

DOI: <https://doi.org/10.32782/2524-0072/2023-50-49>

УДК 657.3

ПРОГНОЗУВАННЯ ВЕЛИЧИНИ ДОХОДУ ВІД РЕАЛІЗАЦІЇ ПОСЛУГ ПІДПРИЄМСТВ ТРАНСПОРТНОЇ ГАЛУЗІ

FORECASTING THE AMOUNT OF INCOME FROM THE SALE OF SERVICES OF TRANSPORT COMPANIES

Прокопович Леонід Борисович

кандидат економічних наук, доцент,

Національний університет кораблебудування імені адмірала Макарова

ORCID: <https://orcid.org/0000-0003-2561-8862>

Prokopovich Leonid

Admiral Makarov National University of Shipbuilding

У статті досліджується проблема підвищення достовірності величини чистого доходу від реалізації послуг підприємств транспортної галузі. В роботі проаналізовано характер існуючої залежності рівня чистого доходу від реалізації групи факторів. Серед факторів розглядалися ϵ : операційні витрати, операційні доходи, інші витрати та інші доходи. Запропоновані регресійні моделі побудовані за допомогою метода найменших квадратів. В якості альтернативи до моделей побудованих за допомогою метода найменших квадратів були побудовані моделі за допомогою методів: k -ближніх сусідів та бінарного дерева рішень. В процесі моделювання були обрані найліпші моделі в кожній групі: лінійна модель за методом найменших квадратів на основі початкових даних, лінійна модель за методом найменших квадратів на основі головних компонент, двофакторна модель на основі бінарного дерева рішень. За результатами проведеного дослідження визначені переваги отриманої регресійної моделі на основі машинного навчання (бінарного дерева рішень) для прогнозування величини чистого доходу від реалізації послуг підприємств транспортної галузі.

Ключові слова: доходи, витрати, регресія, методи машинного навчання, дерево рішення, k -ближніх сусідів.

The article examines the problem of increasing the reliability of the value of net income from the sale of services of enterprises in the transport industry. The paper analyzes the nature of the existing dependence of the level of net income on the implementation of a group of factors. The factors considered are: operating expenses, operating income, other expenses and other income. It turned out that operating expenses and income have the greatest impact on the dependent variable. They were used as factors in the subsequent stages of model building. The proposed regression models are built using the least squares method. To build some of the models, the primary data were standardized and transformed using the principal components method. To build some models, we used the transformation of primary data using the decimal logarithm. The significance of the internal parameters of the models obtained by the least squares method was tested by the value of their p -values. As an alternative to the models built using the least squares method, models were built using machine learning methods. Multivariate regression models were calculated using the following methods: k -nearest neighbors and binary decision tree. In the process of modeling, the best models in each group were selected: a linear least-squares model based on the initial data, a linear least-squares model based on the principal components, and a two-factor model based on a binary decision tree. The qualitative characteristics of the obtained regression models (coefficient of determination, Fisher's criterion, average approximation error) were compared. As a result, the best model was obtained, based on the value of the average approximation error, i.e., a multivariate regression model built using the decision tree method. The multivariate regression model based on a binary decision tree has a very small approximation error and exceeds the previously obtained models in all the parameters studied. Based on the results of the study, the advantages of the obtained regression model based on machine learning (binary decision tree) for predicting the amount of net income from the sale of services of transport enterprises are determined.

Keywords: income, expenses, regression, machine learning methods, decision tree, k -nearest neighbors.

Постановка проблеми. Одним із важливих фінансових показників ефективності діяльності будь-якого підприємства є дохід від реалізації продукції (робіт, послуг). Кризові тренди економіки країни ускладнюють процес управління діяльністю підприємства, погіршуючи прогнозну точність фінансових показників, що були основою для прийняття управлінських рішень. Вказане обумовлює необхідність: дослідження факторів, які впливають на величину та поведінку доходів; пошук сучасних підходів до прогнозування рівня даного показника фінансової діяльності підприємства.

Аналіз останніх досліджень і публікацій.

Питанням вивчення фінансових результатів підприємств присвячено роботи багатьох фахівців, серед яких можна вказати на роботи: Волошина О. А., Кордзаія І. А., Даценко С. М., Ульченко А. М. [2], Єпіфанова І. Ю., Гумєнюк В. С. [5], Мартиненко В. П., Кушик І. Л. [6], Поддєрьогін А. М., Скочій С. В. [7]. Безпосередньо проблеми прогнозування та оцінки величини доходів підприємств досліджували: Артюх-Пасюта О. В., Мілька А. І. [1], Гайбура Ю. А., Загнітко Л. А. [3], Драченко А. І., Юрчишена Л. В. [4]. Проте, проблематика аналізу та прогнозування величини доходів підприємств в умовах погіршення кризових явищ в економіці країни не втрачають актуальності.

Метою дослідження є підвищення достовірності оцінювання рівня чистого доходу від реалізації послуг підприємств транспортної галузі.

Вклад основного матеріалу дослідження. В якості первинних даних була використана інформація люб'язно викладена у публікації Драченко А. І., Юрчишена Л. В. [4]. Проте, перш чим розкрити результати проведеного дослідження, необхідно вказати на деякі проблемні моменти згаданого джерела. Справа у тому, що рівняння регресії (див. формулу 1 [4]) викликає декілька зауважень. По-перше, згідно табл. 3 [4] обрані для побудови регресійної моделі фактори мають

сильний кореляційний зв'язок, а значить використання методу найменших квадратів, без усунення мультиколінеарності невірне. По-друге, величина p -значення для першого внутрішнього параметру моделі (« Y -перетин», див. табл. 4 [4]) вище граничної величини (0,05), а значить даний параметр в моделі є несуттєвим. Вказане, свідчить про проблемність знайденого регресійного рівняння та стало однією з причин проведення нового дослідження.

В якості залежної змінної (Y) досліджується чистий дохід від реалізації, млрд. грн. Фактори, які впливають на величину залежної змінної: X_1 – операційні доходи, млрд. грн.; X_2 – інші доходи, млрд. грн.; X_3 – операційні витрати, млрд. грн.; X_4 – інші витрати, млрд. грн.

Перш чим приступати до відбору факторів для побудови регресійної моделі первинні дані були перевірені на викиди за допомогою побудови діаграм типу «ящик з вусами». Інформація рис. 1 свідчить про відсутність викидів у початкових даних.

Розрахунок кореляційної матриці, див. табл. 1, каже, що фактори X_1 та X_3 мають дуже високу силу впливу на величину залежної змінної, так як величина коефіцієнту кореляції Пірсона вище 0,9. Фактор X_4 має слабкий зв'язок з чистим доходом від реалізації, а фактор X_2 – помірну силу взаємозв'язку (за шкалою Чеддока). Проте, незважаючи на дуже велику силу взаємозв'язку факторів X_1 та X_3 із залежною змінною, вказані фактори також мають дуже високий взаємозв'язок (коефіцієнт кореляції дорівнює 0,95), що вказує на наявність мультиколінеарності. З метою, перевірки на допустимість використання обох факторів для побудови двофакторної регресійної моделі, за допомогою способу найменших квадратів, був зроблений розрахунок зворотної матриці від матриці парних кореляцій факторів, див. табл. 2.

Значення на головній діагоналі зворотної матриці (11,03) перевищують 4,0, що підтверджує критично високий рівень мультиколіне-

Таблиця 1

Кореляційна матриця

	Y	X1	X2	X3	X4
Y	1	0,991639	0,352894	0,972344	0,124088
X1	0,991639	1	0,292046	0,953577	0,036535
X2	0,352894	0,292046	1	0,217923	0,912166
X3	0,972344	0,953577	0,217923	1	0,061733
X4	0,124088	0,036535	0,912166	0,061733	1

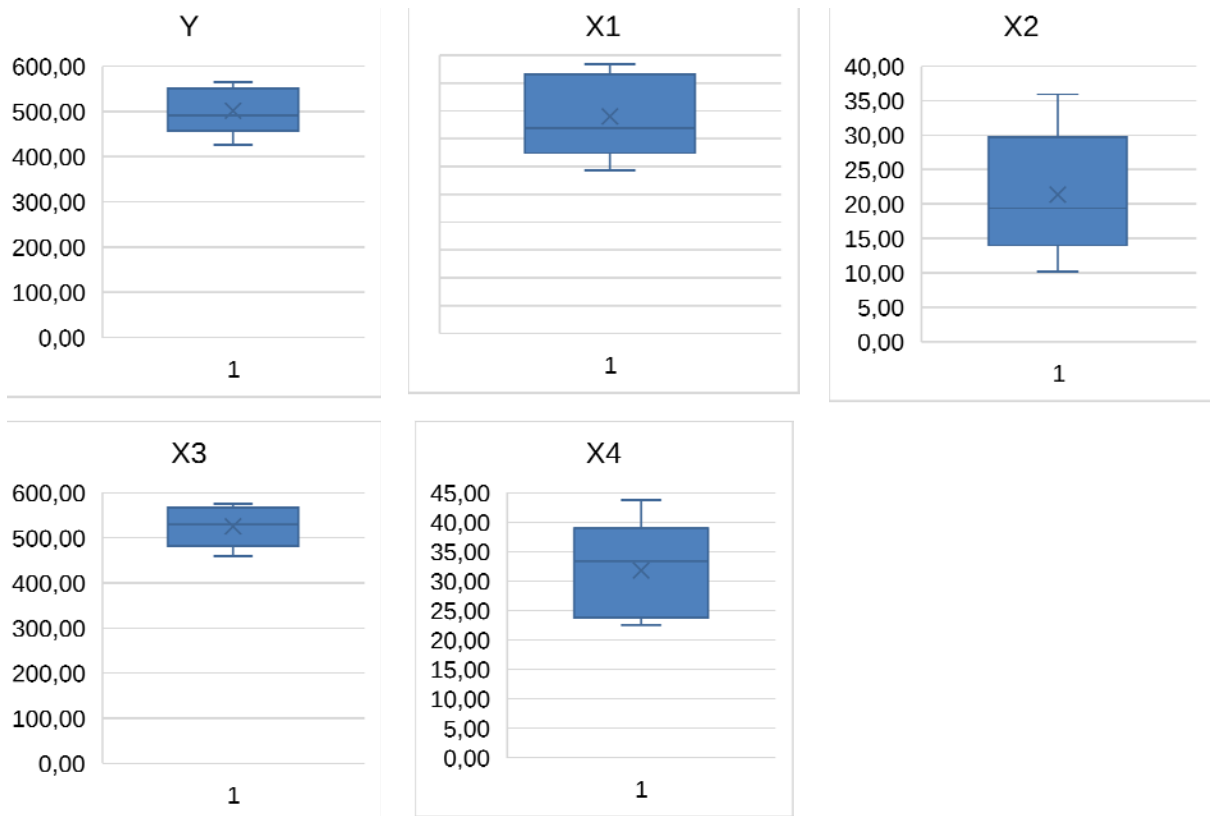


Рис. 1. Перевірка даних на викиди

арності між факторами X1 та X3. У зв'язку з чим, при використанні для побудови регресійної моделі способу найменших квадратів можна діяти у двох напрямках. По-перше, побудувати однофакторні моделі. По-друге, спробувати уникнути мультиколінеарності, наприклад, за допомогою методу головних компонент. Обидва напрямки були досліджені.

Таблиця 2

Зворотна матриця, R⁻¹

	X1	X3
X1	11,0265	-10,5146
X3	-10,5146	11,0265

В межах першого напрямку, були побудовані однофакторні регресійні моделі, за допомогою способу найменших квадратів, за рівняннями 1-8. Для побудови моделей за рівняннями 1 та 8 було використано перетворення первинних даних за допомогою десятичного логарифму.

$$\hat{y} = a_0 + a_1x + \varepsilon, \tag{1}$$

$$\hat{y} = a_1x + \varepsilon, \tag{2}$$

$$\hat{y} = a_0 + a_1x + a_2x^2 + \varepsilon, \tag{3}$$

$$\hat{y} = a_1x + a_2x^2 + \varepsilon, \tag{4}$$

$$\hat{y} = a_0 + a_1 \ln x + \varepsilon, \tag{5}$$

$$\hat{y} = a_1 \ln x + \varepsilon, \tag{6}$$

$$\hat{y} = a_0x^{a_1} + \varepsilon, \tag{7}$$

$$\hat{y} = x^{a_1} + \varepsilon, \tag{8}$$

Результати розрахунків внутрішніх параметрів (a₀, a₁, a₂) отриманих моделей наведені у табл. 3. З метою перевірки внутрішніх параметрів моделі на суттєвість були досліджені розраховані величини р-значень (див. табл. 4).

За інформацією табл. 4 видно, що з 16 моделей 5 не пройшли перевірку на суттєвість параметрів (модель за рівнянням³ для фактору X1 та моделі 1, 3, 4, 7 за фактором X3) та були усунуті від подальшого дослідження.

Для порівняння моделей була сформована порівняльна таблиця (див. табл. 5).

Інформація величин коефіцієнту детермінації (R²) у табл. 5 показує, що одна модель добра (для моделі за рівнянням 5 для фактору X3 R² = 0,94, що менше 0,95), а всі інші моделі – мають високу точність апроксимації (R² > 0,95). Розрахункова величина критерія Фішера (F) перевищує його табличне значення для всіх моделей, що підтверджує їх

Таблиця 3

Параметри моделі (X1, X3)

Номер рівняння	Внутрішні параметри		
	a_0	a_1	a_2
<i>для фактору X1</i>			
1	231,1422374	6,931224683	-
2	-	12,6824549	-
3	30,02467568	17,52959047	-0,135394334
4	-	19,09390731	-0,155173908
5	-473,6678012	267,3069293	-
6	-	137,7699419	-
7	69,1640414	0,541786184	-
8	-	1,700363983	-
<i>для фактору X3</i>			
1	-89,50741607	1,125052193	-
2	-	0,955692357	-
3	322,1193958	-0,476506593	0,001547623
4	-	0,77371238	0,000342329
5	-3120,043315	578,4426127	-
6	-	80,17653656	-
7	0,31129213	1,178754374	-
8	-	0,992382535	-

Таблиця 4

p-значення (X1, X3)

Номер рівняння	p-значення для параметру параметри		
	a_0	a_1	a_2
<i>для фактору X1</i>			
1	0,001521275	0,000916614	-
2	-	1,53149E-05	-
3	0,782451595	0,072036757	0,166380721
4	-	1,08646E-05	0,00027436
5	0,001472829	0,000171801	-
6	-	1,72014E-06	-
7	5,10064E-05	0,000495371	-
8	-	4,55997E-07	-
<i>для фактору X3</i>			
1	0,355874044	0,005497966	-
2	-	1,73491E-07	-
3	0,84217183	0,939233286	0,799425746
4	-	0,017756827	0,344625276
5	0,009304728	0,006094258	-
6	-	1,72086E-05	-
7	0,332830652	0,005339536	-
8	-	1,31462E-10	-

Таблиця 5

Порівняння моделей

Номер рівняння	Коефіцієнт детермінації, R ²	Критерій Фішера, F		Середня помилка апроксимації, А у %
		розрахунковий	табличний	
<i>для фактору X1</i>				
1	0,983347639	177,155	10,128	1,16
2	0,993616211	622,587	7,709	7,83
4	0,999952344	31474,170	9,552	0,57
5	0,994533842	545,832	10,128	0,61
6	0,997859027	1864,309	7,709	3,89
7	0,988938945	268,222	10,128	0,90
8	0,998897482	3624,059	7,709	18,53
<i>для фактору X3</i>				
2	0,999319899	5877,482	7,709	2,50
5	0,941624204	48,391	10,128	1,67
6	0,993233464	587,144	7,709	7,13
8	0,999981277	213633,0344	7,709	2,58

статистичну значимість. Для перевірки якості моделі та відбору найліпшої з побудованих моделей був здійснений розрахунок середньої помилки апроксимації (А). Допустимі значення вказаного показника для моделей не повинні перевищувати 10%. За даними табл. 5 видно, що середня помилка апроксимації за рівнянням 8 для фактору X1 перевищує граничне значення, тому дана модель була усунена від подальшого порівняння. Всі інші моделі мають добру точність апрок-

симації за даним показником, а найліпшою виявилася модель для фактору X1 за рівнянням 4. Якщо підставити значення внутрішніх параметрів найліпша модель прийме наступний вигляд:

$$\hat{y} = (1,08646E - 05)x_1 + 0,00027436x_1^2 + \varepsilon, \quad (9)$$

В межах другого напрямку, перш чим застосувати до первинних даних метод головних компонент, значення факторів були стандартизовані за допомогою способу MinMax. Після

Таблиця 6

Параметри моделі (PC1)

Номер рівняння	Внутрішні параметри		
	a ₀	a ₁	a ₂
1	501,46	-95,32769221	-
2	-	-95,32769221	-
3	503,1927863	-94,77610933	-7,033137349
4	-	-184,6690549	1139,176181

Таблиця 7

p-значення (PC1)

Номер рівняння	p-значення для параметру параметри		
	a ₀	a ₁	a ₂
1	5,14224E-07	0,000602915	-
2	-	0,85951064	-
3	0,000118041	0,006060807	0,713406272
4	-	0,668529114	0,146764713

застосування методу головних компонент виявилось, що значення дисперсії, які пояснюються головними компонентами, наступні: 97.7%, 2.3%. Приймаючи до уваги вказані значення, було вирішено побудувати однофакторні регресійні моделі для першої головної компоненти (PC1) за рівняннями 1–4. Розраховані значення внутрішніх параметрів моделей наведено у табл. 6., а їх *p*-значення у табл. 7.

За результатами табл. 7 видно, що у моделях за рівняннями 2–4 є несуттєві внутрішні параметри, тому дані моделі були відкинуті. Залишилася модель за рівнянням 1, яка із знайденими параметрами прийме наступний вигляд:

$$\hat{y} = 501,46 - 95,32769221PC_1 + \varepsilon, \quad (10)$$

Дана модель (див. рівняння 10) має високу точність апроксимації (коефіцієнт детермінації дорівнює 0,987394945) та статистичну значимість ($F: 235,000 > 10,128$). Якість даної моделі підтверджує також розрахунок середньої помилки апроксимації, величина якої дорівнює 0,87%. Проте, даний показник, трохи гірший чим у моделі за рівнянням 9 ($A = 0,87\%$), тому найліпшою моделлю побудованою за допомогою способу найменших квадратів виявилася модель 9.

На наступному етапі було прийнято рішення спробувати побудувати регресійні моделі за допомогою методів машинного навчання з вчителем: *k*-ближніх сусідів та регресійного бінарного дерева рішень. Позитивним моментом вказаних методів є незалежність їх від впливу мультиколінеарності між факторами X_1 та X_3 . Для реалізації розрахунків згідно вказаних методів був розроблений скрипт на мові Python та бібліотека Scikit-learn. У обох методів є певна кількість гіперпараметрів. При реалізації моделі *k*-ближніх сусідів в якості гіперпараметру був обраний *n_neighbors* – кількість сусідніх елементів зна-

чення яких буде визначати прогнозну величину залежної змінної. Під час розрахунку моделі за методом дерева рішення в якості гіперпараметру була використана глибина дерева – *max_depth*. Порівняльні показники моделей отриманих за допомогою машинного навчання наведені у табл. 8.

Якщо вивчити отримані величини коефіцієнту детермінації для моделей за методом *k*-ближніх сусідів, то для моделі де гіперпараметр *n_neighbors* дорівнює 2 модель добра ($R^2 > 0,8$), а при *n_neighbors* = 3 модель прийнятна ($0,5 < R^2 < 0,8$). Але, можна побачити, що вказані моделі не є статистично значимими, так як розраховані значення критерія Фішера менше чим його табличне значення. Тому, обидві моделі *k*-ближніх сусідів були усуненні від подальшого дослідження.

Модель на основі дерева рішення (див. рис. 2) має високу точність апроксимації ($R^2 = 0,999$) та є статистично значимим ($983,149 > 9,552$).

Середня помилка апроксимації даної моделі складає 0,196%, що є краще, чим у найліпшої моделі розрахованої за допомогою метода найменших квадратів. Отриману регресійну модель можна формалізувати у наступному рівнянні:

$$\hat{y} = f(x_1, x_3, a_1, a_2, a_3, b_1, \dots, b_i) + \varepsilon, \quad (11)$$

де *a* – внутрішні параметри моделі (умови розбиття дерева у його вузлах які зображені на рис. 2);

b – зовнішні параметри (гіперпараметри) моделі, $i \in [1; 11]$.

Підсумовуючи вказане, можна рекомендувати для прогнозування величини чистого доходу від реалізації робіт та послуг підприємства транспортної галузі використовувати регресійну модель на основі бінарного дерева рішень (див. рівняння 11).

Висновок. В роботі запропоновані регресійні моделі для підвищення достовірності

Таблиця 8

Порівняння моделей на основі машинного навчання

Різновид моделі	Коефіцієнт детермінації, R^2	Критерій Фішера, F		Середня помилка апроксимації, A у %
		розрахунковий	табличний	
<i>k</i> -ближніх сусідів (<i>n_neighbors</i> = 2)	0,874	6,946	9,552	2,677
<i>k</i> -ближніх сусідів (<i>n_neighbors</i> = 3)	0,652	1,869		5,253
дерево рішення (<i>max_depth</i> = 2)	0,999	983,149		0,196

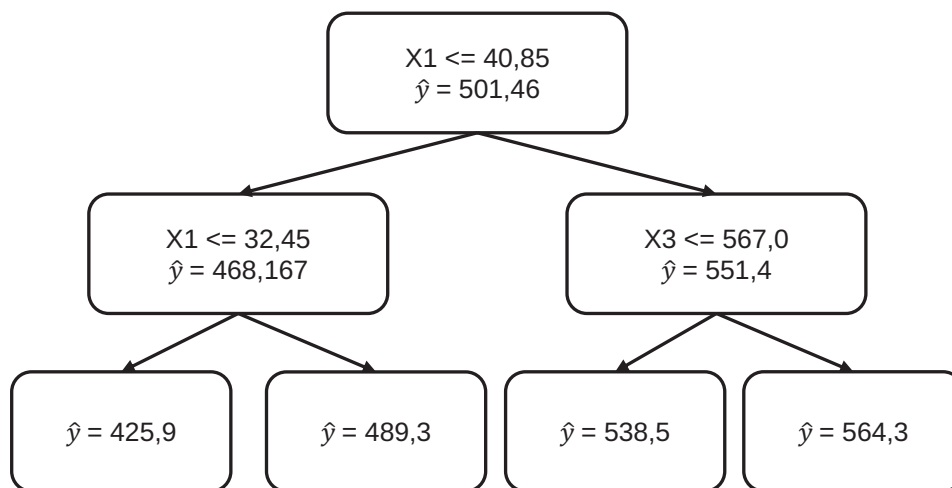


Рис. 2. Регресійне дерево рішення (max_depth = 2)

оцінювання рівня чистого доходу від реалізації робіт та послуг підприємств транспортної галузі. Були розроблені однофакторні та багатоваріантні регресійні моделі за допомогою методів: найменших квадратів, k-ближніх сусідів та регресійного бінарного дерева рішень. В якості факторів, що впливають на залежну змінну в побудованих моделях було використано операційні доходи, млрд. грн., операційні витрати, млрд. грн. За результатом порівняння показників отриманих моделей (коефіцієнту

детермінації, критерія Фішера, середньої помилки апроксимації) була обрана ліпша регресійна модель, якою виявилася модель на основі дерева рішень.

Перспективи подальших досліджень полягають у можливості підвищити якість оцінки величини чистого доходу від реалізації послуг підприємств транспортної галузі України за рахунок застосування інших методів машинного навчання (наприклад, за допомогою нейронних мереж або XGBoost).

СПИСОК ВИКОРИСТАНИХ ДЖЕРЕЛ:

1. Артюх-Пасюта О. В., Мілька А.І. Оцінка достовірності інформації у фінансовій звітності підприємства на основі моделей Дж. Монтьє, М. Д. Беніша та М. Л. Роксас. *Економіка та суспільство*. 2022. № 40. URL: <https://economyandsociety.in.ua/index.php/journal/article/view/1525/1467>.
2. Волошина О. А., Кордзаія І. А., Даценко С. М., Ульченко А. М. Методичні підходи до аналізу доходів підприємства та факторів їх формування. *Ефективна економіка*. 2019. № 6. URL: http://www.economy.nayka.com.ua/pdf/6_2019/49.pdf.
3. Гайбура Ю. А., Загнітко Л. А. Прибутковість підприємства: поняття, фактори, резерви. *Науковий вісник Міжнародного гуманітарного університету*. 2017. № 24-1. С. 99–104. URL: <http://vestnik-econom.mgu.od.ua/journal/2017/24-2-2017/23.pdf>.
4. Драченко А. І., Юрчишена Л. В. Концептуальний підхід до моделювання впливу фінансових показників на прибуток підприємства. *Економіка та суспільство*. 2022. № 44. URL: <https://economyandsociety.in.ua/index.php/journal/article/view/1816/1752>.
5. Єпіфанова І. Ю., Гуменюк В. С. Методика оцінювання прибутковості підприємства. *Економіка та суспільство*. 2017. № 12. URL: https://economyandsociety.in.ua/journals/12_ukr/42.pdf.
6. Мартиненко В. П., Кушик І. Л. Оцінка управління процесом формування прибутковості суб'єкта господарювання. *Інфраструктура ринку*. 2018. № 17. С. 163–168. URL: http://www.market-infr.od.ua/journals/2018/17_2018_ukr/29.pdf.
7. Поддєрьогін А. М., Скокій С. В. Факторний аналіз формування прибутку виробничого підприємства. *Збірник наукових праць Університету державної фіскальної служби України*. 2018. № 2. С. 304–318. URL: <https://core.ac.uk/download/pdf/268453122.pdf>.

REFERENCES:

1. Artyukh-Pasyuta O., Milka A. (2022) Otsinka dostovirnosti informatsii u finansovii zvitnosti pidpriemstva na osnovi modelei J. Montie, M. D. Benisha ta M. L. Roksas [Assessment of the reliability of information in the

company's financial statements based on models J. Monthier, M. D. Benisha and M. L. Roxas]. *Ekonomika i suspilstvo*, no. 40. Available at: <https://economyandsociety.in.ua/index.php/journal/article/view/1525/1467>. (in Ukrainian)

2. Voloshyna O. A., Kordzaia I. A., Datsenko S. M., Ulchenko A. M. (2019) Metodychni pidkhody do analizu dokhodiv pidpriemstva ta faktoriv yikh formuvannia [Methodical approaches to analysis of enterprises income and factors of their formation]. *Efektivna ekonomika*, no. 6. Available at: http://www.economy.nayka.com.ua/pdf/6_2019/49.pdf. (in Ukrainian)

3. Haibura Yu. A., Zahnitko L. A. (2017) Prybutkovist pidpriemstva: poniattia, faktory, rezervy [The enterprise profitability: the concept, factors, reserves]. *Naukovyi visnyk Mizhnarodnoho humanitarnoho universytetu*, no. 24-1, pp. 99–104. Available at: <http://vestnik-econom.mgu.od.ua/journal/2017/24-2-2017/23.pdf>. (in Ukrainian)

4. Drachenko A., Yurchyshena L. (2022) Kontseptualnyi pidkhid do modeliuvannia vplyvu finansovykh pokaznykiv na prybutok pidpriemstva [Conceptual approach to modeling the influence of financial indicators on enterprise profit]. *Ekonomika i suspilstvo*, no. 44. Available at: <https://economyandsociety.in.ua/index.php/journal/article/view/1816/1752>. (in Ukrainian)

5. Yepifanova I. Yu., Humeniuk V. S. (2017) Metodyka otsyniuvannia prybutkovosti pidpriemstva [Methodology of estimating the profitability of an enterprise]. *Ekonomika i suspilstvo*, no. 12. Available at: https://economyandsociety.in.ua/journals/12_ukr/42.pdf. (in Ukrainian)

6. Martynenko V. P., Kushyk I. L. (2018) Otsinka upravlinnia protsesom formuvannia prybutkovosti sub'iekta hospodariuvannia [Assessment of the management of the process for forming the benefit of the economic entity]. *Infrastruktura rynku*, no. 17, pp. 163–168. Available at: http://www.market-infr.od.ua/journals/2018/17_2018_ukr/29.pdf. (in Ukrainian)

7. Poddyerogin A. M., Skochii S. V. (2018) Faktornyi analiz formuvannia prybutku vyrobnychoho pidpriemstva [Factor analysis of production company's profit making]. *Zbirnyk naukovykh prats Universytetu derzhavnoi fiskalnoi sluzhby Ukrainy*, no. 2, pp. 304–318. Available at: <https://core.ac.uk/download/pdf/268453122.pdf>. (in Ukrainian)