture ~ Investments
Finance ~ Agriculture + Industry
HousingFund ~ Investments
Industry ~ HousingFund
Employment ~ Industry

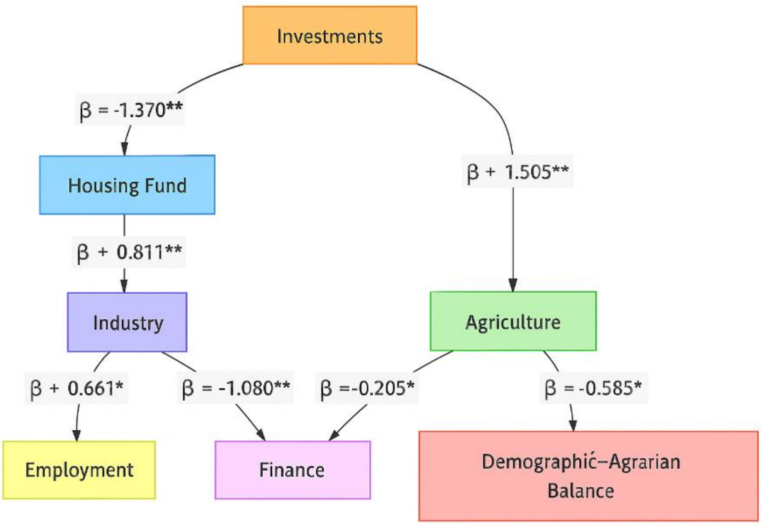

Fig. 4. Diagram of relationships among 7 factors

In fig. 3, adjacent to each path, the following metrics are shown:

Estimate – the standardized regression coefficient indicating the effect size;

Std. Err. – the standard error of the estimate; lower values indicate higher precision;

z-value – the Wald statistic testing how far the coefficient deviates from zero;

p-value – the significance level, with all paths deemed significant at p < 0,01.

Let’s examine and analyze these metrics in greater detail. Each indicator loading onto the latent factor is significant (p ≤ 0,05).

Main structural relationships and their interpretation. The structural part of the model comprises seven directed paths between latent factors:

1. *Agriculture → DemographicAgriculturalBalance* (Estimate = –0,585, p < 0,05). This indicates that in regions with higher demographic indicators and agricultural potential, the economy experiences a structural shift characterized by a relative decline in the role of agriculture. This pattern can be explained by processes of urbanization and the migration of labor resources into other sectors of the economy. Beyond this, such shifts often mirror the classic Lewis dual-sector model, where surplus rural

labor is absorbed into industrial and service sectors. Policy responses might include diversifying rural economies, investing in agro-processing industries, and bolstering social infrastructure to manage urban migration pressures.

2. *Agriculture → Finance* (Estimate = –0,205 (weak connection), $p < 0,05$). This may indicate that agrarian-oriented regions, on average, exhibit a less developed financial sector.

3. *Industry → Finance* (Estimate = –1,080, $p < 0,05$). Since the Finance factor shows a strong direct relationship with debt indicators, this suggests that industrial enterprises in the regions are highly indebted. A high level of industrial development accompanied by a high share of overdue debt (as indicated by the loadings on the Finance factor) can create macroeconomic imbalances that negatively affect the short-term stability of the financial system. This warrants further in-depth analysis.

4. *Investments → Agriculture* (Estimate = +1,505, $p < 0,05$). A coefficient exceeding 1 indicates a high sensitivity of the agricultural sector to investment volume: small changes in investment levels lead to more pronounced shifts in the agrarian sector.

5. *Investments → HousingFund* (Estimate = –1,370, $p < 0,05$). A negative relationship points to a «crowding-out effect»: as investments in the economy increase – for example, channelled into industrial infrastructure, major construction projects, or modernization – resources allocated to maintaining and expanding the housing stock (such as capital repairs, adding new residential units, or affordable-housing programs) may be reduced. In practical terms, this means that in regions experiencing intensified investment activity, housing-sector indicators tend to weaken simultaneously, highlighting the need for a more detailed analysis of how public and private investment funds are redistributed across sectors.

6. *HousingFund → Industry* (Estimate = +0,811, $p < 0,05$). Conceptually, stronger housing-sector performance (larger housing stock, more residential premises, and vibrant small-enterprise activity in construction) appears to stimulate industrial output. Improved housing infrastructure can boost labor availability, raise demand for building materials and durable goods, and spur related manufacturing, all of which drive up overall industrial activity.

7. *Industry → Employment* (Estimate = +0,061, $p < 0,05$). Although modest in size, this coefficient confirms that stronger industrial activity contributes to higher employment levels. The relatively small magnitude suggests that while industry does drive job creation, other sectors or macro-factors (e.g., services, government policy, automation) also play important roles in shaping overall employment.

All paths are statistically significant ($p < 0,05$), underscoring the robustness and interpretability of the proposed structural scheme.

Model fit assessment. The model fit indices are summarized in table 5.

Table 5
Model fit indices

Metric	DoF	DoF Baseline	chi2	chi2 p-value	chi2 Baseline	CFI	GFI	AGFI	NFI	TLI	RMSEA	AIC	BIC	LogLik
Value	182	210	229,3	0,01	2448,8	0,98	0,91	0,89	0,91	0,98	0,045	75,5	215,2	11,3

For the assessment of model-data fit, traditional fit indices were used, based on a comparison of the observed and implied covariance structures.

Interpretation of model fit indices:

DoF (Degrees of Freedom). The number of independent constraints in the model. In the table: **DoF = 182** for the tested model and **DoF Baseline = 210** for the baseline model. Higher DoF indicates a more «free» model; very low DoF may signal over-identification.

Chi-square (χ^2) and p-value. A small χ^2 and a high p-value (>0,05) indicate good fit. In this case **p = 0,01**, so the model differs statistically from a perfect fit; however, with large samples such deviations are common and should be interpreted together with other indices.

Comparative Fit Index (CFI). CFI = 0,98, exceeding the conventional threshold (≥ 0,95), which indicates excellent model fit.

Goodness-of-Fit Index (GFI). GFI = 0,91, above the usual cutoff (≥ 0,90), indicating good fit.

Adjusted Goodness-of-Fit Index (AGFI). AGFI = 0,89, slightly below the conventional threshold (~0,90). This suggests a modest degree of over-parameterization that could be simplified; AGFI adjusts GFI for model complexity.

Normed Fit Index (NFI). NFI = 0,91, above the common threshold (≥ 0,90), which is considered acceptable. NFI compares the χ^2 of the tested model to that of the baseline model.

Tucker–Lewis Index (TLI). TLI = 0,98, exceeding the recommended cutoff (≥ 0,95). TLI penalizes model complexity and rewards parsimony.

Root Mean Square Error of Approximation (RMSEA). RMSEA = 0,0452, below the conventional threshold (≤ 0,05), indicating a close approximate fit of the model to the data.

Information criteria and log-likelihood. Relatively low values of **AIC = 75,47** and **BIC = 215,22**, together with a higher log-likelihood (**LL = 11,26**), support the conclusion that this specification achieves a favorable balance between fit and parsimony compared with alternative models.

Taken together, these indices demonstrate that the model effectively captures the relationships among latent and observed variables while maintaining statistical stability and interpretability. Further refinement could involve accounting for temporal lags or conducting multi-group analyses across different types of regions. Overall, the proposed SEM can serve as an analytical tool for identifying priority areas in territorial development, evaluating regional policy effectiveness, and informing evidence-based economic planning [3].

Future research directions include:

− incorporating time-lag effects to assess the long-term impact of investments;

− adding supplementary indicators (environmental quality, innovation capacity, human capital);

− employing of updated regional statistical data to increase the accuracy of estimates and to test the robustness of the identified latent factors;

− applying spatial econometric models to explore interregional interactions.

This study introduces a structural equation model to explore the nuanced interdependencies among regional investment activity, infrastructure provision, agricultural productivity, industrial output, financial performance, population density, and employment.

The framework combines a measurement model – linking observed indicators to their underlying latent constructs – and a structural model – estimating the directional effects between those constructs – thereby allowing us to assess both construct reliability and the causal interplay among key economic drivers.

Thus, the model identifies the most critical drivers of sustainable development and competitiveness in regional economies. SEM has proven effective for analyzing intricate socio-economic systems, and the findings can underpin targeted management of regional development in Belarus.

The practical significance of this research lies in its potential to optimize regional policy by accounting for territorial specifics, guiding the targeted allocation of investments for balanced growth, and monitoring the effectiveness of government development programs.

References

1. Chitai G. O. Analysis of the dynamics of integral economic indicators of the regions of the Republic of Belarus. *Economics, Modeling, Forecasting*, 2018, no. 12, pp. 43–51 (In Russ.).

2. Igolkina A. G., Meshcheryakov G. Semopy: A python package for structural equation modeling. *Structural Equation Modeling: A Multidisciplinary Journal*, 2020, vol. 27, iss. 6, pp. 952–963. DOI: 10.1080/10705511.2019.1704289.

3. Kline R. B. *Principles and Practice of Structural Equation Modeling*. New York, Guilford Press, 2023, 494 p.

Information about the author

Yuliya A. Osipova, Scientist, M. Sc. (Phys.-Math.), The Economy Research Institute of the Ministry of Economy of the Republic of Belarus.
E-mail: osipova@niei.by
https://orcid.org/0009-0005-3109-6775

Информация об авторе

Осипова Юлия Александровна, научный сотрудник, магистр физико-математических наук, Научно-исследовательский экономический институт Министерства экономики Республики Беларусь.
E-mail: osipova@niei.by
https://orcid.org/0009-0005-3109-6775