

УДК 519.67:339.5

DOI: <https://doi.org/10.32782/2415-8801/2019-6.2>

Дзюбановська Н.В.

кандидат економічних наук,
доцент кафедри прикладної математики,
Тернопільський національний економічний університет

Єрмоєнко В.О.

кандидат фізико-математичних наук, доцент,
доцент кафедри прикладної математики,
Тернопільський національний економічний університет

Сенів Г.В.

викладач кафедри прикладної математики,
Тернопільський національний економічний університет

ЗАСТОСУВАННЯ МЕТОДІВ БІНАРНОЇ КЛАСИФІКАЦІЇ ДО ОЦІНЮВАННЯ МІЖНАРОДНОЇ ТОРГІВЛІ

У статті розглянуто приклад застосування методів бінарної класифікації до оцінювання міжнародної торгівлі країн Європейського Союзу засобами пакету програм STATISTICA 10, також проаналізовано, як демографічна ситуація у країні та добробут населення впливають на показник додатного сальдо торговельного балансу товарів і послуг країн ЄС. Для моделювання як залежну змінну вибрано дихотомічну величину індексу ймовірності додатного сальдо торговельного балансу країн ЄС. З використанням статистичних даних країн ЄС протягом 2008–2016 рр. За результатами аналізу побудовано дві ймовірнісні моделі бінарного вибору для встановлення залежності індексу ймовірності додатного сальдо товарів і послуг країн ЄС від показників демографічної ситуації у країні та матеріального забезпечення населення (на основі логістичної функції (Logit-модель) та функції стандартного нормального розподілу (Probit-модель)).

Ключові слова: бінарна класифікація, ймовірність, міжнародна торгівля, модель, оцінювання, регресія.

ПРИМЕНЕНИЕ МЕТОДОВ БИНАРНОЙ КЛАССИФИКАЦИИ К ОЦЕНКЕ МЕЖДУНАРОДНОЙ ТОРГОВЛИ

Дзюбановская Н.В., Ерёмченко В.А., Сенив Г.В.

В статье рассмотрен пример применения методов бинарной классификации к оценке международной торговли стран Европейского Союза средствами пакета программ STATISTICA 10, также проанализировано, как демографическая ситуация в стране и благосостояние населения влияют на показатель положительного сальдо торгового баланса товаров и услуг стран ЕС. Для моделирования в качестве зависимой переменной выбрана дихотомическая величина индекса вероятности положительного сальдо торгового баланса стран ЕС с использованием статистических данных стран ЕС за период 2008–2016 гг. По результатам анализа построены две вероятностные модели бинарного выбора для установления зависимости индекса вероятности положительного сальдо товаров и услуг стран ЕС от показателей демографической ситуации в стране и материального обеспечения населения (на основе логистической функции (Logit-модель) и функции стандартного нормального распределения (Probit-модель)).

Ключевые слова: бинарная классификация, вероятность, международная торговля, модель, оценка, регрессия.

THE APPLICATION OF BINARY CLASSIFICATION METHODS FOR EVALUATION OF INTERNATIONAL TRADE

Dziubanovska Nataliia, Yeromenko Valerii, Seniv Halyna

In this article an example of applying binary classification methods to assessment European Union countries' international trade was considered by using the STATISTICA 10 software package. International trade is a complex socio-economic process that undergoes constant structural and dynamic changes. These changes are caused by numerous factors and some of them are categorical variables. Binary classification methods are used to model the process that is described with such parameters. The purpose of this research is to evaluate the countries' trade and to consider the impact of both quantitative and categorical factors on the trade process by using binary classification methods. The impact of the countries' demographic situation and the well-being of the population on the positive balance of EU countries' trade of goods and services was analyzed. The dichotomous value index of the probability of the positive EU countries' trade balance of goods and services was selected as a dependent variable for modeling. The EU countries' data statistics during 2008–2016 were used for researching. Based on the analysis, it could be conclude that the positive balance of EU countries' trade is more

influenced by the immigration of the population in the country rather than total number of the population or the migration of the population. The simulation's results also confirm that the positive balance of the country's trade depends on the purchasing power of the population. Such indicators as GDP per capita, average income, net income, wages and salaries are significant factors of influencing on main trade's indicators. According to the results of the analysis, two probabilistic models of binary choice for evaluation of the dependence of the probability of trade's positive balance of goods and services from the countries' demographic situation and the well-being of the countries' population were constructed (on the basis of logistic function (Logit-model) and on the basis of standard normal distribution function (Probit-model)). The adequacy of the constructed models was verified by means of ROC analysis and analysis of the models' residuals.

Keywords: binary classification, evaluation, international trade, model, probability, regression.

Постановка проблеми. Для оцінювання динаміки економічних процесів та дослідження впливу чинників на цей процес використовують чималу кількість економічних моделей. Більшість із них взаємодіє лише з кількісними величинами, хоча на практиці часто зустрічаються ситуації, в яких залежна змінна – дихотомічна чи деякі фактори впливу – категоріальні. Міжнародна торгівля як складний соціально-економічний процес зазнає постійних структурних та динамічних змін, що зумовлені численною кількістю чинників, частина яких є категоріальними. Для моделювання процесу заданого такими параметрами застосовують методи бінарної класифікації. Великою популярністю серед них користуються регресійні моделі бінарного вибору *Logit* і *Probit*.

Аналіз останніх досліджень і публікацій. Регресійні моделі бінарного вибору надзвичайно часто використовують у медичних дослідженнях. Проте останні два десятиліття їх усе частіше застосовують до оцінювання економічних процесів, зокрема для знаходження ймовірності настання кризових явищ, банкрутства тощо. У процесі дослідження соціально-економічних процесів питанням використання методів бінарної класифікації приділяють увагу такі науковці, як: І.Г. Лук'яненко [1], О.К. Малютін [2], Н. Міщук [3], В.І. Сушко, Т.С. Павлюк [4], Ю.І. Швабу [5] та ін.

Постановка завдання. Метою дослідження є оцінювання торгівлі країн та дослідження впливу як кількісних, так і категоріальних факторів на процес торгівлі із застосуванням методів бінарної класифікації.

Виклад основного матеріалу дослідження. Під час побудови регресійної моделі бінарного вибору залежною змінною вибирають бінарну величину, а чинники впливу можуть бути як кількісними, так і категоріальними. Таким чином, у моделях бінарного вибору будується регресійна модель залежності ймовірності того, що результативна дихотомічна змінна набуде значення 0 або 1 за заданих значень факторів [6]. У регресійних моделях бінарного вибору *Logit* і *Probit* використовують відповідно логістичну функцію або функцію стандартного нормального розподілу.

Модель бінарного вибору на основі логістичної функції називається логістичною регресією, або *Logit*-моделлю. Модель бінарного вибору на основі функції стандартного нормального розподілу називається *Probit*-моделлю [7].

Значними перевагами застосування методів бінарної класифікації є статистична надійність результатів та легкість їх інтерпретації.

Розглянемо приклад застосування регресійних моделей бінарного вибору до оцінювання міжнародної торгівлі країн ЄС [8]. З'ясуємо, за яких умов спостерігатиметься додатне сальдо торговельного балансу для певної країни, і виберемо фактори, що впливатимуть на це.

Відомо, що добробут і купівельна спроможність населення мають значний вплив на торгівлю країни, крім того, чисельність і міграційні процеси у країні теж відображаються на числових показниках обсягів експортно-імпортних операцій. Отже, проаналізуємо як демографічна ситуація у країні та матеріальне забезпечення населення впливають на показник додатного сальдо торговельного балансу.

Для моделювання як залежну змінну виберемо Y – індекс імовірності додатного сальдо товарів і послуг країн ЄС протягом 2008–2016 рр. [9], де

$$Y = \begin{cases} 1, & \text{сальдо додатне;} \\ 0, & \text{сальдо від'ємне.} \end{cases}$$

Як фактори впливу розглянемо: чисельність населення (X_1), загальну національну концепцію зайнятості населення, % (X_2), еміграцію населення, % (X_3), імміграцію населення, % (X_4), ВВП на душу населення, євро (X_5), середній дохід населення, євро (X_6), чистий прибуток населення, євро (X_7), заробітну плату населення, євро (X_8), використання населенням Інтернету, % (X_9).

Застосувавши статистичні дані країн ЄС протягом 2008–2016 рр. [9] та модуль *Generalized Linear / Nonlinear* програмного продукту *STATISTICA 10*, отримуємо *Logit*-модель для залежної біноміальної змінної індексу ймовірності додатного сальдо товарів і послуг країн ЄС (рис. 1).

За результатами моделювання бачимо, що значущими є фактори X_4 , X_5 , X_6 , X_7 , X_8 і X_9 . Тобто із запропонованих факторів на показник додатного сальдо торгівлі впливають імміграція населення (X_4), ВВП на душу населення (X_5), середній дохід населення (X_6), чистий прибуток населення (X_7), заробітна плата населення (X_8) та використання населенням Інтернету (X_9).

На рис. 2 зображено параметри для побудови рівняння логістичної регресії.

Імовірність того, що сальдо торговельного балансу буде додатним, визначається за формулою:

$$Y = \frac{e^Z}{1 + e^Z},$$

де Z – лінійна регресійна модель, що описує зв'язок між незалежними факторами (рис. 2):

$$Z = -9,74 + 1,1X_4 + 0,00044X_5 - 0,00035X_6 - 0,00041X_7 + 0,111X_9$$

Зважаючи на те, що коефіцієнт для змінної X_8 надзвичайно малий, даним фактором можна знехтувати. Це тому, що ми використовували загальні статистичні дані заробітної плати населення країн ЄС без ранжування за посадовими окладами, віковими категоріями тощо.

Для аналізу якості побудованої логістичної моделі функціональним є застосування *ROC*-аналізу (*Receiver Operator Characteristic*) (рис. 3).

Effect	Saldo, Y - Test of all effects (Spreadsheet8) Distribution : BINOMIAL, Link function: LOGIT Modeled probability that Saldo, Y = yes		
	Degr. of Freedom	Wald Stat.	p
Intercept	1	12,55206	0,000396
Population, X1	1		
Total employment national concept, X2	1	0,01045	0,918592
Emigration, X3	1	0,14837	0,700101
Immigration, X4	1	6,74097	0,009422
GDP per capita, X5	1	14,46390	0,000143
Median equivalised net income, X6	1	6,78218	0,009207
Net earnings, X7	1	6,29483	0,012109
Wages and salaries, X8	1	5,24076	0,022064
Internet use, X9	1	32,10806	0,000000

Рис. 1. Результати побудованої моделі *Logistic regression*

Effect	Saldo, Y - Parameter estimates (Spreadsheet8) Distribution : BINOMIAL, Link function: LOGIT Modeled probability that Saldo, Y = yes							
	Level of Effect	Column	Estimate	Standard Error	Wald Stat.	Lower CL 95, %	Upper CL 95, %	p
Intercept		1	-9,73653	2,74819	12,55206	-15,123	-4,3502	0,000396
Population, X1		2	0,00000	0,00000	2,82549	-0,000	0,0000	0,092779
Total employment national concept, X2		3	5,65910	55,36843	0,01045	-102,861	114,1792	0,918592
Emigration, X3		4	0,14436	0,37477	0,14837	-0,590	0,8789	0,700101
Immigration, X4		5	1,09309	0,42101	6,74097	0,268	1,9183	0,009422
GDP per capita, X5		6	0,00044	0,00012	14,46390	0,000	0,0007	0,000143
Median equivalised net income, X6		7	-0,00035	0,00014	6,78218	-0,001	-0,0001	0,009207
Net earnings, X7		8	-0,00041	0,00016	6,29483	-0,001	-0,0001	0,012109
Wages and salaries, X8		9	-0,00000	0,00000	5,24076	-0,000	-0,0000	0,022064
Internet use, X9		10	0,11100	0,01959	32,10806	0,073	0,1494	0,000000
Scale			1,00000	0,00000		1,000	1,0000	

Рис. 2. Параметри оцінювання *Logit*-моделі

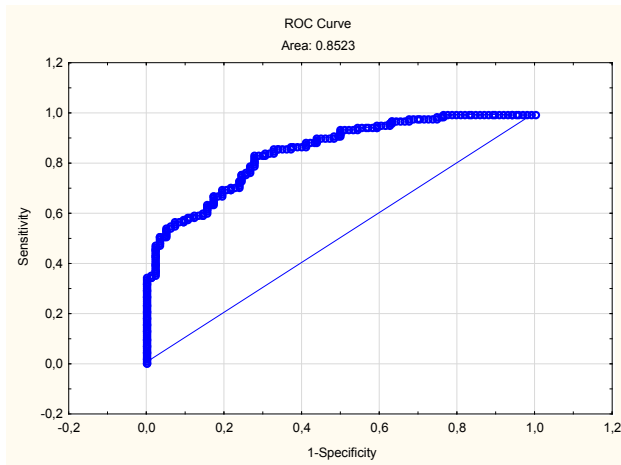


Рис. 3. ROC-крива оцінювання якості побудованої *Logit*-моделі

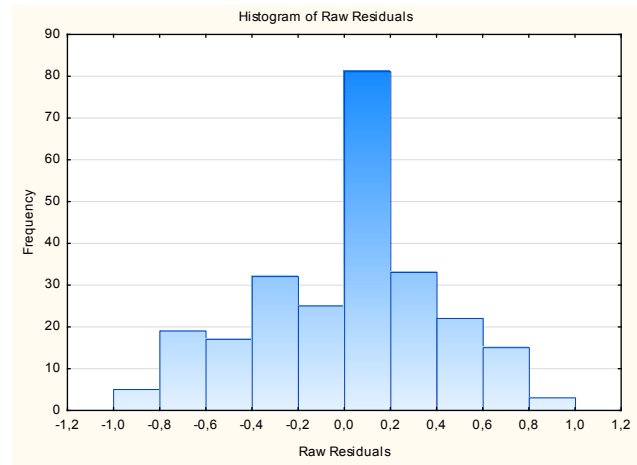


Рис. 4. Гістограма розподілу залишків *Logit*-моделі

Площа під *ROC*-кривою визначає якість у задачах бінарної класифікації. Отже, кількісну оцінку якості моделі отримуємо за допомогою показника *AUC* (*Area under ROC curve*), що дорівнює 0,8523 (рис. 3). Це вказує на доволі високу точність отриманого результату.

Окрім *ROC*-кривої, також дослідимо розподіл залишків моделі. Із гістограми (рис. 4) доходимо висновку, що залишки *Logit*-моделі нормально розподілені. На це також вказує нормальний імовірнісний графік розподілу залишків моделі (рис. 5).

Отже, доходимо висновку, що побудована логістична регресійна ймовірнісна модель є значущою.

Ще однією моделлю бінарного вибору є *Probit*-модель. Застосуємо ці ж параметри для її побудови, що і для *Logit*-моделі. Її відмінність від логістичної моделі полягає лише в тому, що для розрахунку ймовірності використовується нормальна функція розподілу (функція Гаусса).

Застосувавши модуль *Generalized Linear / Nonlinear* програмного продукту *STATISTICA 10*, одержимо *Probit*-модель для оцінювання ймовірності додатного

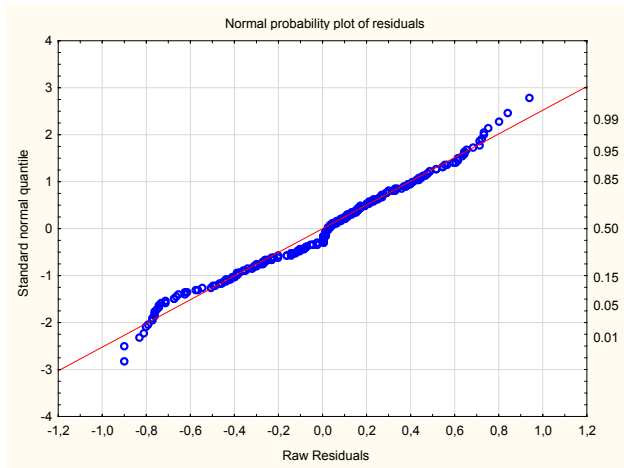


Рис. 5. Нормальний імовірнісний графік залишків Logit-моделі

сальдо товарів і послуг країн ЄС. Результати моделювання зображено на рис. 6.

На основі результатів моделювання доходимо висновку, що на показник додатного сальдо торгівлі впливають ті ж фактори, що і під час моделювання логістичної регресії.

Користуючись параметрами оцінювання Probit-моделі (рис. 7) та врахувавши, що для розрахунку ймовірності додатного сальдо торговельного балансу використовується нормальна функція розподілу, одержимо:

$$Y = \frac{1}{\sqrt{2\pi}} e^{-\frac{1}{2}Z^2},$$

де Z – лінійна залежність між незалежними факторами (рис. 7):

$$Z = -6 + 0,66X_4 + 0,00026X_5 - 0,00021X_6 - 0,00024X_7 + 0,66X_8.$$

Нехтуємо фактором X₈, ураховуючи, що коефіцієнт надзвичайно малий.

Побудуємо ROC-криву для аналізу якості Probit-моделі (рис. 8). Бачимо, що AUC=0,8519 (рис. 8), тобто точність результату моделювання, доволі висока.

Це також можна оцінити за допомогою гістограми розподілу залишків моделі (рис. 9) та нормального ймовірнісного графіка залишків (рис. 10).

Отже, за результатами проведеного дослідження, застосувавши лінійні логістичні моделі *logit* і *probit*, можна з'ясувати, що перевага обсягів експорту над імпортом країни залежить від таких факторів, як імміграція населення, ВВП на душу населення, середній дохід населення, чистий прибуток населення, заробітна плата населення та використання населенням Інтернет-ресурсів.

Висновки з проведеного дослідження. На основі аналізу доходимо висновку, що на показник додатного сальдо значний вплив має імміграція населення. За результатами попередніх досліджень було виявлено, що на обсяги експорту та імпорту країни впливають і чисельність населення, і міграційні процеси у країні [10]. Проте бачимо, що на показник додатного сальдо більше впливатиме імміграційна ситуація у країні. Тобто якщо країна позиціонує

Effect	Saldo, Y - Test of all effects (Spreadsheet8) Distribution : BINOMIAL, Link function: PROBIT Modeled probability that Saldo, Y = yes		
	Degr. of Freedom	Wald Stat.	p
Intercept	1	14,22628	0,000162
Population, X1	1		
Total employment national concept, X2	1	0,06345	0,801118
Emigration, X3	1	0,14137	0,706921
Immigration, X4	1	7,37517	0,006613
GDP per capita, X5	1	15,36490	0,000089
Median equivalised net income, X6	1	7,07315	0,007825
Net earnings, X7	1	6,39671	0,011433
Wages and salaries, X8	1	5,57765	0,018191
Internet use, X9	1	36,93227	0,000000

Рис. 6. Результати побудованої моделі Probit regression

Effect	Saldo, Y - Parameter estimates (Spreadsheet8) Distribution : BINOMIAL, Link function: PROBIT Modeled probability that Saldo, Y = yes							
	Level of Effect	Column	Estimate	Standard Error	Wald Stat.	Lower CL 95, %	Upper CL 95, %	p
Intercept		1	-6,00616	1,59240	14,22628	-9,1272	-2,88512	0,000162
Population, X1		2	0,00000	0,00000	3,14952	-0,0000	0,00000	0,075950
Total employment national concept, X2		3	8,07680	32,06346	0,06345	-54,7664	70,92002	0,801118
Emigration, X3		4	0,08395	0,22329	0,14137	-0,3537	0,52159	0,706921
Immigration, X4		5	0,66177	0,24368	7,37517	0,1842	1,13938	0,006613
GDP per capita, X5		6	0,00026	0,00007	15,36490	0,0001	0,00039	0,000089
Median equivalised net income, X6		7	-0,00021	0,00008	7,07315	-0,0004	-0,00006	0,007825
Net earnings, X7		8	-0,00024	0,00010	6,39671	-0,0004	-0,00005	0,011433
Wages and salaries, X8		9	-0,00000	0,00000	5,57765	-0,0000	-0,00000	0,018191
Internet use, X9		10	0,06666	0,01097	36,93227	0,0452	0,08816	0,000000
Scale			1,00000	0,00000		1,0000	1,00000	

Рис. 7. Параметри оцінювання Probit regression

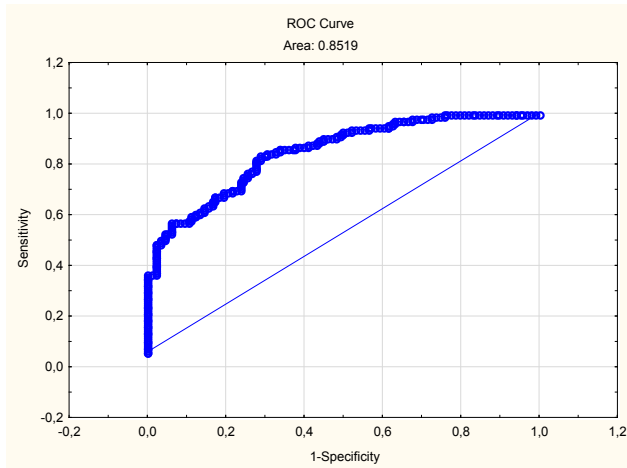


Рис. 8. ROC-крива оцінювання якості побудованої *Probit*-моделі

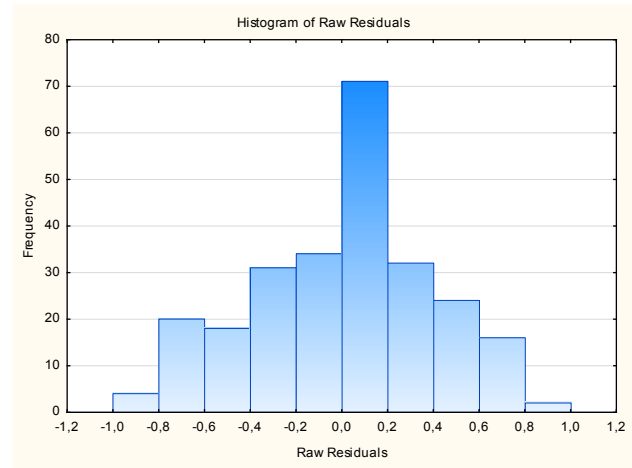


Рис. 9. Гістограма розподілу залишків *Probit*-моделі

себе як країна-експортер, збільшення кількості імігрантів сприятиме переважанню експорту над імпортом.

Також результати моделювання підтвердили гіпотезу про вплив купівельної спроможності населення на показник додатного сальдо країни. Так, такі показники, як ВВП на душу населення, середній дохід, чистий прибуток та заробітна плата населення, є значущими чинниками впливу на основні показники торгівлі. Крім того, що значна частина торговельних угод укладається на рівні урядів країн, населення теж значно впливає на зміну обсягів торговельних операцій. Так, чим більше матеріальне забезпечення та купівельна спроможність населення, тим більшими будуть показники зовнішньої торгівлі країни.

Перевагою застосування моделей бінарного вибору до оцінювання міжнародної торгівлі є те, що вони дають змогу дослідити, чи відбудеться подія для конкретного випробування. Тобто під час оцінювання торгівлі маємо можливість оперувати не лише факторами, що задані у кількісному вигляді, а й факторами, які визначені бінарними величинами.

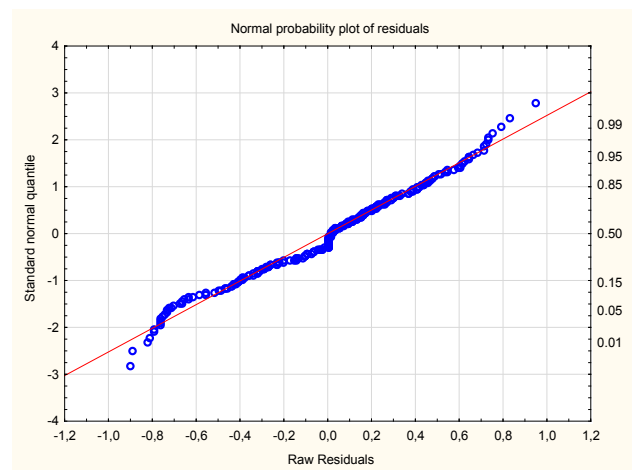


Рис. 10. Нормальний імовірнісний графік залишків *Probit*-моделі

1. Лук'яненко І.Г. Оцінка ймовірності настання кризових явищ у фінансовому секторі України. *Бізнес Інформ*. 2011. № 5(2). С. 51–54. URL: https://www.business-inform.net/pdf/2011/5_2/51_54.pdf (дата звернення: 18.12.2019).

2. Малютін О.К. Імовірнісна індикативна модель виявлення кризових явищ інвестиційного розвитку національної економіки. *Економіка та держава*. 2012. № 11. С. 31–34. URL: http://www.economy.in.ua/pdf/11_2012/10.pdf (дата звернення: 18.12.2019).

3. Міщук Н. Використання економетричних моделей для оцінки стійкості банків в Україні. *Вісник Львівського університету. Серія «Економіка»*. 2009. Вип. 42. С. 10–22. URL: <http://publications.lnu.edu.ua/bulletins/index.php/economics/article/viewFile/7031/7039> (дата звернення: 18.12.2019).

4. Сушко В.І., Павлюк Т.С. Класифікація моделей оцінки ймовірності банкрутства підприємств. *Економіка: теорія та практика*. 2014. № 1. С. 72–83.

5. Швабу Ю.І. Модель податкового планування з метою активізації економічних процесів в Україні. *Агросвіт*. 2011. № 20. С. 58–62. URL: http://www.agrosvit.info/pdf/20_2011/13.pdf (дата звернення: 18.12.2019).

6. Логистическая регрессия / Центр статистического анализа. URL: <https://www.statmethods.ru/statistics-metody/logisticheskaya-regressiya/> (дата звернення: 25.10.2019).

7. Пробит-модель регрессии / Центр статистического анализа. URL: <https://www.statmethods.ru/statistics-metody/probit-model-regressii/> (дата звернення: 25.10.2019).

8. Eurostat. URL: <http://www.ec.europa.eu/> (дата звернення: 20.12.2019).

9. Дзюбановська Н.В. Прагматизм оцінювання міжнародної торгівлі країн: методи і моделі : монографія. Тернопіль : ТНЕУ, 2019. 298 с.

10. Dziubanovska N. Multifactor models for studying the EU countries' international trade. *Economic Annals-XXI*. 2019. Vol. 175(1–2). P. 29–34. DOI: <https://doi.org/10.21003/ea.V175-05> (дата звернення: 06.12.2019).

1. Lukianenko I. H. (2011) Otsinka ymovirnosti nastannia kryzovykh yavlyshch u finansovomu sektori Ukrainy [Assessment of the probability of crisis occurrence in the financial sector of Ukraine]. *Biznes Inform.* 5(2), 51–54. Retrieved from: https://www.business-inform.net/pdf/2011/5_2/51_54.pdf (accessed 18 December 2019). (in Ukrainian)
2. Maliutin O. K. (2012) Imovirnisna indykatyvna model vyivlennia kryzovykh yavlyshch investytsiinoho rozvytku natsionalnoi ekonomiky [A probabilistic indicative model for the identification of crisis phenomena of investment development of the national economy]. *Ekonomika ta derzhava*, 11, 31–34. Retrieved from: http://www.economy.in.ua/pdf/11_2012/10.pdf (accessed 18 December 2019). (in Ukrainian)
3. Mishchuk N. (2009) Vykorystannia ekonometrychnykh modelei dlia otsinky stiikosti bankiv v Ukraini [The application of econometric models to assess the resilience of banks in Ukraine]. *Visnyk Lvivskoho universytetu. Seriya ekonomika*, 42, 10–22. Retrieved from: <http://publications.lnu.edu.ua/bulletins/index.php/economics/article/viewFile/7031/7039> (accessed 18 December 2019). (in Ukrainian)
4. Sushko V. I., Pavliuk T. S. (2014) Klasyfikatsiia modelei otsinky ymovirnosti bankrutstva pidpriemstv [Classification of bankruptcy probability estimation models of enterprises]. *Ekonomika: teoriia ta praktyka*, 1, 72–83. (in Ukrainian)
5. Shvabu Yu. I. (2011) Model podatkovoho planuvannia z metoiu aktyvizatsii ekonomichnykh protsesiv v Ukraini [Model of tax planning for the purpose of activation of economic processes in Ukraine]. *Ahrosvit*, 20, 58–62. Retrieved from: http://www.agrosvit.info/pdf/20_2011/13.pdf (accessed 18 December 2019). (in Ukrainian)
6. Lohystycheskaia rehressyia [Logistic Regression]. *Tsentr statystycheskoho analyza*. Retrieved from: <https://www.statmethods.ru/statistics-metody/logisticheskaya-regressiya/> (accessed 25 October 2019). (in Russian)
7. Probit-model rehressyy [Probit model of regression]. *Tsentr statystycheskoho analyza*. Retrieved from: <https://www.statmethods.ru/statistics-metody/probit-model-regressii/> (accessed 25 October 2019). (in Russian)
8. Eurostat. Retrieved from: <http://www.ec.europa.eu/> (accessed 20 December 2019).
9. Dziubanovska N. V. (2019) Prahmatyzm otsiniuvannia mizhnarodnoi torhivli krain: metody i modeli [The pragmatism of assessment countries' international trade: methods and models]. Ternopil : TNEU. (in Ukrainian)
10. Dziubanovska N. Multifactor models for studying the EU countries' international trade. *Economic Annals-XXI*. 2019. Vol. 175(1-2). P. 29–34. DOI: <https://doi.org/10.21003/ea.V175-05> (accessed 06 December 2019).

E-mail: dzyubanovska.n@gmail.com