

Stable development of travel business in the Odessa region: estimate using multivariate statistical analysis

Valerii Matskul^{1*}, Anatoliy Kovalyov¹, Valentyn Chernyshev¹, Diana Okara²

¹ Odessa National Economic University, 65082 Odessa, Ukraine

² Odessa State Academy of civil engineering and architecture, 65029 Odessa, Ukraine

Abstract. The article examines the level of sustainable development of the travel business in the Odessa region of Ukraine. This level is evaluated as a latent indicator, which is characterized by a certain set of signs-symptoms. The methodology used for modeling is Partial Least Squares-Path Modelling (or PLS-PM) and the method Time-Wise Multi-Way Principal Component Analysis. This approach allowed us to identify and evaluate latent and explicit factors affecting the sustainable development of travel business in the Odessa region.

Key words: travel business, sustainable development, Time-Wise Multi-Way Principal Component Analysis, Partial Least Squares-Path Modelling

1 Introduction

In recent years, there has been a reorientation of the main goals of the business. Previously, the main task of business was to obtain the greatest possible profit, now stable development is taking priority. A lot of research has been devoted to the issues of sustainable development of the travel business, among which we mention works [1-3]. But many economic indicators (such as development stability, competitiveness, investment attractiveness, efficiency, etc.) are latent, clearly unmeasured. Therefore, it becomes necessary to use special methods for their evaluation. Among recent studies, we note the work [4], in which the level of tourist service is estimated using some generalizing indicator (in fact, the well-known Porter competitiveness polygon). This work is devoted to assessing the level of sustainable development of the tourism business in the Odessa region using special methods of multivariate statistical analysis. It should be noted that the Odessa region has strong potential for sustainable development. Among the many factors influencing this, we note the following: the presence of an extended line of the Black Sea coast, the presence of a developed infrastructure (sea and river ports, an international airport), sanatoriums, rest houses, cultural attractions, etc. Unfortunately, in 2020, the tourism business in the Odessa region (due to COVID-19) practically ceased to function normally.

2 Materials and methods

Data for analysis is available on the website of the State Statistics Service of Ukraine (regional statistics) and Main Department of Statistics in Odessa region [9,10]. The time period for which the data will be available starting from 2010 and ending in 2019. Preliminary data processing was carried out in MS Excel spreadsheets. When modelling and computing was used the following tools: a package for component analysis (PCA-PM) of Statistica12 [8], and program XLSTAT (a package XLSTAT-PLSPM for analysis using the least partial square method PLS-PM: Partial Least Squares Path Modelling or Projection on the Latent Structures Path Modelling).

PLS-PM methods started to be widely used in a variety of applied researches of the 70-s years of the twentieth century because of the work of Herman Wold and his son Sven Wold, in this works the basic principles of the technique of modelling PLS-PM were established. PLS-PM is a tool for modelling of interconnections between latent variables. The method PLS-PM is used for analyses of high dimensional data in the conditions of ill-structured environment. It is widely applied, in particular, in the economy to evaluate such latent indexes as a utility, level of economic development, etc. Following, the task of modelling by the method of PLS-PM can be described as follows: let

X is a $n \times p$ data matrix: $X = \begin{pmatrix} x_{11} & x_{12} & \dots & x_{1p} \\ x_{21} & x_{22} & \dots & x_{2p} \\ \vdots & \vdots & \dots & \vdots \\ x_{n1} & x_{n2} & \dots & x_{np} \end{pmatrix}$, where n is the number of observations (objects), and p is

* Corresponding author: valerii.matskul@gmail.com

the number of variables (characteristics). The matrix X can be divided into l blocks $X_j, j = \overline{1, l}$, each of which X_j is associated with a latent variable LV_j , the estimate of which is denoted $Y_j = LV_j$. All interconnections between variables can be divided into two types: interconnections between latent variables LV_j and blocks X_j (these interconnections form the outer model) and interconnections between the latent variables with each other (these interconnections form the internal model). Internal model can be written in the form of a system of linear equations: $LV_j = \beta_0 + \sum_{i \rightarrow j} \beta_{ji} LV_i + \varepsilon_j^{in}$, where LV_i is all latent variables that affect the latent variable LV_j ; β_{ji} are structural coefficients that characterize the force and the direction of connection between latent variables LV_i and LV_j ; β_0 - free member; ε_j^{in} - a random deviation of the internal model. The internal model imposed by the following requirements:

1. The system of latent equations should be recursive so it has to correspond to loop-free graph.
2. Internal model is recursive, so the following requirement is true: $E(LV_j \setminus LV_i) = \beta_{0i} + \sum_{i \rightarrow j} \beta_{ji} LV_i$.
3. Random deviation are independent of the explanatory variables, that is, the following requirement is satisfied: $cov(LV_j, \varepsilon_j) = 0$.

Thus, at the simulation by using the method of projection on the latent structure the requirements for the statistical distributions of variables and random deviations are absent. The external model describes two type of interconnection between latent and explicit variables: reflective and formative. Reflective type is the most widespread type of external models in which the latent variable is the cause of implicit variables, that is, explicit variables reflect latent. The external model can be written in the form of the system of linear equations: $x_{jk} = \lambda_{0jk} + \lambda_{jk} LV_j + \varepsilon_{jk}^{out}$, where λ_{jk} are load coefficients, λ_{0jk} are free members, ε_{jk}^{out} is random deviation of the external model. Formative type is a kind of external model when explicit variables are "cause" of latent variables, i.e. they «form» latent variables. External formative model can be represented in the form of a system of linear equations: $LV_j = \lambda_{0j} + \lambda_{jk} x_{jk} + \varepsilon_j^{out}$, where λ_{jk} are load coefficients, λ_{0j} are free members, ε_j^{out} is random deviation of the external model. Latent variables can not be measured directly. Therefore we enter the notion of evaluation of latent variable $Y_j = LV_j$, which is a linear combination of the corresponding implicit variables: $Y_j = \sum_k w_{jk} x_{jk}$, where w_{jk} - external weight (weight multipliers) of the model. Thus, variables LV_j and $Y_j = LV_j$ characterize one and the same entity, but are used in different purposes: the first is for theoretical, the other is for practical.

PLS-PM modelling can be represented as a three-steps algorithm:

- a) calculating of external weights w_{jk} to obtain estimates of the latent variables;
- b) computation of structural coefficients β_{jk} of internal model;
- c) calculation of loads λ_{jk} of external model.

The first step is the main and the most complicate step of PLS-PM methodology, that is iteration, its purpose is to calculate the values of external measures and, as a result, ratings Y_j . The result of this step is that abstract latent variables materialize. The second step is solution of a system of linear equations by the method of least squares, the third is calculation of correlations $\lambda_{jk} = cor(x_{jk}, Y_j)$.

3 Results and Discussion

Building PLS-PM for the model of sustainable development of travel business for Odessa region. The following table presents indicators of sustainable travel business development:

Table 1. Indicators of sustainable travel business development

Indicator	Description of the indicator	Notation
Number of tourists	Total number of tourists per calendar year	x_{11}
Wage of employees in the service sector	Total wage of employees in the service sector per calendar year	x_{12}
Number of travel company	Number of travel company in region	x_{13}
Tax revenues from the travel industry	Total tax revenues from the travel industry for calendar year	x_{14}
Subsidies for the population	Total subsidies for the population per calendar year	x_{21}
State cost on social services for the population	The sum of state cost on the protection of public safety, health and other social services for a year	x_{22}
The unemployment rate	The proportion of the total number of unemployed persons to the economically active population	x_{23}
Emissions of pollutants into atmospheric air	Quantity of the emissions of pollutants into atmospheric air in region	x_{31}
Wastes generation	Quantity of the wastes generation in region	x_{32}
State cost on the protection of the environment	The sum of state cost for the protection of the environment per calendar year	x_{33}
Current expenditures on environmental	Costs on environmental per calendar year	x_{34}
Gross Regional Product	Value of the GRP per calendar year	x_{41}
State spending on economic development	State spending on economic development per calendar year	x_{42}
Utility Costs	The amount of the state cost for utility bills populated	x_{43}
Total income	Total population income for a calendar year	x_{44}

Note that these indicators of sustainable development of the tourism business were selected from a large number of factors affecting the stable development of the region.

Following the work [5], we build a hierarchical structural model. The original model is represented in the following Figure 1. Latent variables: LV_1 - level of sustainable development of the tourism business, LV_2 – level of social condition, LV_3 - level of the environmental protection, LV_4 – condition of economics are denoted by ovals, and explicit variables (indicators from a table) by rectangles.

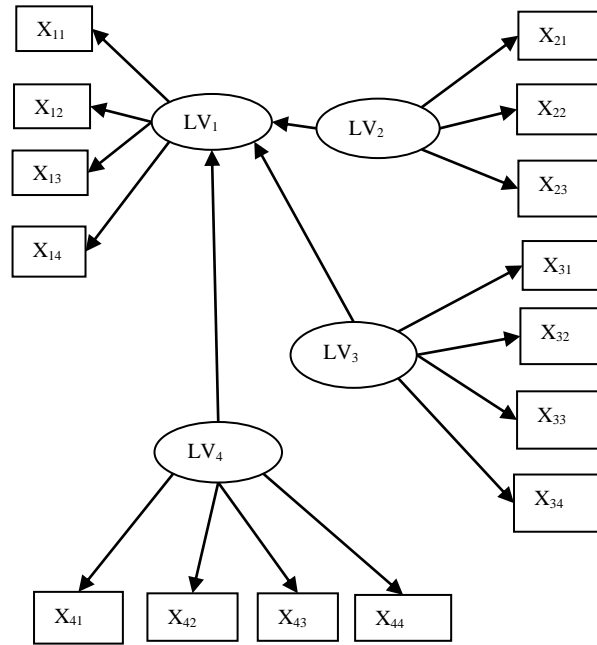


Fig. 1. Schematic representation of the basic model.
Source: Authors with XLSTAT-PM.

The relationship of the latent variables is the internal model, and the relationships between latent and explicit variables represent an external model. Let's write these models in analytical form. The internal model: $LV_1 = \beta_0 + \beta_2 LV_2 + \beta_3 LV_3 + \beta_4 LV_4 + \varepsilon_1$, where LV_j - latent variables; $\beta_2, \beta_3, \beta_4$ - structural coefficients, which characterize the strength and direction of the interconnection between latent variables, β_0 - free member and ε_1 is residual. The external model looks like the following system:

$$\begin{cases} x_{11} = \lambda_{01} + \lambda_{11} LV_1 + \varepsilon_{11} \\ \dots \\ x_{44} = \lambda_{04} + \lambda_{44} LV_4 + \varepsilon_{44} \end{cases};$$

where x_{11}, \dots, x_{44} are explicit variables, $\lambda_{11}, \dots, \lambda_{44}$ are coefficients of loads, $\lambda_{01}, \dots, \lambda_{04}$ - free members, $\varepsilon_{11}, \dots, \varepsilon_{44}$ - residuals. Note that at structural coefficients simulating by means of the partial least squares (PLS) we use the estimations of latent variables: $Y_j = LV_j = \sum_k w_{jk} x_{jk}$. For our model, these estimations can be presented in the form of the following system of equations:

$$\begin{cases} Y_1 = w_{11} x_{11} + w_{12} x_{12} + w_{13} x_{13} + w_{14} x_{14} \\ Y_2 = w_{21} x_{21} + w_{22} x_{22} + w_{23} x_{23} \\ Y_3 = w_{31} x_{31} + w_{32} x_{32} + w_{33} x_{33} + w_{34} x_{34} \\ Y_4 = w_{41} x_{41} + w_{42} x_{42} + w_{43} x_{43} + w_{44} x_{44} \end{cases};$$

where w_{11}, \dots, w_{44} are external weights of the model.

Let's check correctness of breaking the X matrix of raw data into 4 blocks X₁, X₂, X₃, X₄ (which corresponds to latent components LV₁, LV₂, LV₃, LV₄). The method Time-Wise Multi-Way Principal Component Analysis (TMPCA) is applicable. We calculate using the appropriate module of the Statistica. Note that variables should be standardized before

analysis ($z_{ij} = \frac{x_{ij} - \overline{x_{ij}}}{\sigma_j}$, $i = \overline{1,10}$; $j = \overline{1,15}$). This brings all variables to dimensionless values with mean values of

0 and standard deviations of 1. For the analysis, we use the following results of the module TMPCA:

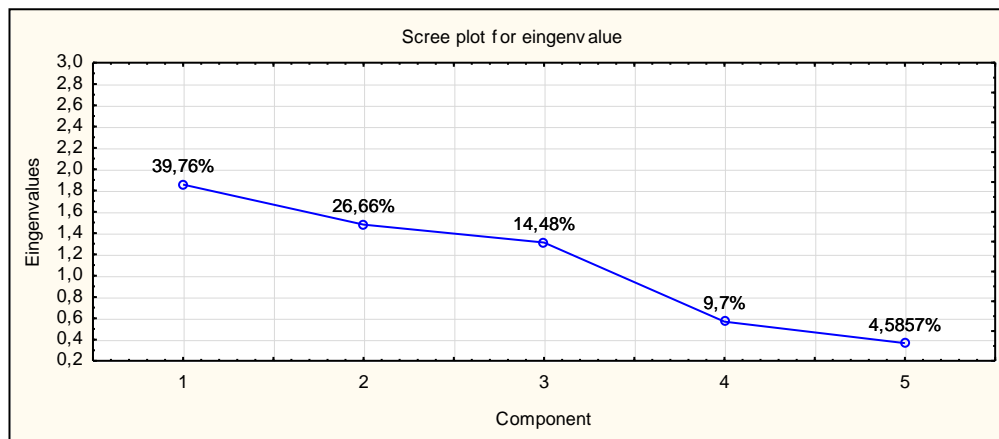


Fig. 2. Screen plot for eigenvalues.
Source: Authors with Statistica.

Table 2. Eigenvalues and percentage contribution of PCA at the total variation of components LV₁, LV₂, LV₃, LV₄.

Components	eigenvalue	% total variance	accumulated %
LV ₁	3,67872	39,76	39,76
LV ₂	2,499577	26,66	66,42
LV ₃	2,3032	14,48	80,9
LV ₄	1,8732	9,7	90,6

Source: Authors with Statistica.

Figure 2 and Table 2 show that the components-factors LV₁, LV₂, LV₃, LV₄ selected for modeling provide more than 90% of the total variation.

Analyses of the model by PLS-PM includes the following items: a) verification of the internal consistency of the blocks; b) verification of significance of variables in the external model; c) verification of cross correlations of variables in block with latent variables of other blocks; d) verification of the internal model; e) verification of the model qualification by index of conformity of model to data.

Step a). The following criteria are used to verify internal consistency in blocks: Cronbach-alpha (α) and Dillon-Goldstein (ρ) coefficients (see, for example, [5-8]), as well as eigenvalues λ_i of the correlation matrix of explicit variables (indicators).

Table 3. Checking internal consistency in blocks

Block	α	ρ	λ_1	λ_2
LV ₄	0.8843	0.9123	1.87	0.72
LV ₃	0.1009	0.0912	2.3	0.73
LV ₂	0.2012	0.5809	2.50	0.38
LV ₁	0.8547	0.9035	3.68	0.70

Source: Authors with XLSTAT-PM.

The Table 3 shows that the two blocks (LV₄ and LV₁) have good $\alpha > 0,7$ and $\rho > 0,7$ coefficients. But the LV₃ and LV₂ blocks have poor internal consistency ($\alpha, \rho < 0,7$). The following Table 4 shows the correlation coefficients of explicit and latent variables for each of the blocks.

Table 4. The correlation coefficients of explicit and latent variables for each of the blocks

Block	Variables			
	x_{31}	x_{32}	x_{33}	x_{34}
LV ₃	-0,8234	-0,9123	0,9546	0,8756
	x_{21}	x_{22}	x_{23}	
LV ₂	0.8947	0.9035	-0,8759	

Source: Authors with XLSTAT-PM.

In the block LV₃ two variables (x_{31} and x_{32}) have a negative correlation with latent variable. In the LV₂ also there is a variable with negative correlation (x_{23}). This leads to internal inconsistencies in blocks, since these indicators are so-called de-stimulators (the higher their values, the worse). For the correct application of multivariate statistical analysis methods, it is recommended to make all indicators unidirectional (as a rule, stimulants). Therefore, we will make a replacement: $x_{31}'=1- x_{31}$, $x_{32}'=1- x_{32}$, $x_{23}'=1- x_{23}$. After modification of variables we obtain the index of cooperation of external model, are presented in the following Table 5:

Table 5. Cooperation indexes of the external model

Block	α	ρ	λ_1	λ_2
LV ₄	0.8802	0.9113	1.87	0.72
LV ₃	0.8854	0.9463	2.3	0.73
LV ₂	0.8567	0.9347	2.50	0.38
LV ₁	0.8573	0.9086	3.68	0.70

Source: Authors with XLSTAT-PM.

All three of the cooperation criterions satisfy the necessary ($\alpha, \rho > 0,7$; $\lambda_1 > 1, \lambda_2 < 1$).

Step b). The second phase is verification of the external model:

Table 6. The value of the coefficients of the external model

Block	Variable	External weight w_j	Load λ_{1j}
LV ₄	x_{41}	0.5347	0.2678
	x_{42}	0.4342	0.8473
	x_{43}	0.3245	0.4565
	x_{44}	0.3641	0.3172
	$1- x_{31}$	0.3274	0.9163

LV ₃	1- x_{32}	0.3047	0.8132
	x_{33}	0.2679	0.9024
	x_{34}	0.2473	0.8271
LV ₂	x_{21}	0.3274	0.9045
	x_{22}	0.4382	0.9134
	1- x_{23}	0.3645	0.8356
LV ₁	x_{11}	0.3468	0.9768
	x_{12}	0.2173	0.7688
	x_{13}	0.1012	0.5467
	x_{14}	0.4037	0.9736

Source: Authors with XLSTAT-PM.

Variables are considered significant if the load factor $\lambda_{1j} > 0,7$. By the results of the model validation it is necessary to eliminate variables x_{13} , x_{41} , x_{43} , x_{44} . After exclusion of nonsignificant variables we get the following test results of external model:

Table 7. The value of the coefficients of the external model after modifications

Block	Variable	External weight w_j	Load λ_{1j}
LV ₄	x_{42}	0.3769	0.9151
LV ₃	1- x_{31}	0.3453	0.9273
	1- x_{32}	0.3048	0.8601
	x_{33}	0.2869	0.9041
	x_{34}	0.2283	0.8171
LV ₂	x_{21}	0.3394	0.9032
	x_{22}	0.4192	0.9081
	1- x_{23}	0.3586	0.8576
LV ₁	x_{11}	0.4071	0.9879
	x_{12}	0.2412	0.7697
	x_{14}	0.4394	0.9677

Source: Authors with XLSTAT-PM.

After checking the load factors λ_{1j} , that is, after checking the interconnection of explicit variables with latent variable of the corresponding block, you need to check the cross-loading, i.e. to determine the strength of the connection between explicit variables and latent variables of other blocks. This will exclude such indicators whose

connection with latent variable of another block is greater than with latent variable of the corresponding block. The Table 8 shows that the strength of the connection of all explicit variables with latent variables of corresponding blocks is greater than with latent variables of other blocks, that is, all variables are "loyal" to their blocks.

Table 8. The value of the coefficients of cross load

Block	Variable	Block			
		LV ₄	LV ₃	LV ₂	LV ₁
LV ₄	x_{42}	0.9071	-0.0750	0.0934	0.4374
LV ₃	1- x_{31}	0,3149	0.9273	0.5768	0.8091
	1- x_{32}	0.2133	0.8081	0.5638	0.7283
	x_{33}	-0.5463	0.9023	0.6475	0.7205
	x_{34}	-0.3148	0.8087	0.6405	0.5586
LV ₂	x_{21}	-0.0384	0.5768	0.9120	0.6879
	x_{22}	-0.1061	0.7495	0.9071	0.8412
	1- x_{23}	-0.0834	0.5205	0.8586	0.7225
LV ₁	x_{11}	-0.2412	0.8364	0.8031	0.9879
	x_{12}	0.0945	0.4273	0.6814	0.7907
	x_{14}	-0.2927	0.2489	0.8506	0.9687

Source: Authors with XLSTAT-PM.

The next step is to check the quality of the internal model. This Figure 3 is a graphic representation of the internal model with the specified values of structural coefficients:

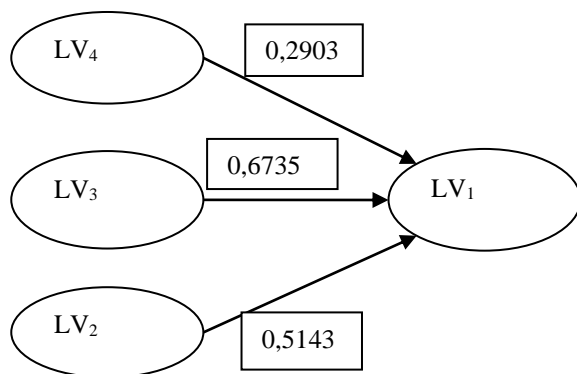


Fig. 3. Internal model with the structural coefficients.

Source: Authors with XLSTAT-PM.

In the Table 9 are estimates β_i of the equations of structural model, as well as the results of the criterion-statistics.

Table 9. Estimation of the parameters β_i of a structural model

	Estimation	Standard error	t-statistics	Pr(> t)
Free member	2,1628E-17	0.0354054	2,53E-16	1
LV ₄	0.29028279	0.1157102	1.598248	0.12
LV ₃	0.67353833	0.1663135	3.849672	0.009

LV ₂	0.51436382	0.1306488	2.968084	0.003
-----------------	------------	-----------	----------	-------

Criterion of t-statistics is held for blocks LV₃ and LV₂ ($\Pr(>|t|) < 0.05$), but not to block LV₄ ($\Pr(>|t|) > 0.05$). Following the criterion of t-statistics, the latent factor LV₄ should be excluded from consideration. But we decided to leave it in the model. After all, the impact of the level of economic development of the region on the environment and the level of social development is obvious. Our hierarchical model did not take into account the relationship between latent factors LV₄, LV₃ and LV₂. The following Table 10 presents summary statistics for the structural model.

Table 10. Total statistic of internal model

Block	R^2	Ratio explanation variation	Ratio rejecting variation AVE
LV ₄	0.00	0.79	0.89
LV ₃	0.00	0.85	0.75
LV ₂	0.00	0.89	0.79
LV ₁	0.87	0.85	0.67

Showing R^2 for target block LV₁ more than 87%. The proportion of variation characterizes share variations of a block, which reproduced latent variable this block. This characteristic for the all blocks far exceeds 50%, that positively characterises the model. In the last column of a given indicator, which characterizes the dispersion excluded share (average particle dispersion indicators unit that explains the latent variable dispersion in the block that contains the error of measurement). Fashion AVE for the all blocks exceeds 50%, therefore, this criterion of the internal model is also considered satisfactory. The fifth stage is the calculation of the single factor of quality conformity model data - GoF (Goodness-of-Fit [6]). Coefficient characterizes the quality as an internal model of the system, and external, and serves as an indicator of the reliability of the forecast model (predictive reliability model is considered to be high, if the coefficient of GoF > 70%). Our model coefficient GoF=82%. Therefore, the constructed model can be used for forecasting. Finally, let's analyze the final version of our model. The internal model can be written as the following equation:

$$LV_1 = 0,5143LV_2 + 0,6735LV_3 + 0,2903LV_4 + \varepsilon_1^{in}.$$

Estimates of latent variables are a system of equations:

$$\left\{ \begin{array}{l} LV_1 = 0,41x_{11} + 0,24x_{12} + 0,44x_{14} \\ LV_2 = 0,34x_{21} + 0,42x_{22} + 0,36(1 - x_{23}) \\ LV_3 = 0,35(1 - x_{31}) + 0,30(1 - x_{32}) + 0,29x_{33} + 0,23x_{34} \\ LV_4 = 0,38x_{42} \end{array} \right.$$

4 Conclusion

Two factors have the most significant direct impact on the level of stable development of the tourism business in the Odessa region (LV₁): LV₂ – level of social condition (comfort) and LV₃ - level of the environmental protection (ecological state of the region). The strength of their influence is 0,5143 and 0,6735 respectively. A weaker direct effect (with a strength of 0,2903) on the endogenous factor LV₁ is exerted by the exogenous factor LV₄ - condition of economics. Obviously, its influence is direct (through factors LV₂ and LV₃). According to the results obtained, the level of social comfort (LV₂) can be assessed using three factors: x_{21} - total subsidies for the population; x_{22} - the sum of state cost on the protection of public safety, health and other social services; x_{23} - the proportion of the total number of unemployed persons to the economically active population (de-stimulating effect). The level of the environmental protection (LV₃) can be assessed using 4 factors: x_{31} - emissions of pollutants into atmospheric air and x_{32} - wastes generation (de-stimulating effect); x_{33} - state cost on the protection of the environment and x_{34} - current expenditures on environmental. On improving the state of the economy (factor LV₄) is significantly influenced by x_{42} - state spending on economic development. The level of stable development of tourism business (LV₁) determined by the direct impact of

three factors: x_{11} - the number of tourists; x_{12} - wage of employees in the service sector; x_{14} - tax revenues from the travel industry. On the model you can count latent indexes for each year and create a stimulant to predict future values. Sustainability of development can be assessed by comparing the values of estimates for different periods. For example, if the estimation of the latent variables (LV_2, LV_3, LV_4) current period previous period ratings are smaller, then the corresponding block is considered to be sustainable. Or for every variables you can define a baseline with which to conduct the comparison.

References

1. L. Butowski. Sustainable Tourism – A Model Approach, Visions for Global Tourism Industry – Creating and Sustaining Competitive Strategies [online]/L. Butowski (2012). Available at: <http://www.intechopen.com/books/visions-for-global-tourism-industry-creating-and-sustaining-competitive-strategies/sustainable-tourism-a-modeling-approach> (2019)
2. R. Casagrandi. A theoretical approach to tourism sustainability [online]/R. Casagrandi, S. Rinaldi. Conservation Ecology, **1**, 6, (2002). Available at: <http://home.deib.polimi.it/rinaldi/ENS/fr5.a.pdf> (2019).
3. J. B. Greenwood. Sustainable development in a tourism destination context: a Plimsoll model of sustainability in Tyrell Country, North Carolina / J.B. Greenwood. Raleigh, NC. (2006)
4. M. Sahaidak, V. Kostynets and I. Kostynets. Modeling of development of the internal market of tourist services in Ukraine, based on the principles of cluster management. SHS Web of Conferences, **65**, 04010 (2019)
5. M. Tenenhaus, V.E. Vinzi, Y.M. Chatelin and C. Lauro. PLS Path Modeling. Computational Statistics & Data Analysis, **48**, (2005)
6. M. Tenenhaus, S. Amato and V.E. Vinzi A Global Goodness-of-Fit Index for PLS Structural Equation Modelling. Proceeding of the XLII SIS Scientific Meeting, **1** (2004)
7. M. Wtzels, G. Odekerken-Schroder and C. Van Oppen. Using PLS Path Modeling for Assessing Hierarchical Construct Models: Guidelines and Empirical Illustration. MIS Quarterly, **31** (2009)
8. G. Sanchez. PLS Path Modeling with R/Gaston Sanchez. Berkeley, California (2013)
9. State Statistics Service of Ukraine [online]. Available at: <http://www.ukrstat.gov.ua> (2020)
10. Main Department of Statistics in Odessa region [online]. Available at: <http://www.od.ukrstat.gov.ua> (2020)