Electronic scientific and practical journal

INTELLECTUALIZATION OF LOGISTICS AND SUPPLY CHAIN MANAGEMENT

#6 (2021)
April '21

WWW.SMART-SCM.ORG
ISSN 2708-3195
DOI.ORG/10.46783/SMART-SCM/2021-6

ISSN 2708-3195
9 772708 319005

Electronic scientific and practical collection
INTELLECTUALIZATION OF LOGISTICS AND SUPPLY CHAIN MANAGEMENT

Electronic scientific and practical publication in economic sciences

ISSN 2708-3195
DOI: https://doi.org/10.46783/smart-scm/2021-6

Released 6 times a year

№ 6 (2021)
April 2021

Kyiv - 2021

Founder: Viold Limited Liability Company

The electronic scientific and practical journal is registered in international scientometric data bases, repositories and search engines. The main characteristic of the edition is the index of scientometric data bases, which reflects the importance and effectiveness of scientific publications using indicators such as quotation index, h-index and factor impact (the number of quotations within two years after publishing).

In 2020, the International Center for Periodicals (ISSN International Center, Paris) included the Electronic Scientific and Practical Edition "Intellectualization of Supply Chain Management" in the international register of periodicals and provided it with a numerical code of international identification: ISSN 2708-3195 (Online).

Recommended for dissemination on the Internet by the Academic Council of the Department of Logistics NAU (No. 7 of February 26, 2020). Released 6 times a year. Editions references are required. The view of the editorial board does not always coincide with that of the authors.

t.me/smart_scm
facebook.com/Smart.SCM.org
twitter.com/ScmSmart

DOI: https://doi.org/10.46783/smart-scm/2021-6
e-mail: support@smart-scm.org

тел.: (063) 593-30-41
https://smart-scm.org

Contents

UDC 338.27:656.1
DOI: https://doi.org/10.46783/smart-scm/2021-6-1
JEL Classification: C51, C52, R12, R41.
Received: 22 February 2021

Savchenko L.V. PhD of Technical Sciences, Associate Professor, Associate Professor of Logistics Department of National Aviation University (Ukraine)

ORCID – 0000-0003-3581-6942
Researcher ID – Q-5323-2018
Scopus author id: –57208225385

Semeriahina M.M. Senior Lecturer of Logistics Department, National Aviation University (Ukraine)

ORCID – 0000-0001-7490-6874
Researcher ID – S-7158-2018
Scopus author id: –

Shevchenko I.V. PhD of Economic Sciences, Associate Professor of higher mathematics department of the National Aviation University (Ukraine)

ORCID – 0000-0001-7910-0490
Researcher ID –
Scopus author id: –

MODELING OF REGIONAL FREIGHT FLOWS OF ROAD TRANSPORT IN UKRAINE

Lidiia Savchenko, Myroslava Semeryagina, Iryna Shevchenko "Modeling of regional freight flows of road transport in Ukraine". *A transport system can be defined as a complex system characterized by a random value of transport demand, variable weather and climatic factors, a set of characteristics of transport infrastructure, and a complex system of interconnections. One of the key modes of transport providing freight transport both in domestic and international traffic is road. Its mobility and the ability to deliver cargo from door to door is a unique competitive advantage over other modes of transport.*

To create an effective logistics infrastructure that meets the demand for domestic freight transport, first of all, information is needed on the needs for transport between regions of the country. Thus, it is necessary to look for mathematical approaches to modeling freight flows, combining their practical implementation using widely used software products (for example, MS Excel).

The purpose of the paper is to build effective multifactor regression models of demand for input and output transportation of goods by road for each region of Ukraine according to publicly available statistical data of the State Statistics Service of Ukraine.

The modern approach to modeling cargo flows requires fast processing of a large amount of statistical data. In addition, the method should be as universal as possible and capable of quick and simple changes under conditions of a change in statistical data. From this point of view, the most acceptable option can be considered to be the modeling of freight traffic using regression models based on correlation and regression analysis. In *general, the task is to find the dependence of the demand for transportation on the factors that influence it. Such factors in the existing models are connected with various macroeconomic indicators, as well as the distance of delivery.*

The data of regional statistics of the State Statistics Service of Ukraine and data of the "Lardi-Trans" website as the most widely used by freight carriers and shippers were taken as the initial data for modeling.

A list of factors has been found that significantly influence the demand for freight transport by road between regions of Ukraine. A rating of influencing factors has been compiled, among which are the gross regional product, regional volumes of foreign trade in goods (imports) and gross regional product per one inhabitant of the region. The absolute values of the correlation coefficients are in the range 0.351-0.974. The lowest correlation coefficient is between the transportation distance and the demand for delivery, which proves a negligible relationship between the volume of regional transportation and the distance of delivery.

Multivariate regression models with thirteen, five and two factors of influence on demand are built. Accuracy parameter values are acceptable for all model variants. The normalized R-squared of the obtained models does not fall below 84%, and the average approximation error does not rise above 1.6%, which is an excellent performance of the models.

Keywords: demand for freight transportation, regional transportation of goods by road, domestic transportation, modeling the demand for transportation, correlation-regression analysis, linear multivariate regression.

Лідія Савченко, Мирослава Семерягіна, Ірина Шевченко "Моделювання регіональних вантажопотоків автомобільного транспорту в Україні". *Транспортну систему можна визначити як складну систему, яка характеризується випадковою величиною транспортного попиту, змінними погодно-кліматичними факторами, набором характеристик транспортної інфраструктури та складною системою взаємозв'язків. Одним з ключових видів транспорту, що забезпечує вантажні перевезення як у внутрішньому, так і міжнародному сполученні, є автомобільний. Його мобільність та можливість доставки вантажу «від дверей до дверей» є унікальною конкурентною перевагою перед іншими видами транспорту.*

Для створення ефективної логістичної інфраструктури, що забезпечує попит на внутрішні вантажні перевезення, перш за все необхідна інформація про потреби в перевезеннях між регіонами країни. Таким чином, необхідно шукати математичні підходи до моделювання вантажопотоків, комбінуючи їх практичну реалізацію з використанням широко використовуваних програмних продуктів (наприклад, MS Excel).

Метою статті є побудова ефективних багатофакторних регресійних моделей попиту на вхідні та вихідні перевезення вантажів автомобільним транспортом для кожної області України за загальнодоступними статистичними даними Державної служби статистики України.

Сучасний підхід до моделювання вантажопотоків вимагає швидкої обробки значної кількості статистичних даних. Крім того, метод має бути максимально універсальним та здатним до швидких та простих змін у разі зміни наявних статистичних даних.

Найбільш прийнятним варіантом з цієї точки зору може вважатися моделювання вантажопотоків з використанням регресійних моделей на основі проведення кореляційно-регресійного аналізу. У загальному вигляді задача полягає в знаходженні залежності попиту на перевезення від факторів, що на нього визначальний вплив. Такими факторами є макроекономічні показники, а також відстань перевезення.

У якості вихідних даних взяті дані регіональної статистики та дані сайта «Ларді-Транс» як найбільш широко використовуваного у вантажоперевізників, відправників та замовників перевезень.

Знайдено фактори, що становлять значний вплив на попит на вантажні перевезення автотранспортом з та до областей України. Складено рейтинг факторів впливу, серед яких на перших позиціях Валовий регіональний продукт; Регіональні обсяги зовнішньої торгівлі товарами (імпорт) та Валовий регіональний продукт у розрахунку на одну особу. Абсолютні значення коефіцієнтів кореляції перебувають у діапазоні 0,351-0,974. Найнижчий коефіцієнт кореляції - між

відстанню перевезень та попитом на них, що доводить незначний зв'язок між обсягами регіональних перевезень та відстанню доставки.

Побудовано моделі багатофакторної регресії з тринадцятьма, п'ятьма і двома факторами впливу на попит. Значення параметрів точності є прийнятними для всіх варіантів моделей. Нормований R-квадрат отриманих моделей не опускається нижче 84%, а середня помилка апроксимації не піднімається вище 1,6%, що є відмінними показниками моделей.

Ключові слова: попит на вантажні перевезення, регіональні перевезення вантажів автомобільним транспортом, внутрішні перевезення, моделювання попиту на перевезення, кореляційно-регресійний аналіз, лінійна багатофакторна регресія.

Лидия Савченко, Мирослава Семерягина, Ирина Шевченко "Моделирование региональных грузопотоков автомобильного транспорта в Украине". *Транспортную систему можно определить как сложную систему, которая характеризуется случайной величиной транспортного спроса, переменными погодно-климатическими факторами, набором характеристик транспортной инфраструктуры и сложной системой взаимосвязей. Одним из ключевых видов транспорта, обеспечивающим грузовые перевозки как во внутреннем, так и в международном сообщении, является автомобильный. Его мобильность и возможность доставки груза «от двери до двери» является уникальным конкурентным преимуществом перед другими видами транспорта.*

Для создания эффективной логистической инфраструктуры, обеспечивающей спрос на внутренние грузовые перевозки, прежде всего необходима информация о потребностях в перевозках между регионами страны. Таким образом, необходимо искать математические подходы к моделированию грузопотоков, комбинируя их практическую реализацию с использованием широко используемых программных продуктов (например, MS Excel).

Целью статьи является построение эффективных многофакторных регрессионных моделей спроса на входные и выходные перевозки грузов автомобильным транспортом для каждой области Украины по общедоступными статистическим данным Государственной службы статистики Украины.

Современный подход к моделированию грузопотоков требует быстрой обработки большого количества статистических данных. Кроме того, метод должен быть максимально универсальным и способным к быстрым и простым изменениям в случае перемены статистических данных. Наиболее приемлемым вариантом с этой точки зрения может считаться моделирование грузопотоков с использованием регрессионных моделей на основе проведения корреляционно-регрессионного анализа. В общем виде задача состоит в нахождении зависимости спроса на перевозки от факторов, оказывающих на него влияние. Такими факторами в существующих моделях указываются различные макроэкономические показатели, а также расстояние перевозки.

В качестве исходных данных для моделирования взяты данные региональной статистики Государственной службы статистики Украины и данные сайта «Ларди-Транс» как наиболее широко используемого у грузоперевозчиков и грузоотправителей.

Определен перечень факторов, составляющих значительное влияние на спрос на грузовые перевозки автотранспортом между областями Украины. Составлен рейтинг факторов влияния, среди которых на первых позициях валовый региональный продукт, региональные объемы внешней торговли товарами (импорт) и валовый региональный продукт в расчете на одного жителя области. Абсолютные значения коэффициентов корреляции находятся в диапазоне 0,351-0,974. Самый низкий коэффициент корреляции - между расстоянием перевозок и спросом на них, что доказывает незначительную взаимосвязь между объемами региональных перевозок и расстоянием доставки.

Построены модели многофакторной регрессии с тринадцатью, пятью и двумя факторами влияния на спрос. Значения параметров точности являются приемлемыми для всех вариантов моделей. Нормированный R-квадрат полученных моделей не опускается ниже 84%, а средняя ошибка аппроксимации не поднимается выше 1,6%, что является отличными показателями моделей.

Ключевые слова: спрос на грузовые перевозки, региональные перевозки грузов автомобильным транспортом, внутренние перевозки, моделирование спроса на перевозки, корреляционно-регрессионный анализ, линейная многофакторная регрессия.

Introduction. The transport system can be defined as a complex system characterized by a random variable of transport demand, variable weather and climatic factors, a set of characteristics of transport infrastructure and a complex system of relationships. The main purpose of the transport system is to meet the demand of the population, business and government agencies for transport services. The correspondence between the capabilities of the transport system and the demand for its services is determined by the balance of demand and capacity of the transport system. In this regard, it is very important to accurately determine the demand for transport services.

In the text of the National Transport Strategy of Ukraine for the period up to 2030 [1] the priority of the industry as one of the most important in the national economy is identified. The volume of freight traffic directly reflects the financial and economic condition of the country and its regions, as well as a marker of trends in the business environment.

One of the key modes of transport that provides freight in both domestic and international traffic is road. Its mobility and ability for door-to-door delivery is a unique competitive advantage over other modes of transport. Transportation by other modes of transport mostly needs the involvement of road in the first and last stages of the delivery process. Thus, modeling the demand for transportation of goods by road is an urgent task, the correctness of the results of which depends on the quality of the whole process of transportation.

To create an effective logistics infrastructure that meets the demand for domestic freight transport, first of all, information is needed on the demand for transport between regions of the country.

Determination of real freight traffic is associated with a number of difficulties. The most accurate is the method of direct accounting, which consists of a direct full survey of cargo-generating and cargo-absorbing points of the region. This method provides the most complete data for the characterization of traffic flows in a certain period of time. However, its disadvantage is the high labor intensity of data collection and processing. Unfortunately, the collection of such data involves direct interviewing, questioning each actual and potential point of departure and destination of cargo, which in reality is possible only in a small area (no more than a city microdistrict). The accounting of transported goods according to the nomenclature in organizations producing and consuming products, and in road transport enterprises, would certainly make it possible to easily collect the necessary information on freight traffic. However, at the moment in Ukraine there are no such reports for all enterprises, or access to them is limited. In addition, there is the problem of biased applications of consignors, lack of accounting for the frequency of transport and the weight of packages. The inaccuracy of accounting in the performed volumes of transportations in road transport enterprises also creates additional difficulties in determining the real traffic flows.

Thus, it is necessary to look for mathematical approaches to modeling freight traffic, combining their practical implementation using widely used software products (for example, MS Excel).

Analysis of the latest research. Basic principles and approaches to forecasting freight and passenger transportation are presented by the authors in [3]. Authors [7] considered the factors of influence on the evolution of transport systems, that is, they forecasted the trend regarding the volumes of transportation in dependence from factors of external environment. Analysis of trends in freight traffic in containers by the author [2] was carried out using econometric models built on the basis of time series analysis and correlation-regression analysis. Significant number of works (for example, [4]) dedicated to forecasting freight and passenger flows on the railway. It should be noted that despite identical objects (cargo), the principle of forecasting demand for carriage by rail and road should be quite different due to the nature of functioning of these means of transport and the difference of the fields of their rational application.

Formulation of the purpose of the study. The purpose of the paper is to build effective multifactor regression models of demand for incoming and outgoing transportation of goods by road for each region of Ukraine according to publicly available statistical data of the State Statistics Service of Ukraine.

Presentation of the main research. The modern approach to modeling of cargo flows requires fast processing of a impressive amount of statistical data. In addition, the method should be as universal as possible and capable of quick and easy changes under the conditions of changes in current statistics.

From this point of view, the most acceptable option can be considered to be the modeling of freight traffic with regression models based on correlation and regression analysis.

Correlation-regression analysis is a set of statistical and mathematical methods used for quantitative analysis of the links between socio-economic phenomena and processes. A random variable is used as a dependent variable in regression analysis, and non-random variables are used as an independent variable.

Regression analysis is used when the relationships between variables can be quantified as some combination of these variables. The resulting combination is used to predict the value that can take the target (dependent) variable, which is calculated on a given set of values of input (independent) variables. In the simplest case, standard statistical methods, such as linear regression, are used.

The regression model includes the following parameters and variables:

- Unknown parameters denoted as {β};
- Independent variables, {X};
- Dependent variable, {Y}.

The function $y = f\left(x_1, x_2, ..., x_n\right)$, which describes the dependence of the conditional mean value of the result characteristic (dependent variable) on the given values of arguments (independent variables) $x_1, x_2, ..., x_n$, is called the regression equation.

Linear regression is described by a linear relationship between the studied variables:

$$y = \beta_0 + \beta_1 x_1 + ... + \beta_n x_n, \quad (1)$$

where y is the dependent variable;

$x_1, x_2, ..., x_n$ – independent variables;

$\beta_0 ... \beta_n$ - regression coefficients.

The solution of the mathematical equations of the relationship between the dependent and independent factors involves the calculation of their unknown parameters from the initial data - the coefficients β_0, $\beta_1 ...$, β_n. Determination of unknown regression coefficients [5], according to which the square of the deviation of the observed (statistical) values of the performance indicator y_c is minimized from the model (obtained by the constructed regression equation) values $y_p = f(x, \beta)$. The objective function, respectively, is the expression:

$$\sum_{i=1}^{n} (y_{ci} - y_{pi})^2 \rightarrow \min \quad (2)$$

The set of coefficients *{β}*, which will provide the minimum of the objective function (2), is taken to describe the dependence of the resulting parameter y on the factors *{X}*.

Consider the stages of regression analysis.

1. Task formulation. At this stage, preliminary hypotheses about the dependence of the studied phenomena are formed. A set of factors is selected that can affect the resulting indicator.

2. Collection of statistical data. Arrays of dependent variable and independent variables are obtained.

3. Formulation of a hypothesis about the form of relationships. Choosing the form of the regression function.

4. Calculation of numerical values of parameters β of the regression equation.

5. Evaluation of the accuracy of regression analysis. Calculating the resulting error of the regression model.

6. Interpretation of the obtained results. The obtained results of regression analysis are compared with previous hypotheses. The correctness and plausibility of the obtained results are evaluated.

7. Prediction of unknown values of the dependent variable.

When conducting regression analysis, a well-grounded choice of not only the type (mathematical form) of the dependencies used, but also the factors themselves, is of great importance. That is why it is desirable to carry out correlation-regression analysis. In addition to the above steps of regression analysis it must be added the stage of estimating the correlation between the dependent and each independent variable. The degree of relationship (correlation) between the dependent and independent variables is determined using the correlation coefficient:

$$r_{xy} = \frac{\sum_i (x_i - \bar{x})(y_i - \bar{y})}{\sqrt{\sum_i (x_i - \bar{x})^2 (y_i - \bar{y})^2}}, \qquad (3)$$

where r_{xy} is the correlation coefficient between the dependent variable *y* and the independent variable *x;*

x_i, y_i – the i-th value of the independent and dependent variables;

$\bar{x}$, $\bar{y}$ – the average value of the independent and dependent variables. Only those independent variables that have high values of the correlation coefficient (usually more than 0.5 in absolute value) are left in the regression model.

In general, the task is to find the dependence of the demand for transportation on the factors that have a decisive influence on it. These factors are macroeconomic indicators. For freight transportation, this is the gross domestic product, the volume of production by industry, the volume of imported and exported goods. When forecasting passenger traffic, the main factors are the size, mobility of the population, income and tariffs for transport services. The time factor is highlighted as significant, in which all the ongoing economic and social processes and factors influencing them are accumulated [3]. In the work [7] as a factor of the external environment was taken the GDP, which reflects the efficiency of the economy. The GDP factor is also used to predict freight traffic on the railroad by the author [4], indicating a clear mutual influence of these two factors on each other.

Since the purpose of the paper is to model the demand for road freight transport between the regions of Ukraine, the data of regional statistics [8] should be taken as initial data. As for the resulting factor - the demand for transportation, we use the data of the site "Lardi-Trans" as the most widely used by carriers, shippers and customers (Table 1).

Table 1. Demand for transportation to and from the regions of Ukraine (daily statistics "Lardi-Trans" [6])

Region of Ukraine	Number of orders for transportation from the region	Number of orders for transportation to the region
Vinnytsia	128	114
Volyn	29	56
Dnepropetrovsk	318	162
Donetsk	34	41
Zhytomyr	96	71

Transcarpathian	17	57
Zaporozhye	163	153
Ivano-Frankivsk	8	44
Kyiv	1069	653
Kirovograd	39	170
Luhansk	33	35
Lviv	47	107
Mykolayivska	61	45
Odessa	117	101
Poltava	107	189
Rivne	70	58
Sumy	22	41
Ternopil	61	147
Kharkiv	166	149
Kherson	98	69
Khmelnytsky	43	62
Cherkasy	45	44
Chernivtsi	19	28
Chernihiv	24	21
Total, orders	2814	2617
Average value, orders	117.3	109.0

Source: Compiled by the authors

Since correlation-regression analysis is used in the article to model demand, we must find factors that have a significant impact on the resulting indicator (demand). The State Statistics Service of Ukraine systematically presents information on the regions of Ukraine, which could potentially have an impact on the demand for road freight:

1) Gross regional product (million UAH), latest data for 2018;

2) Gross regional product per capita (UAH), latest data for 2018;

3) Number of legal entities by region, latest data as of January 1, 2021;

4) Population as of December 1, 2020 (current population);

5) Population as of December 1, 2020 (permanent population);

6) Freight transportation by road in the region in 2020, thousand tkm;

7) Freight transportation by road in the region in 2020, thousand tons;

8) Regional volumes of foreign trade in goods in January-October 2020 (exports, thousand US dollars);

9) Regional volumes of foreign trade in goods in January-October 2020 (imports, thousand US dollars);

10) Regional volumes of foreign trade in services for 9 months of 2020 (exports, thousand US dollars);

11) Regional volumes of foreign trade in services for 9 months of 2020 (imports, thousand US dollars);

12) Volumes of manufactured construction products and indices of construction products in 2020, UAH million.

Another factor used in modeling the demand and volume of traffic for both cargo and passengers is always the distance of transportation. To model the demand for transportation to and from the regions, the average distances of transportation between all regions were calculated (Fig. 1).

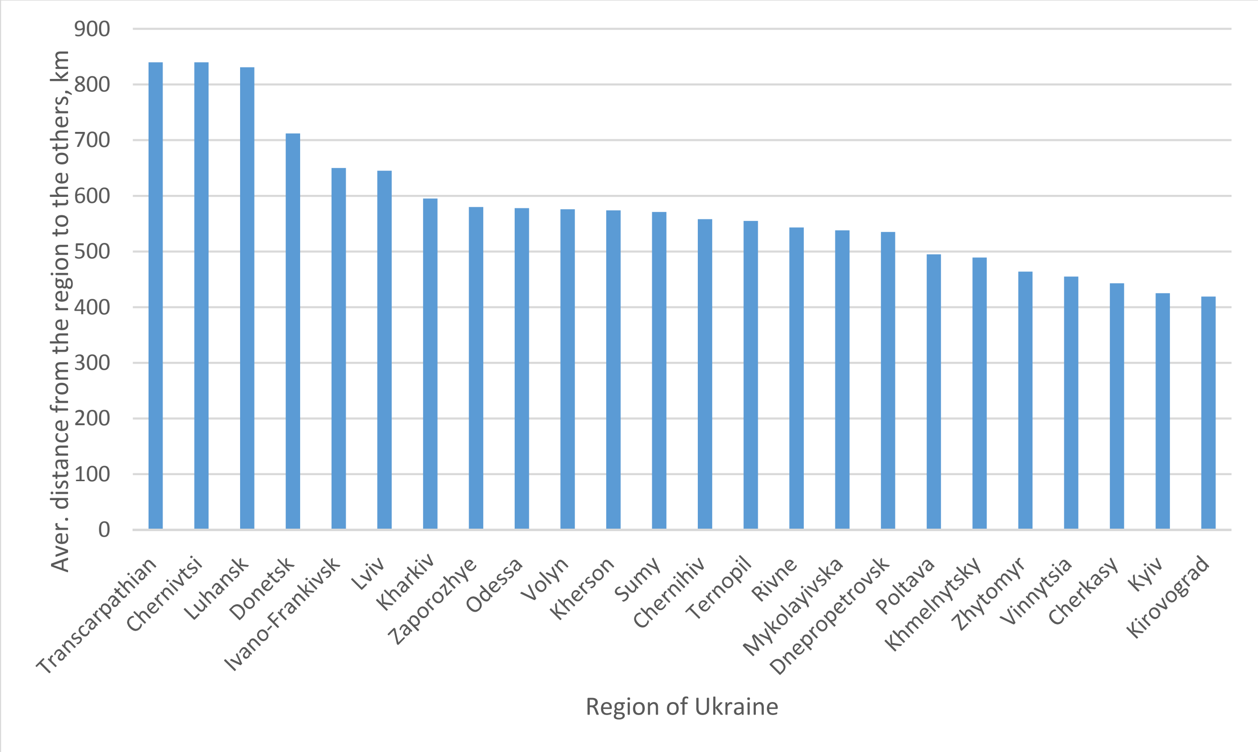


Figure 1 – Rating of regions of Ukraine in relation to the average distance of delivery to other regions

Source: compiled by the authors

Analyzing Fig. 1, it can be assumed that in regions with large average distances of connections with other regions (Transcarpathian, Chernivtsi, Luhansk, etc.), domestic transportation or transportation between neighboring oblasts will be more developed. Whereas regions that have relatively small average distances of connections with other regions (Kirovohrad, Kyiv, Cherkasy, etc.) will have a more extensive network of connections with all regions of Ukraine. However, this assumption must be verified by correlation analysis.

Here are all the statistics for each region of Ukraine, which are publicly available and updating with some regularity (Table 2).

Table 2. Regional statistics that may have an impact on the volume of road transport to and from the region of Ukraine

Region	Gross regional product (UAH million), latest data for 2018.	Gross regional product per capita (UAH), latest data for 2018.	Number of legal entities by region, latest data as of January 1, 2021	Population per 1 December 2020 (current population)	Population as of December 1, 2020 (permanent population)	Freight transport by road in the region in 2020, thousand tkm	Freight transport by road in the region in 2020, thousand tons	Regional volumes of foreign trade in goods in January-October 2020 (exports, thousand US dollars)	Regional volumes of foreign trade in goods in January-October 2020 (imports, thousand US dollars)	Regional volumes of foreign trade in services for 9 months of 2020 (exports, thousand US dollars)	Regional volumes of foreign trade in services for 9 months of 2020 (imports, thousand US dollars)	Volumes of manufactured construction products and indices of construction products in 2020, UAH mln.	The average distance from the region to other areas, km
	1	2	3	4	5	6	7	8	9	10	11	12	13
Vinnytsia	111498	71104	33012	1530930	1523845	1004447.4	5428.9	1164104.9	452562.5	117159.4	28888.1	10731.3	455.2
Volyn	60448	58297	22897	1028062	1025334	2054934.2	4792.9	530265.9	1056962.4	62854.3	21724.3	2513.7	576.0
Dnepropetrovsk	369468	114784	103645	3146125	3142816	2846710.3	20889.4	6235536.8	3728578.2	121241.4	157767.0	17756.9	535.4
Donetsk	192256	45959	91822	4103490	4090605	477847.7	19150.4	3237089.0	1211423.0	59970.2	56381.5	10122.7	712.4
Zhytomyr	77110	62911	32046	1196996	1197765	462201.4	3044.5	539426.5	426725.7	56840.2	7775.6	2075.2	463.6
Transcarpathian	52445	41706	24137	1250767	1247934	4218056	5301.9	1095201.8	1012004.7	217563.2	18478.8	1905.8	840.4
Zaporozhye	147076	85784	48725	1669239	1668450	915655.6	3960.4	2392645.1	993194.3	127978.3	20489.13	2723	579.7
Ivano-Frankivsk	78443	57033	29459	1362132	1359406	1108184.9	9199.2	628021.8	508437.9	42472.9	17698.5	3743.3	650.1
Kyiv	1031229	395618	413128	4751881	4704795	8339948	40753.5	11589891.1	21351356.7	3150834.8	1938283.14	55490.6	425.1
Kirovograd	64436	67763	25348	921695	915280	651488.9	4643.6	730598.0	203125.9	17095.0	8467.56	1366.7	418.7
Luhansk	35206	16301	41344	2122914	2118317	400129	1225.3	107562.8	170199.2	15862.1	30637.28	669.6	831.1
Lviv	177243	70173	74475	2499711	2481341	4050253	12155.3	1882326.6	2792077.1	424529	56875.32	14142	644.6
Mykolayiv	79916	70336	49939	1109932	1109217	1085824.9	6835.8	1757653.7	646778.1	254409.14	14062.93	3139.7	537.9
Odessa	173241	72738	86456	2370134	2359074	2076435.8	8818.6	1078699.4	1642397.9	621113.15	209865.26	27925.5	578.0
Poltava	174147	123763	34608	1373517	1365679	1560473.2	7403.7	1807864.4	949267.0	28788.66	65018.35	8146.4	495.2
Rivne	56842	49044	23918	1149221	1148161	1883216.6	4061.2	387869.8	300628.3819	52473.21	19584.03	3265.4	543.3
Sumy	68489	62955	25128	1055053	1052861	681840	1624	725067.9	696884.7	17656.52	17122.73	1660.2	571.4
Ternopil	49133	46833	22771	1031521	1028270	918505.8	3796	366078.7	329292.3	76222.39	7539.4	2561	555.4
Kharkiv	233321	86904	83170	2637037	2621401	2491370.5	10655.2	1160753.8	1455952.1	284570.98	37023.87	14356.2	594.7
Kherson	55161	52922	29680	1018484	1017052	755484.4	3199.1	225241.8	292176.2	23716.08	12421.23	1241.5	574
Khmelnytsky	75646	59583	30474	1245167	1242004	1095859.6	6241.3	496033.6	400786.1	17999.82	11766.18	6472.1	489.4
Cherkasy	93315	76904	29848	1180189	1176560	1599965.3	5663.3	685448.9	349155.3	30404.99	15523.28	2519.8	442.6
Chernivtsi	33903	37441	16356	897295	894230	653127.3	1155.3	130633.4	137583.8	33409.22	2671.01	2056.6	840
Chernihiv	70624	69725	23062	978434	969892	964744.7	1338.6	652492.2	289754.0	22400.8	24164.24	2440.2	558.0

Source: Compiled by the authors

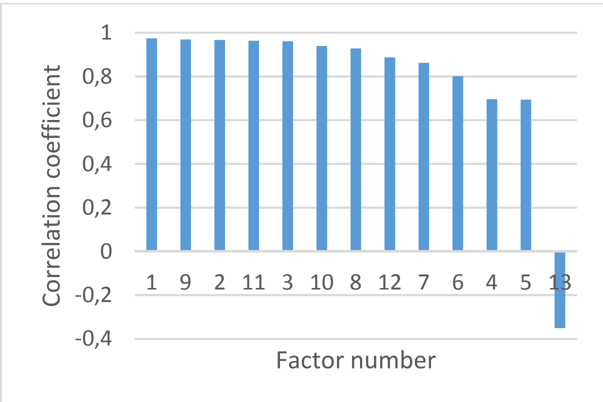
The correlation between the factors and daily demand for transportation to and from the regions of Ukraine is evaluated in Table. 3. To do this, the formula (3) was used, which in Excel is implemented with the CORREL function.

Table 3. Correlation coefficients between the volume of traffic to and from the regions of Ukraine and regional statistics

Region	Gross regional product (UAH million), latest data for 2018.	Gross regional product per capita (UAH), latest data for 2018.	Number of legal entities by region, latest data as of January 1, 2021	Population per 1 December 2020 (current population)	Population as of December 1, 2020 (permanent population)	Freight transport by road in the region in 2020, thousand tkm	Freight transport by road in the region in 2020, thousand tons	Regional volumes of foreign trade in goods in January-October 2020 (exports, thousand US dollars)	Regional volumes of foreign trade in goods in January-October 2020 (imports, thousand US dollars)	Regional volumes of foreign trade in services for 9 months of 2020 (exports, thousand US dollars)	Regional volumes of foreign trade in services for 9 months of 2020 (imports, thousand US dollars)	Volumes of manufactured construction products and indices of construction products in 2020, UAH mln.	The average distance from the region to other areas, km
	1	2	3	4	5	6	7	8	9	10	11	12	13
Transportation from the region	0.974	0.967	0.961	0.696	0.694	0.801	0.862	0.928	0.969	0.939	0.963	0.887	-0.351
Transportation to the region	0.930	0.954	0.909	0.627	0.624	0.784	0.813	0.866	0.929	0.909	0.924	0.848	-0.409

Source: Compiled by the authors

Let's make a rating of independent variables (factors of influence) on demand for transportation from and to regions of Ukraine (Fig. 2). For convenience of visual representation of material, the name of the factor is replaced by its number (Tab. 3).



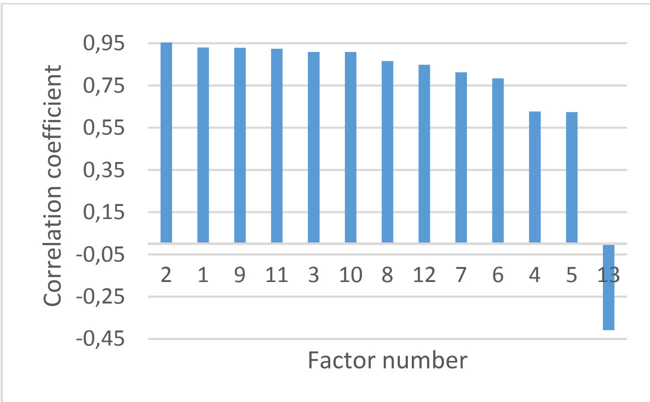


Figure 2 – Correlation coefficients of demand for freight transport by road from thirteen factors: a) from the regions of Ukraine; b) to the regions of Ukraine

Source: compiled by the authors

It can be seen that the difference between the rating of the most important factors influencing the demand for transportation from and to the regions is insignificant and is observed only in the first three factors. For the demand for transportation from the regions, the most influential factors are 1) Gross regional product; 2) Regional volumes of foreign trade in goods (imports); 3) Gross regional product per capita. Whereas for the demand for transportation to the regions the most influential factors are 1) Gross regional product per capita; 2) Gross regional product; 3) Regional volumes of foreign trade in goods (imports). Absolute values of correlation coefficients are in the range of 0.351-0.974 (demand for transportation from the regions) and 0.409-0.954 (demand for transportation to the regions). It can prove that all thirteen factors have a significant impact on the demand for transportation both from and to the Ukrainian regions. The lowest and rather insignificant correlation coefficient between the demand for transportation and the distance is surprising. After all, the distance of transportation has always been one of the determining factors in modeling the demand for transportation of both goods and passengers. The research conducted here proves that the formation of demand for regional transportation in Ukraine is almost without the influence of the factor of the distance between the points of departure and receipt. In addition, the distance factor is the only one of others that has a negative correlation coefficient. This indicates the inverse relationship between the demand for transportation and its distance. This conclusion is quite logical, because with increasing distance, the demand for transportation should fall, and vice versa. Such interdependence is based on the desire to save money and time, because long-distance transportation is more expensive than short-distance one.

For the next stage - regression analysis - we will transfer the daily demand obtained from the one-time statistics of "Lardi-Trans" into the annual one, multiplying it by the number of working days in a year (251). A regression analysis was carried out using the "Regression" tool of the "Data Analysis" add-in for MS Excel spreadsheets. The analysis results are shown in Fig. 3 and Fig. 4.

RESULTS

Regression statistics	
Multiple R	0.996
R-square	0.993
Normalized R-square	0.983
Standard error	6955.482
Observations	24

ANOVA

	df	*SS*	*MS*	*F*	Significance of F
Regression	13	65869905471.2	5066915805.5	104.734	0.000
Residuals	10	483787245.3	48378724.5		
Total	23	66353692716.5			

	Coefficients	Standard error	t-statistic	P-Value	Lower 95%	Upper 95%	Lower 99.0%	Upper 99.0%
Y-intersection	61602.722	23999.435	2.567	0.028	8128.649	115076.795	-14458.031	137663.475
Variable X 1	0.66859	0.179	3.728	0.004	0.269	1.068	0.100	1.237
Variable X 2	-0.65344	0.272	-2.401	0.037	-1.260	-0.047	-1.516	0.209
Variable X 3	-0.36900	0.485	-0.761	0.464	-1.450	0.712	-1.906	1.168
Variable X 4	-0.97183	0.611	-1.591	0.143	-2.332	0.389	-2.907	0.964
Variable X 5	0.96527	0.616	1.567	0.148	-0.407	2.338	-0.987	2.918
Variable X 6	-0.00116	0.003	-0.343	0.738	-0.009	0.006	-0.012	0.010
Variable X 7	-1.75311	0.847	-2.070	0.065	-3.640	0.134	-4.437	0.931
Variable X 8	0.00001	0.004	0.003	0.998	-0.008	0.008	-0.012	0.012
Variable X 9	-0.00671	0.006	-1.033	0.326	-0.021	0.008	-0.027	0.014
Variable X 10	0.05463	0.043	1.272	0.232	-0.041	0.150	-0.082	0.191
Variable X 11	0.07424	0.051	1.466	0.173	-0.039	0.187	-0.086	0.235
Variable X 12	-1.25481	0.733	-1.712	0.118	-2.888	0.378	-3.578	1.068
Variable X 13	-41.664	21.869	-1.905	0.086	-90.390	7.063	-110.972	27.644

Figure 3 – Regression analysis - data from the regions
Source: compiled by the authors

RESULTS

Regression statistics	
Multiple R	0.966
R-square	0.933
Normalized R-square	0.846
Standard error	12459.303
Observations	24

ANOVA

	df	*SS*	*MS*	*F*	Significance of F
Regression	13	2.17E+10	1.67E+09	10.7468	0.0003
Residuals	10	1.55E+09	1.55E+08		
Total	23	2.32E+10			

	Coefficients	Standard error	t-statistic	P-Value	Lower 95%	Upper 95%	Lower 99.0%	Upper 99.0%
Y-intersection	55527.9	42990.012	1.292	0.226	-40259.784	151315.650	-80719.139	191775.004
Variable X 1	0.494	0.321	1.538	0.155	-0.222	1.210	-0.524	1.512
Variable X 2	-0.315	0.488	-0.646	0.533	-1.401	0,771	-1.860	1.230
Variable X 3	-1.120	0.869	-1.290	0.226	-3.056	0.815	-3.874	1.633
Variable X 4	0.569	1.094	0.520	0.614	-1.868	3.007	-2.897	4.036
Variable X 5	-0.564	1.104	-0.511	0.620	-3.023	1.895	-4.062	2.933
Variable X 6	-0.004	0.006	-0.683	0.510	-0.018	0.009	-0.023	0.015
Variable X 7	-0.183	1.517	-0.121	0.906	-3.563	3.198	-4.991	4.625
Variable X 8	0.000	0.007	0.066	0.949	-0.015	0.016	-0.021	0.022
Variable X 9	-0.005	0.012	-0.409	0.691	-0.031	0.021	-0.042	0.032
Variable X 10	0.103	0.077	1.334	0.212	-0.069	0.274	-0.141	0.347
Variable X 11	0.029	0.091	0.321	0.755	-0.173	0.231	-0.258	0.317
Variable X 12	-1.593	1.313	-1.213	0.253	-4.518	1.332	-5.754	2.568
Variable X 13	-40.10	39.173	-1.024	0.330	-127.382	47.185	-164.250	84.052

Figure 4 – Regression analysis - data to the regions
Source: compiled by the authors

The results obtained allow us to make the following conclusions. Since the values of the multiple correlation coefficient are 0.996 and 0.966 (for the demand for transportation from and to the region, respectively), which is more than the accepted boundary value of 0.7, we can say about a strong relationship between the dependent value (demand for freight transportation by road between regions of Ukraine) and selected thirteen factors. The values of the adjusted coefficient of determination mean that, respectively, 98.3 and 84.6% of the variations in the demand for transportation from and to the region are explained by the variation of the selected thirteen factors, and the remaining 1.7 and 15.4% are explained by other factors unaccounted in this regression model. The values obtained are high and acceptable for decision making. The next block is analysis of variance. The Fisher test is used to check the statistical significance of the regression equation as a whole. The actual values of the criterion are 104.7 and 10.8 with the corresponding significance levels of about 0. Thus, the actual values of the Fisher test significantly exceed the significance levels, which means that the regression equation can be recognized as statistically significant with a very high probability. The next section contains information about the values of the regression coefficients. The resulting models are follows:

- annual demand for transportation from the region:

$$y = 61602,7 + 0,67x_1 - 0,65x_2 - 0,37x_3 - 0,97x_4 + 0,97x_5 - 0,001x_6 -$$
$$-1,75x_7 + 0,0001x_8 - 0,01x_9 + 0,06x_{10} + 0,07x_{11} - 1,26x_{12} - 41,66x_{13}$$

- annual demand for transportation to the region:

$$y = 55527,9 + 0,49x_1 - 0,32x_2 - 1,12x_3 + 0,57x_4 - 0,56x_5 - 0,004x_6 -$$
$$-0,18x_7 + 0,0001x_8 - 0,01x_9 + 0,1x_{10} + 0,03x_{11} - 1,59x_{12} - 40,1x_{13}$$

The free coefficient of the equation can be considered as the value of the demand for regional transportation, which does not depend on the selected factors. Surprisingly, the coefficients in the equations do not always coincide in sign with the corresponding correlation coefficients. The obtained values are checked for statistical significance using the Student's t-test. The p-value column contains the significance levels at which the regression coefficients are considered statistically significant different from zero. The limiting value in practice is usually considered to be 0.05 (95% probability). If the actual p-value is less than the limiting, the regression coefficient is considered statistically significant. It can be seen that such a condition is satisfied for a small number of factors. According to p-values, only the coefficients at x_1 and x_2 are statistically significant for modeling demand from the region (values 0.004 and 0.037). None of the regression coefficients that model demand to regions are statistically significant. To increase the significance of the regression coefficients, the number of factors should be reduced by removing those of them that have the greatest pairwise correlation. It should be noted that the regression coefficients can also be obtained using the LINEST function in MS Excel. However, the function does not provide additional information about the resulting model; thus, information about the R-square estimates, Fisher's tests, t-statistics, etc. will have to be obtained using additional calculations, including using MS Excel functions. To reduce the number of independent regression variables, we will carry out a correlation analysis - calculate the matrix of pair correlation coefficients. For this purpose, we use the Correlation tool of the Data Analysis in MS Excel Spreadsheet Add-in. The correlation matrix is presented in Table 4 and Table 5.

Table 4. Pairwise correlations matrix (demand for transportation from the region)

	Transportation from the region	1	2	3	4	5	6	7	8	9	10	11	12	13
Transportation from the region														
1	0.974													
2	0.967	0.963												
3	0.961	0.983	0.937											
4	0.696	0.804	0.631	0.817										
5	0.694	0.801	0.628	0.814	1.000									
6	0.801	0.833	0.808	0.826	0.612	0.609								
7	0.862	0.940	0.847	0.927	0.891	0.890	0.788							
8	0.928	0.962	0.905	0.932	0.811	0.810	0.782	0.946						
9	0.969	0.972	0.962	0.980	0.714	0.711	0.866	0.886	0.919					
10	0.939	0.933	0.934	0.966	0.673	0.669	0.860	0.832	0.850	0.978				

11	0.963	0.950	0.957	0.970	0.673	0.670	0.822	0.846	0.881	0.990	0.984			
12	0.887	0.929	0.870	0.933	0.804	0.801	0.810	0.887	0.854	0.895	0.910	0.889		
13	-0.351	-0.306	-0.436	-0.238	0.013	0.015	-0.104	-0.225	-0.263	-0.257	-0.239	-0.270	-0.273	

Source: compiled by the authors

Table 5. Pairwise correlations matrix (demand for transportation to the region)

	Transportation to the region	1	2	3	4	5	6	7	8	9	10	11	12	13
Transportation to the region														
1	0.930													
2	0.954	0.963												
3	0.909	0.983	0.937											
4	0.627	0.804	0.631	0.817										
5	0.624	0.801	0.628	0.814	1.000									
6	0.784	0.833	0.808	0.826	0.612	0.609								
7	0.813	0.940	0.847	0.927	0.891	0.890	0.788							
8	0.866	0.962	0.905	0.932	0.811	0.810	0.782	0.946						
9	0.929	0.972	0.962	0.980	0.714	0.711	0.866	0.886	0.919					
10	0.909	0.933	0.934	0.966	0.673	0.669	0.860	0.832	0.850	0.978				
11	0.924	0.950	0.957	0.970	0.673	0.670	0.822	0.846	0.881	0.990	0.984			
12	0.848	0.929	0.870	0.933	0.804	0.801	0.810	0.887	0.854	0.895	0.910	0.889		
13	-0.409	-0.306	-0.436	-0.238	0.013	0.015	-0.104	-0.225	-0.263	-0.257	-0.239	-0.270	-0.273	

Source: compiled by the authors

It can be observed that from all factors, only the 13th is independent from the other factors (although there is some connection with factor 2).

We offer the following method of successive exclusion of factors from the model:

1. Find the largest number in the matrix of pairwise correlations.

2. Of the two factors, the pair of which has the maximum value of pair correlation, leave in the model the factor that has the greatest value of the correlation with the final (dependent) factor.

3. Repeat the procedure of steps 1-2 until the regression coefficients are significant according to the p-value or a sufficient level of accuracy is achieved (for example, an acceptable average approximation error is obtained):

$$\overline{\varepsilon} = \frac{1}{n}\sum_{i}\left|\frac{y_{\phi i} - y_{mi}}{y_{\phi i}}\right| \cdot 100\%,$$

where $y_{\phi i}$ - i-th value of the statistical series of the dependent quantity;

y_{mi} - i-th value of the theoretical series of the dependent quantity, calculated using the obtained regression equation.

There are three variants of the regression model were considered:

- with the maximum number of factors (thirteen);
- with five factors;
- with two factors.

The reduction of the obtained basic model with thirteen factors was carried out using the described above algorithm. The linear regression equations obtained using the above technique, as well as the characteristics of the models, are grouped in Table 6 and Table 7.

Table 6. Regression analysis from the regions

	13 factors	5 factors	2 factors
List of factors in the model	1-13	1, 4, 6, 12, 13	1, 13
Regression model	$y = 61602,7 + 0,67x_1 - 0,65x$ $-0,37x_3 - 0,97x_4 + 0,97x_5 -$ $-0,001x_6 - 1,75x_7 + 0,0001x_8$ $-0,01x_9 + 0,06x_{10} + 0,07x_{11} -$ $-1,26x_{12} - 41,66x_{13}$	$y = -956,3 + 0,35x_1 - 0,02x$ $-0,004x_6 - 0,08x_{12} + 20,24$	$y = 6867,2 + 0,25x_1 - 25,7$
Normalized R-square, %	98.3	96.6	94.8
Checking Fisher's criterion	Regression equation as a whole is statistically significant	Regression equation as a whole is statistically significant	Regression equation as a whole is statistically significant
Significant factors according to p-value	1. 2	1. 4	1
Average approximation error, %	1.18	1.54	1.14

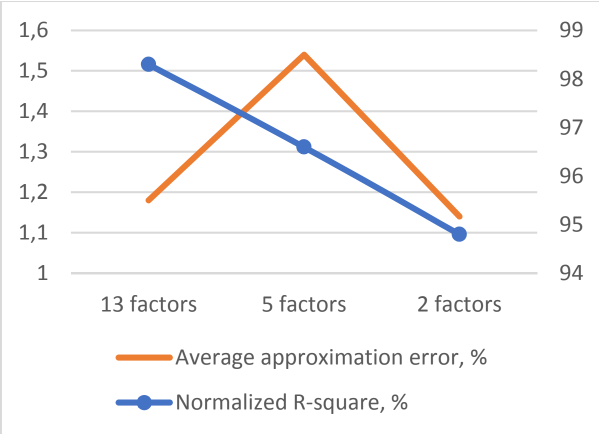
Source: compiled by the authors

Table 7. Regression analysis to the regions

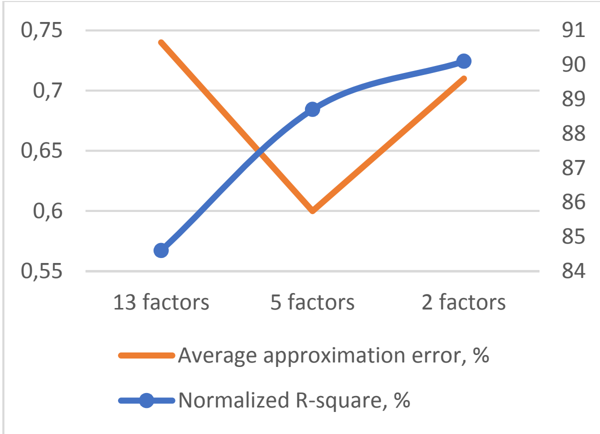
	13 factors	5 factors	2 factors
List of factors in the model	1-13	2, 4, 6, 12, 13	2, 13
Regression model	$y = 55527,9 + 0,49x_1 - 0,32$ $-1,12x_3 + 0,57x_4 - 0,56x_5 -$ $-0,004x_6 - 0,18x_7 + 0,0001$ $-0,01x_9 + 0,1x_{10} + 0,03x_{11} -$ $-1,59x_{12} - 40,1x_{13}$	$y = -4376 + 0,39x_2 + 0,0008$ $+0,0005x_6 + 0,11x_{12} - 3,79x_1$	$y = -7684,2 + 0,43x_2 + 2,2$
Normalized R-square, %	84.6	88.7	90.1
Checking Fisher's criterion	Regression equation as a whole is statistically significant	Regression equation as a whole is statistically significant	Regression equation as a whole is statistically significant
Significant factors according to p-value	-	2	2
Average approximation error, %	0.74	0.60	0.71

Source: compiled by the authors

The normalized R-square and the average approximation error were analyzed for the three scenarios (Fig. 5)



a)



b)

Figure 5 – Values of parameters of regression model accuracy for three scenarios of modeling of demand for freight transportations by road transport: a) from the regions; b) to the regions

Source: compiled by the authors

It can be observed that the values of the accuracy parameters are acceptable for all six variants of the models. The normalized R-square does not fall below 84%, and the average approximation error does not rise above 1.6%. If the decision on the best variant of the regression equation is made on the basis of the average error of approximation, then for modeling of demand from the regions it is necessary to choose two-factor model, and for modeling of demand for the regions - five-factor. If the optimality criterion is a normalized R-square, the best choice for modeling the demand from the regions will be a 13-factor model, and for modeling the demand for the region - a two-factor.

The perspectives of the research should be the choice of the best regression equation that would describe the demand for regional transportation in Ukraine. After that, it will be possible to create a matrix of demand for transportation between all regions of Ukraine.

Conclusions. Regional freight transportation by road ensures the satisfaction of demand for goods within the country and is necessary for the smooth operation of manufacturing and service enterprises. Forecasting the demand for transportation between regions and the subsequent planning and organization of freight flows are important economic tasks. In this regard, reliable mathematical models are needed to predict regional freight traffic by road transport based on annual (quarterly) statistical data for the regions of Ukraine.

The performed correlation-regression analysis made it possible to establish the functional dependences of the demand for road transportation of goods to and from the Ukrainian region, namely, the multiple regression equations. The main factors that have the greatest impact on interregional transportation are determined, and a comparative analysis of models with thirteen, five and two linear regression factors is carried out. Indicators of errors and reliability allow us to speak about the sufficient accuracy of the model in relation to real data obtained using the platform for the search for cargo and transport "Lardi-Trans".

References

1. «Pro skhvalennia Natsionalnoi transportnoi stratehii Ukrainy na period do 2030 roku». Rozporiadzhennia Kabinetu Ministriv Ukrainy vid 30 travnia 2018 r. № 430-r. [Electronic resource]. – Access mode: https://zakon.rada.gov.ua/laws/show/430-2018-p#Text.
2. Avtomonova L.Yu. Prognozirovanie ob'Yomov konteynernyih perevozok s ispolzovaniem ekonometricheskih modeley. *Sistemnyiy analiz i logistika*. 2018. #1(16). p.60-69.
3. Borisevich V.I., Kandaurova G.A., Kandaurov N.N. i dr. Prognozirovanie i planirovanie ekonomiki: Ucheb. posobie. Mn. Interpresservis; *Ekoperspektiva*. 2001. 380 s.
4. Gulamov A. A. Prognozirovanie ob'Yomov perevozok gruzov na uzbekskoy zheleznoy doroge. *Izvestiya Peterburgskogo universiteta putey soobscheniya*. 2010. #1. [Electronic resource]. – Access mode: https://cyberleninka.ru/article/n/prognozirovanie-obyomov-perevozok-gruzov-na-uzbekskoy-zheleznoy-doroge.
5. Ivashchuk O.T. ta in. Ekonomiko–matematychne modeliuvannia: navchalnyi posibnyk. Ternopil: *Ekonomichna dumka*, 2008. 701 s.
6. Lardi-Trans. [Electronic resource]. – Access mode: https://lardi-trans.com.
7. Lynnyk I. E., Sanko Ya.V. Shchodo vyznachennia vplyvu zmin zovnishnoho seredovyshcha na evoliutsiiu transportnykh system. *Skhidnoukrainskyi zhurnal peredovykh tekhnolohii*. 2012. 5/3 (59). S. 14 – 16.
8. State Statistics Service of Ukraine. [Electronic resource]. – Access mode: http://ukrstat.gov.ua.