

Т. В. Томашевська,

кандидат технічних наук, доцент,
доцент кафедри комп'ютерних наук та інформаційних систем,
Державний торговельно-економічний університет,
E-mail: t.tomashevaska@knute.edu.ua
Researcher ID: I-8901-2018,
ORCID: <http://orcid.org/0000-0001-5001-9226>;

Н. О. Гордійко,

кандидат технічних наук, доцент,
доцент кафедри прикладної фізики,
Національний технічний університету України
"Київський політехнічний інститут імені Ігоря Сікорського",
E-mail: n.hordiiko-ipt@ill.kpi.ua
ORCID: <https://orcid.org/0000-0002-0925-5160>

Побудова прогнозних моделей динаміки споживчої активності на основі методів машинного навчання та Big Data платіжних систем

У статті досліджено можливості використання великих даних платіжних систем для побудови прогнозних моделей динаміки споживчого попиту як ключового чинника формування роздрібного товарообороту. Актуальність дослідження зумовлена необхідністю отримання оперативних і достовірних індикаторів економічної активності в умовах високої невизначеності та нестабільності економічного середовища, зокрема внаслідок широкомасштабного вторгнення РФ в Україну у 2022 році.

Обґрунтовано доцільність використання великих даних як альтернативного джерела інформації, що дозволяє мінімізувати часові лаги, притаманні офіційній статистиці, та підвищити точність оцінювання поточного стану економіки. У процесі дослідження сформовано систему відносних показників, що відображають різні аспекти функціонування платіжної інфраструктури та поведінки споживачів, зокрема коефіцієнт активності платіжних карток, забезпеченість POS-терміналами, інтенсивність використання платіжної інфраструктури та рівень поширення безконтактних технологій. На основі зазначених індикаторів із застосуванням методу головних компонент побудовано інтегральний індекс споживчої активності, який дозволяє узагальнити багатомірну інформацію та зберегти більшу частину варіації вихідних даних. Встановлено, що перша головна компонента пояснює понад 80% загальної дисперсії. Отримані результати свідчать про високу інформативність побудованого індексу, який може бути інтерпретований як випереджальний індикатор змін у споживчому попиті.

Для оцінювання впливу інтегрального індексу на динаміку споживчого попиту побудовано економетричну модель на основі методу найменших квадратів із урахуванням лагових залежностей. Як результуючу змінну використано індекс фізичного обсягу роздрібного товарообороту, що розглядається як проксі-показник реалізованого споживчого попиту. Результати моделювання засвідчили статистичну значущість інтегрального індексу та високу пояснювальну здатність моделі, що підтверджує наявність стійкого взаємозв'язку між показниками платіжної активності та макроекономічною динамікою.

У роботі також здійснено порівняльний аналіз ефективності класичних економетричних методів і сучасних підходів машинного навчання, зокрема моделі часових рядів Prophet. Модель продемонструвала високу якість апроксимації на навчальній вибірці, проте виявилася неефективною для прогнозування в умовах різких економічних змін. Отримані результати показали, що застосування зазначених методів не забезпечує покращення якості прогнозування порівняно з базовою економетричною моделлю. Моделі машинного навчання демонструють низьку здатність до узагальнення на тестових вибірках, що пов'язано з обмеженим обсягом даних, а також наявністю структурних зламів у часових рядах.

Зроблено висновок, що ключовим фактором підвищення точності прогнозування є не стільки складність моделі, скільки якість та інформативність вхідних даних. Використання великих даних платіжних систем до-

© Стаття опублікована на умовах відкритого доступу за ліцензією CC BY-NC-ND license
(<https://creativecommons.org/licenses/by-nc-nd/4.0/>)

© This is an open access article distributed under the terms of CC BY-NC-ND license
(<https://creativecommons.org/licenses/by-nc-nd/4.0/>)

зволяє формувати випереджальні індикатори, які адекватніше відображають поточний стан економіки та забезпечують можливість оперативного реагування на зміни у споживчій поведінці.

Практична значущість дослідження полягає у можливості застосування запропонованого підходу для моніторингу та прогнозування споживчого попиту в режимі, близькому до реального часу, що є особливо важливим для прийняття управлінських рішень в умовах кризових явищ.

Ключові слова: великі дані, моделювання споживчого попиту, роздрібний товарооборот, інтегральний індекс споживчої активності, методи машинного навчання, модель Prophet.

Вступ. У сучасних умовах стрімкої цифровізації економічних процесів та зростання обсягів доступних даних особливої актуальності набуває проблема побудови ефективних моделей прогнозування споживчої активності. Традиційно для аналізу економічної динаміки використовуються класичні статистичні методи, однак їхнє застосування дедалі частіше виявляє низку обмежень у контексті роботи з високимірними, неоднорідними та швидкоплинними даними.

По-перше, класичні підходи, що базуються на припущеннях лінійності, стаціонарності та нормальності розподілів, не завжди здатні адекватно відобразити складну природу сучасних економічних процесів. Споживча активність формується під впливом великої кількості факторів – від макроекономічних умов до поведінкових характеристик індивідів, які часто мають нелінійний і стохастичний характер. У таких умовах традиційні моделі можуть втрачати точність та прогностичну здатність.

По-друге, суттєвим недоліком класичних статистичних інструментів є їхня інерційність. Дані, що використовуються для аналізу, як правило, мають значний часовий лаг, а результати їхньої обробки відображають зміни у споживчій поведінці із запізненням у кілька місяців або навіть півроку. Це значно знижує практичну цінність даних для прийняття оперативних управлінських рішень, особливо в умовах високої невизначеності.

Натомість аналіз великих масивів транзакційних даних платіжних систем демонструє принципово інший рівень чутливості до змін економічної поведінки. Так, у ході досліджень було зафіксовано різкий обвал активності платіжних карток у Херсонській та Донецькій областях, що проявлявся, зокрема, у фактичній відсутності транзакцій через платіжні термінали. Важливо, що такі зміни фіксуються великими даними практично миттєво, тоді як офіційна статистика продовжує відображати застарілі тенденції протягом тривалого часу.

Крім того, аналіз просторової та структурної організації платіжної інфраструктури дає змогу виявляти приховані дисбаланси. Зокрема, спостерігається диспропорція між кількістю встановлених платіжних терміналів і фактичним обсягом транзакцій, що може бути індикатором неефективного використання інфраструктури або потенційно свідчити про наявність схем тіньового обігу коштів.

Ще одним важливим аналітичним показником є коефіцієнт активності платіжних карток (співвідно-

шення кількості активних карток та загальної кількості карток в обігу), який може розглядатись як проксі-індикатор рівня довіри населення до банківської системи в окремих регіонах. Зміни цього показника дозволяють оперативно відстежувати поведінкові реакції споживачів на кризові явища та макроекономічні шоки.

Зазначена проблематика набуває особливої актуальності в українській економіці, яка перебуває у стані перманентної турбулентності внаслідок повномасштабного військового вторгнення. За таких умов виникає нагальна потреба у формуванні індикаторів, здатних забезпечити своєчасну, релевантну та науково обґрунтовану оцінку поточного стану споживчої активності та її короткострокових перспектив. Одним із перспективних напрямів вирішення цієї проблеми є використання методів машинного навчання у поєднанні з аналізом великих даних, зокрема даних платіжних систем. Такі підходи дозволяють працювати з масивами транзакцій у режимі, наближеному до реального часу, виявляти приховані закономірності та формувати точніші та більш адаптивні прогностичні моделі.

Аналіз останніх досліджень та публікацій. Розвиток інформаційно-комунікаційних технологій призвів до суттєвого зростання кількості інформації, яку можна використовувати для вирішення нагальних економічних задач. Усвідомлення цього факту та можливості, які відкриває обробка й аналіз великих наборів даних для забезпечення прийняття науково обґрунтованих рішень, спричинили зростання уваги дослідників до вивчення ролі великих даних в економіці та дослідження способів використання результатів аналізу великих даних для стратегічного управління.

Загалом питаннями впливу цифрових технологій на розвиток соціально-економічних процесів займалася низка зарубіжних та українських учених, серед яких В. Лукато (W. Lucato) [1], Л. Айнав (L. Einav) та Дж. Левін (J. Levin) [2], С. Онешко і Л. Пащук [3], М. Дзямучич, Т. Шматковська, О. Борисюк [4]. Серед загальних питань цифрової трансформації окремо розглядається значення великих даних у цифровій трансформації економіки. У роботі [5] Л. В. Конг (L. W. Cong), Б. Лі (B. Li) та К. Т. Чжан (Q. T. Zhang) досліджують використання даних із нетрадиційних джерел для новітніх досліджень в економіці та суміжних галузях. Науковці показали позитивні результати при використанні текстового аналізу в корпоративних фінансах, інвестиціях та макроекономічних

прогнозах, аналізі цифрових слідів із соціальних мереж та мобільних пристроїв. Н. Сапотницька зі співавторами, Л. Птащенко та А. Добропас розглядають, як великі дані стають інструментом у прийнятті рішень, та оцінюють ефективність їхнього використання в економічних дослідженнях [6; 7].

Отже, і в українській, і в зарубіжній літературі парадигма великих даних розглядається як основа розвитку економіки, з одного боку, та як інструментарій для підтримки прийняття рішень у завданнях практичного управління бізнесом – з іншого.

Використання великих даних для конкретних задач економічного аналізу та моделювання розглядається в роботах [8; 9; 10]. Так, у роботі [8] проаналізовано анонімізовані транзакції за банківськими картами для оцінки економічних показників Іспанії та показано, що з допомогою таких наборів можна спрогнозувати основні статистичні показники (ВВП, рівень безробіття, тривалість життя тощо). Причому автори продемонстрували, що запропонований ними підхід дозволяє оцінити такі характеристики для досить дрібних територіальних об'єднань, для яких офіційна статистика відсутня. Застосування нейронних мереж для аналізу складної структури даних щодо банківських транзакцій продемонстровано у публікації [9]. Така структура передбачає використання методів машинного навчання, оскільки класичний статистичний аналіз не може забезпечити отримання релевантних результатів. А в роботі [10] досліджено вплив платіжних технологій на купівельну поведінку користувачів і продемонстровано залежність купівельної активності від використання банківських карток.

Серед українських дослідників питання використання великих даних у фінансовій діяльності підприємств досліджували С. Супруненко, А. Чорновол і В. Гаврилюк [11]. Однак учені зосередили свою увагу на загальних питаннях управління економічними процесами.

Матеріали та методи. Для України дослідження можливості використання даних щодо карткових платежів має особливе значення. Унаслідок широкомасштабного вторгнення РФ відбулися суттєві зміни у структурі та розміщенні підприємств і населення. Також оновлення статистичних звітів підприємств щодо їхньої діяльності відбулося тільки влітку 2025 року. Усе це може суттєво викривлювати дані офіційної статистики. Тому розробка підходів до використання великих даних як індикатора економічного стану країни є актуальним завданням.

Дослідження проводилося на основі узагальнених даних Національного банку України (НБУ), зокрема на даних звітів про розвиток безготівкових платежів та платіжної інфраструктури [12]. Ці дані є результатом обробки великих масивів неструктурованих звітів (логів) платіжних систем і технологічними слідами, які демонструють роботу роздрібною торгівлі й електронної комерції. Розглядалися щомі-

сячні дані за період з січня 2020 року по жовтень 2025 року. Цей період охоплює два критичні шоки в економіці, пов'язані з відновленням комерційної діяльності після пандемії COVID-19 та повномасштабним вторгненням, що поглиблює аналіз впливу цифрових інструментів на споживчу активність населення. Набір даних містить такі показники:

- кількість електронних платіжних засобів, термін дії яких не закінчився (в обігу);
- кількість електронних платіжних засобів, за якими протягом звітного періоду була здійснена хоча б одна видаткова транзакція;
- кількість пристроїв самообслуговування (із них кількість банкоматів, кількість програмно-технічних комплексів самообслуговування (ПТКС));
- кількість платіжних терміналів (із них торговельних, безконтактних, банків та небанківських установ, із них безконтактних);
- кількість пунктів продажу товарів / наданих послуг.

Для аналізу впливу активності використання банківських карток на споживчий попит вжито індекс фізичного обсягу роздрібного товарообороту [13]. Методологічною основою дослідження стали методи машинного навчання (Machine Learning, ML). Для визначення індексу споживчої активності був побудований узагальнений індекс на основі методу головних компонент (Principal component analysis, PCA), що також дозволило визначити найбільш важливі показники впливу на узагальнений індекс.

Для ідентифікації зв'язку між показниками платіжної інфраструктури (великими даними) та рівнем споживчої активності в роботі використано класичний метод найменших квадратів (Ordinary Least Squares, OLS). Однак, урахувавши специфіку досліджуваного періоду (2020–2025 рр.), що характеризується надзвичайною волатильністю через пандемію та повномасштабну війну, стандартні припущення OLS щодо гомоскедастичності та відсутності автокореляції залишків часто порушуються.

Дані економічної активності під час війни демонструють явище гетероскедастичності – ситуацію, коли дисперсія випадкової помилки моделі не є сталою. Різкі шоки (обстріли інфраструктури, блекаути, міграційні хвилі) створюють викиди та нерівномірну мінливість транзакцій. Використання стандартного OLS за таких умов призводить до зміщення оцінок стандартних помилок, що робить *t*-статистики та *p*-values недостовірними, а висновки про значущість чинників – хибними.

Для отримання статистично спроможних висновків у роботі застосовано оцінювач НАС (Heteroskedasticity and Autocorrelation Consistent), також відомий як процедура Ньюї – Веста (Newey – West). Алгоритм автоматично коригує стандартні помилки коефіцієнтів, урахувавши змінну дисперсію залишків у періоди турбулентності. Оскільки транзакційні дані є часовими рядами, поточне значення

часто залежить від попереднього. НАС-корекція враховує часову залежність помилок, що є критичним для місячної статистики НБУ.

Також застосовувався інтелектуальний аналіз на основі моделі Prophet. В основі моделі лежить адитивна модель регресії, яка інтерпретує регресійну модель як комбінацію декількох компонентів. Основною перевагою цього методу є те, що модель Prophet автоматично визначає моменти, коли напрямок тренду різко змінюється, що є критично важливим для економіки України, яка за досліджуваний період пережила декілька криз. Модель гнучко підлаштовується під дані, не втрачаючи їхньої логіки. Для оцінки прогностичної здатності моделі використовувався метод ковзного вікна (Sliding Window), а не випадковий поділ, що є критичним для часових рядів.

Гіпотеза. Зростання кількості активних карток (ті, за якими була хоча б одна операція) є випереджальним індикатором відновлення споживчого попиту.

Результати дослідження. У сучасних умовах цифровізації економіки відбувається стрімке зростання обсягів даних, що генеруються в процесі функціонування фінансових, торговельних та інформаційних систем. Такі масиви інформації прийнято визначати як великі дані (Big Data), що характеризуються значним обсягом, високою швидкістю оновлення та різноманітністю джерел походження. Особливе місце серед джерел великих даних займають платіжні системи, що акумулюють інформацію про транзакційну активність населення, структуру споживчих витрат, рівень фінансової інклюзії та інтенсивність використання платіжної інфраструктури. На відміну від традиційних статистичних даних, такі масиви формуються у режимі, близькому до реального часу, що відкриває нові можливості для побудови оперативних аналітичних і прогнозних моделей.

Важливою перевагою великих даних є відносна неможливість маніпулювання ними. Оскільки інформація генерується автоматично в результаті фактичних транзакцій між економічними агентами, вона значно менше піддається суб'єктивним викривленням порівняно з традиційною статистичною звітністю. Крім того, агреговані дані платіжних систем часто є доступними для дослідників у вигляді відкритих або частково відкритих наборів, що публікуються центральними банками, міжнародними організаціями та фінансовими установами. Це забезпечує прозорість, відтворюваність та порівнянність результатів досліджень.

Застосування великих даних суттєво розширює можливості економічного аналізу, зокрема у сфері прогнозування. Методи машинного навчання та сучасної економетрики дозволяють ефективно працювати з високовимірними масивами інформації, виявляти приховані закономірності та формувати інтегральні індикатори, які адекватніше відображають

складні економічні процеси. Це особливо актуально для аналізу споживчої активності, яка формується під впливом великої кількості факторів і характеризується високою мінливістю.

Для економіки України проблема своєчасності та якості статистичної інформації набуває особливої актуальності. В умовах кризових явищ, спричинених, зокрема, повномасштабною російською агресією у 2022 році, спостерігаються значні труднощі у збиранні й обробці офіційної статистики. Так, можливі затримки у поданні статистичної звітності, обмеження доступу до даних з окремих регіонів, а також зниження точності окремих показників. У таких умовах особливої ваги набуває необхідність використання альтернативних джерел інформації, здатних забезпечити оперативну оцінку економічної ситуації. Споживча активність населення є одним із ключових драйверів економічного розвитку, оскільки вона безпосередньо впливає на обсяги виробництва, зайнятість та податкові надходження. Відповідно, своєчасне виявлення змін у поведінці споживачів дозволяє приймати більш обґрунтовані управлінські рішення як на макро-, так і на мікрорівні.

Отже, використання великих даних платіжних систем у поєднанні з методами машинного навчання й економетричного аналізу створює підґрунтя для формування нових підходів до моделювання та прогнозування споживчої активності, що є особливо важливим в умовах високої невизначеності та нестабільності економічного середовища.

З метою побудови узагальненого показника, що відображає динаміку споживчої активності населення на основі даних платіжних систем, було сформовано систему відносних індикаторів. Застосування саме відносних показників обумовлено необхідністю усунення масштабного ефекту та забезпечення порівнянності даних у часі. Введення нових індексів також пов'язане з тим, що абсолютні показники (кількість карток, терміналів тощо) не завжди коректно відображають інтенсивність економічної активності, тоді як їхні нормалізовані форми дозволяють краще інтерпретувати поведінкові характеристики споживачів.

Коефіцієнт активності карток (*Card_activity_ratio*) визначається за формулою:

$$Card_active_ratio = \frac{Cards_active}{Cards_total}, \quad (1)$$

де *Cards_active* – кількість банківських карток, за якими за досліджуваний період була здійснена хоча б одна транзакція, *Cards_total* – загальна кількість банківських карток.

Показник відображає рівень фактичного використання платіжних інструментів населенням. Зростання цього коефіцієнта свідчить про підвищення довіри до банківської системи, активізацію безготівкових розрахунків та загальне пожвавлення економічної активності.

Кількість POS-терміналів на одну картку обчислюється за формулою:

$$Pos_per_card = \frac{Pos_terminals_total}{Cards_total}, \quad (2)$$

де $Pos_terminals_total$ – кількість POS-терміналів.

Цей індикатор характеризує рівень забезпеченості платіжною інфраструктурою загальної кількості користувачів. Зростання значення показника свідчить про розширення можливостей здійснення безготівкових платежів, що своєю чергою стимулює споживчу активність.

Кількість терміналів на одну торгову точку ($Terminals_per_location$) визначається як:

$$Terminals_per_location = \frac{Pos_terminals_total}{Merchant_locations}, \quad (3)$$

де $Merchant_locations$ – кількість пунктів продажу товарів / надання послуг.

Економічний зміст показника полягає у характеристиці інтенсивності використання платіжної інфраструктури в місцях продажу. Значення, що перевищують одиницю, можуть свідчити про високу концентрацію платіжних пристроїв або дублювання інфраструктури, що інколи пов'язано з особливостями організації бізнесу або можливими неефективностями.

Частка безконтактних терміналів ($Contactless_share$) розраховується як:

$$Contactless_share = \frac{Pos_terminals_contactless}{Pos_terminals_total}, \quad (4)$$

де $Pos_terminals_contactless$ – кількість безконтактних POS-терміналів.

Цей показник ілюструє рівень технологічного розвитку платіжної інфраструктури та поширеність сучасних способів оплати. Зростання частки безконтактних терміналів зазвичай асоціюється з підвищенням швидкості та зручності здійснення платежів, що може опосередковано впливати на споживчу поведінку.

Перед застосуванням методу PCA всі змінні були нормалізовані з метою забезпечення їхньої порівнянності й уникнення домінування показників з більшими масштабами. У результаті застосування зазначеного методу отримано такі значення поясненої дисперсії за головними компонентами: перша – 80,3%; друга – 15,1%; наступні компоненти є незначними.

Кумулятивна частка поясненої дисперсії для перших двох компонент перевищує 95%, що свідчить про високу ефективність зменшення розмірності. Це дозволяє зробити висновок про доцільність використання першої головної компоненти як інтегрального індексу споживчої активності. Навантаження змінних на першу головну компоненту мають такий вигляд:

- pos_per_card – 0,4688;

- atm_total – 0,467;
- $card_activity_ratio$ – 0,455;
- $ptks_total$ – 0,432;
- $contactless_share$ – (-0,412).

Позитивні навантаження мають показники, що характеризують активність використання платіжних карток, забезпеченість платіжною інфраструктурою та доступ до фінансових послуг. Це дозволяє інтерпретувати зростання значення інтегрального індексу як збільшення споживчої активності населення. Водночас показник “частка безконтактних платежів” має від'ємне навантаження. Такий результат свідчить про складні структурні ефекти у споживчій поведінці, зокрема про те, що зростання частки безконтактних операцій не завжди супроводжується збільшенням загального обсягу витрат, а може відображати лише зміну технологічного способу здійснення платежів.

Отже, застосування методу головних компонент дозволило:

- зменшити багатовимірний простір показників до одного інтегрального індексу;
- зберегти понад 80% інформації вихідних даних;
- отримати економічно інтерпретований показник споживчої активності.

Побудований індекс представлений на рис. 1 (авторські розрахунки за даними [12; 13]).

Аналіз динаміки побудованого індексу свідчить про наявність двох шоків криз у споживчій активності населення. Перша криза спостерігалася на початку 2020 року і спричинена проблемами в економіці внаслідок обмежень, викликаних пандемією COVID-19. Однак тоді відбулося доволі швидке відновлення споживчої активності до передкризових рівнів. Друга криза мала місце у першій половині 2022 року, коли відбулося зниження споживчої активності до мінімального рівня за досліджуваний період. Відновлення є достатньо повільним, воно почалося тільки у 2023 році. У 2025 році відбулося досить різке збільшення темпів відновлення споживчої активності населення, однак передвоєнний рівень так і не був досягнутий.

Отриманий індекс авторками було використано на наступному етапі дослідження як пояснювальну змінну у моделях прогнозування динаміки роздрібно-го товарообороту.

З метою аналізу та прогнозування споживчого попиту, який є базовою складовою формування роздрібно-го товарообороту, побудовано економетричну модель на основі методу найменших квадратів. У цьому контексті індекс фізичного обсягу роздрібно-го товарообороту розглядається як агрегований показник, що відображає реалізований споживчий попит в економіці. За пояснювальну змінну використано інтегральний індекс споживчої активності, побудований на основі даних платіжних систем, що інтерпретується як випереджальний індикатор змін у поведінці споживачів.

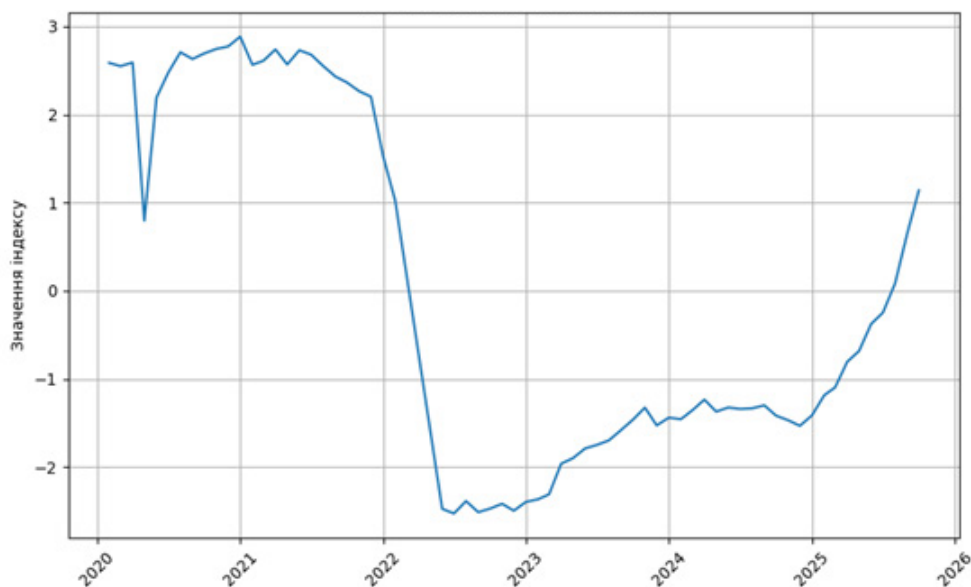


Рис. 1. Динаміка побудованого індексу споживчої активності населення на основі використання банківських карток

З урахуванням динамічної природи споживчого попиту модель специфіковано з включенням лагових змінних:

$$Y_t = \beta_0 + \beta_1 I_t + \beta_2 I_{t-1} + \beta_3 Y_{t-1} + \varepsilon_t, \quad (5)$$

де Y_t – індекс роздрібного товарообороту як проксі-показник реалізованого споживчого попиту; I_t – інтегральний індекс споживчої активності; I_{t-1} – лаг індексу, що відображає відкладений вплив поведінкових

факторів; Y_{t-1} – лаг залежної змінної, який враховує інерційність попиту.

Для підвищення надійності статистичних оцінок застосовано робастні стандартні помилки типу HAC, що дозволяє урахувати можливу автокореляцію та гетероскедастичність залишків. Розрахунок параметрів моделі здійснювався з допомогою розробленої програми на мові Python з використанням бібліотек Statsmodels та Prophet. Параметри моделі та характеристики критеріїв наведені, відповідно, у табл. 1, 2 (авторські розрахунки за даними [12; 13]).

Таблиця 1

Параметри економетричної моделі споживчої активності

Коефіцієнт	Значення	Стандартна похибка	Критерій z	Рівень значущості $P > z $	Довірчий інтервал для $\alpha=0,05$
const	21,6576	10,664	2,031	0,042	[0,756; 42,560]
activity_index	9,2625	3,131	2,959	0,003	[3,126; 15,399]
index_lag1	-8,9999	3,059	-2,942	0,003	[-14,995; -3,004]
retail_lag1	0,7983	0,092	8,713	0,000	[0,619; 0,978]

Таблиця 2

Характеристики економетричної моделі споживчої активності (OLS)

Показник	Значення
Кількість спостережень	64
Omnibus	83,02
$P(\text{Omnibus})$	0,000
R^2	0,706
Скоригований R^2	0,691
F -статистика	40,65
$P(F\text{-статистика})$:	1,81e-14
Log-Likelihood	-231,87
Критерій AIC	471,7
Критерій Дарбіна – Уотсона	2,002
Критерій Жака – Бера	1148,424
$P(\text{Жака – Бера})$	4,20e-250

Отримані значення коефіцієнта детермінації $R^2 = 0,706$ при рівні статистичної значущості моделі в цілому ($p < 0,001$) та коефіцієнт Дарбіна – Уотсона $\approx 2,0$ вказують на якість побудованої моделі та відсутність автокореляції залишків. Отже, інтегральний індекс споживчої активності є статистично значущим фактором, який визначає динаміку споживчого попиту. Модель має вигляд:

$$Y_t = 21,058 + 9,263I_t - 9,0I_{t-1} + 0,798Y_{t-1} + \varepsilon_t \quad (6)$$

Позитивний коефіцієнт при поточному значенні індексу вказує на те, що зростання платіжної активності населення супроводжується збільшенням

обсягів споживання. Водночас наявність від'ємного коефіцієнта при лаговому значенні індексу свідчить про існування ефекту короткострокового насичення попиту. Це означає, що підвищена активність споживачів у попередньому періоді може призводити до часткового зниження попиту в наступному, що відображає міжчасовий перерозподіл витрат. Суттєвий вплив лагового значення залежної змінної підтверджує інерційний характер споживчого попиту, що є типовим для макроекономічних процесів та свідчить про поступовість змін у поведінці споживачів. Результати порівняльного аналізу фактичних і розрахункових значень роздрібного товарообороту наведені на рис. 2 (авторські розрахунки за даними [12; 13]).



Рис. 2. Порівняння фактичних та модельних значень роздрібного товарообороту, отриманих з допомогою економетричної моделі

Отже, побудована модель дозволяє інтерпретувати інтегральний індекс, сформований на основі великих даних платіжних систем, як випереджальний індикатор змін у споживчому попиті. Це створює можливості для оперативнішого прогнозування динаміки роздрібного товарообороту та підвищує ефективність аналітичного забезпечення економічної політики.

З метою порівняння ефективності класичних економетричних підходів із сучасними методами прогнозування часових рядів у дослідженні було застосовано модель Prophet, розроблену компанією Meta Platforms. Prophet є адитивною моделлю часових рядів, яка дозволяє враховувати трендову, сезонну та випадкову компоненти процесу. У загальному вигляді модель може бути представлена як:

$$Y(t) = g(t) + s(t) + \varepsilon(t), \quad (7)$$

де $g(t)$ – трендовий компонент, $s(t)$ – сезонний компонент, $\varepsilon(t)$ – випадкова складова.

У межах цього дослідження модель Prophet використовувалася для прогнозування динаміки індексу

роздрібного товарообороту як проксі-показника споживчого попиту. Вхідними даними слугували часові ряди щомісячних значень показника. Особливістю моделі є її здатність автоматично виявляти тренди та сезонні коливання без необхідності складної попередньої специфікації, що робить її зручною для практичного застосування. Графік порівняння фактичних і змодельованих значень наведений на рис. 3 (авторські розрахунки за даними [12; 13]).

Отримані результати показують, що модель дуже точно відстежує динаміку до 2024 року. Це пояснює високий коефіцієнт детермінації ($R^2 = 0,82$). Модель успішно виявила значне падіння початку 2022 року та наступне відновлення. Однак унаслідок високої волатильності минулого (широкомасштабне вторгнення) прогноз на 2025 рік мав велику похибку, хоча алгоритм ідентифікував зміну структурної логіки ринку, адаптувавши тренд до нових реалій функціонування платіжної системи.

Візуалізація компонентів моделі Prophet (рис. 4, авторські розрахунки за даними [12; 13]) дозволила ізолювати чистий ефект впливу активних карток (Trend) від циклічних коливань (Seasonality).

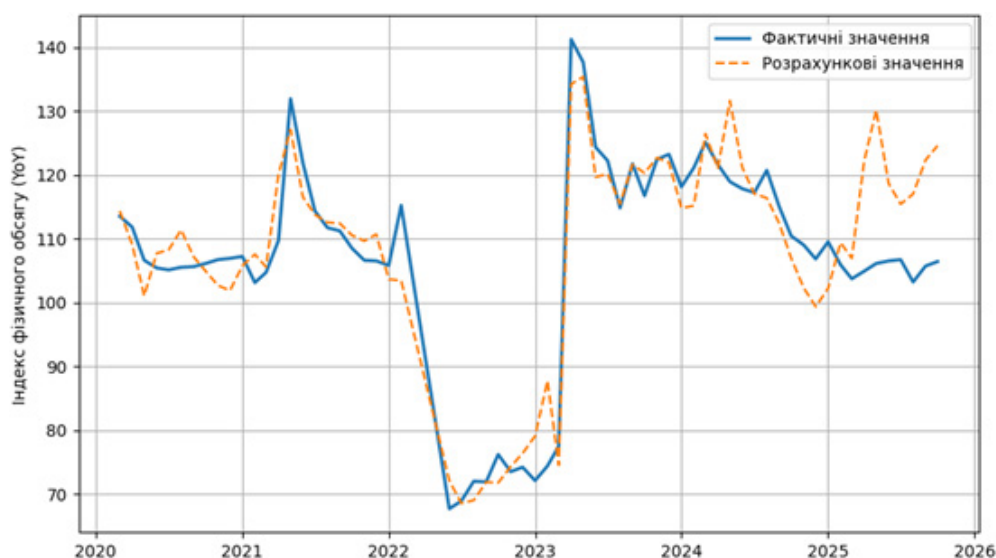


Рис. 3. Порівняння фактичних і модельних значень роздрібногo товарообороту, отриманих з допомогою моделі Prophet

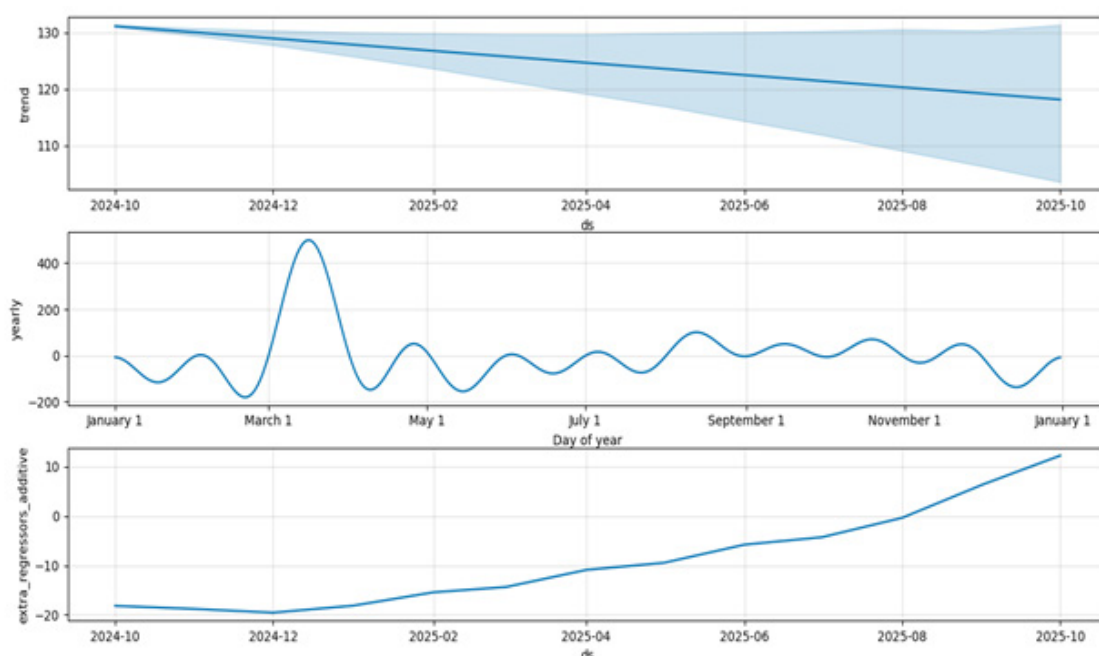


Рис. 4. Виокремлені компоненти моделі Prophet

На першому підграфіку спостерігається стійке зниження тренда споживчої активності. Це зумовлено синергетичним впливом міграційного відтоку населення та зниженням реальної купівельної спроможності в умовах воєнного стану.

На другому підграфіку виділений сезонний компонент, який свідчить про певну зміну споживчих звичок у кризові періоди: найбільший пік спостерігається у березні, також незначне підвищення можна побачити у період серпень – грудень. Таку споживчу активність можна пояснити традиційним зростанням транзакцій у сегментах товарів і послуг повсякденного попиту (Fast-Moving Consumer Goods, FMCG), що корелює з національними особливостями споживчої поведінки в першій декаді місяця; крім

того, підвищення соціальних виплат зазвичай починає відчуватися саме в березні.

Нижній підграфік демонструє вплив індексу активності банківських карток. Він показує, як зміна кількості активних карток штовхає роздрібний товарообіг вгору.

Отримані результати свідчать про те, що модель Prophet добре відтворює історичну динаміку показника, проте має обмежену здатність до узагальнення та прогнозування. На відміну від моделі Prophet, побудована економетрична модель з використанням інтегрального індексу споживчої активності продемонструвала значно кращі результати прогнозування, що пояснюється можливістю таких моделей швидше реагувати на різкі зміни тенденцій у часових рядах.

Встановлено, що зовнішній регресор *activity_index* є критичним параметром, який пояснює різкі амплітудні зсуви в періоди структурних криз. Це підтверджує гіпотезу про високу прогностичну цінність транзакційних великих даних для цілей макроекономічного моніторингу

Висновки та рекомендації. У статті досліджено можливості використання великих даних платіжних систем для побудови прогностичних моделей динаміки споживчого попиту як ключового чинника формування роздрібного товарообороту.

У ході дослідження сформовано систему відносних показників, що характеризують різні аспекти функціонування платіжної інфраструктури та поведінки споживачів. На основі цих показників із застосуванням методу головних компонент побудовано інтегральний індекс споживчої активності, який дозволив узагальнити багатовимірну інформацію та зберегти основну частку варіації вихідних даних. Результати економетричного моделювання засвідчили, що запропонований індекс є статистично значущим фактором формування споживчого попиту. Побудована модель продемонструвала високу пояснювальну здатність та дозволила виявити як інерційні, так і короткострокові коригувальні ефекти у динаміці споживчої поведінки.

Порівняльний аналіз показав, що застосування методів машинного навчання (зокрема, моделі часових рядів Prophet) не забезпечує покращення якості прогнозування у досліджуваних умовах. Отримані результати свідчать про обмежену здатність зазначених підходів до узагальнення за умов відносно невеликої вибірки та наявності структурних зламів у часових рядах.

Виявлено, що з 2024 року відбулося суттєве зниження споживчого попиту, зумовлене кризовими явищами в економіці України. Показано, що індекс активності банківських карток є фактором, який штовхає споживчий попит. Також висвітлені особливості споживчого попиту в Україні у 2022–2024 роках, які демонструють певну зміну купівельних звичок українців, зумовлену як обмеженістю купівельної спроможності, так і прив'язкою до змін у соціальних виплатах. На тлі цього використання великих даних платіжних систем дає змогу сформуванню випереджальних індикаторів, які адекватніше відображають поточний стан економічної активності.

Практична значущість отриманих результатів полягає у можливості використання запропонованого підходу для оперативного моніторингу та прогнозування споживчого попиту, що є особливо актуальним в умовах нестабільності економічного середовища, спричиненого російським вторгненням в Україну у 2022 році. Це відкриває перспективи для підвищення якості аналітичного забезпечення економічної політики та прийняття обґрунтованих управлінських рішень.

Отже, результати дослідження підтверджують доцільність інтеграції методів обробки великих даних та економетричного аналізу для побудови ефективних моделей прогнозування макроекономічних показників, зокрема у сфері споживчої активності.

Подальші дослідження в цій області авторки планують спрямувати на дослідження впливу транзакційних даних на споживчу активність у регіональному розрізі та виявлення аномальних проявів для визначення впливу цифрових трансформацій на вихід економіки з тіні.

Список використаних джерел

1. Model to evaluate the Industry 4.0 readiness degree in industrial companies / W. C. Lucato et al. *IFAC-PapersOnLine*. 2019. Vol. 52, No. 13. P. 1808–1813. DOI: 10.1016/j.ifacol.2019.11.464
2. Einav L., Levin J. Economics in the age of big data. *Science*. 2014. Vol. 346, No. 6210. P. 715–721. DOI: <https://doi.org/10.1126/science.1243089>
3. Oneshko S., Pashchuk L. Industry 4.0 and creative economy (globalization challenges of the time). *Futurity Economics & Law*. 2021. Vol. 1, No. 4. P. 4–11. DOI: <https://doi.org/10.57125/FEL.2021.12.25.01>
4. Дзямулич М. І., Шматковська Т. О., Борисюк О. В. Великі дані та їх роль у формуванні цифрової економіки. *Галицький економічний вісник*. 2021. Т. 70, № 3. С. 16–21. DOI: https://doi.org/10.33108/galicianvisnyk_tntu2021.03
5. Cong L. W., Li B., Zhang Q. T. Alternative data in fintech and business intelligence. *The Palgrave Handbook of FinTech and Blockchain*. 2021. Chapter 9. P. 217–242. DOI: 10.1007/978-3-030-66433-6_9
6. Using Big Data to optimize economic processes in the digital age / N. Sapotnitska et al. *Financial and Credit Activity: Problems of Theory and Practice*. 2023. № 4 (51). С. 164–174. DOI: <https://doi.org/10.55643/fcptr.4.51.2023.4131>
7. Птащенко Л., Добропас А. Використання технології Big Data при наукових дослідженнях в економіці. *Grail of Science*. 2025. № 59. С. 334–340. DOI: <https://doi.org/10.36074/grail-of-science.12.12.2025.034>
8. Predicting regional economic indices using big data of individual bank card transactions / S. Sobolevsky et al. 2017 IEEE International Conference on Big Data. 2017. P. 1313–1318. DOI: 10.1109/BigData.2017.8258061
9. The topology of card transaction money flows / M. Zanin et al. *Physica A: Statistical Mechanics and its Applications*. 2016. Vol. 462. P. 134–140. <https://doi.org/10.1016/j.physa.2016.06.091>

10. See-To E. W. K., Ngai E. W. T. An empirical study of payment technologies, the psychology of consumption, and spending behavior in a retailing context. *Information & Management*. 2019. Vol. 56, No. 3. P. 329–342. DOI: <https://doi.org/10.1016/j.im.2018.07.007>
11. Супруненко С., Чорновол А., Гаврилюк В. Використання аналітики даних для управління фінансовими процесами в цифровому середовищі України. *Економіка та суспільство*. 2024. № 62. DOI: <https://doi.org/10.32782/2524-0072/2024-62-51>
12. Ринок платіжних карток. Набір даних. Єдиний державний веб-портал відкритих даних. URL: <https://data.gov.ua/dataset/e69526a2-26bb-4278-a49b-bdc4163cb662> (дата звернення: 25.03.2025).
13. Індекс фізичного обсягу роздрібного товарообороту України. Банк даних Державної служби статистики України. URL: https://stat.gov.ua/uk/explorer?urn=SSSU%3ADF_SALE_AND_STOCKS_OF_GOODS_RETAIL_A_Q%28~%29&filter=IDX_TRNVR_RTL_ENTRP_VOL (дата звернення: 25.03.2025).

References

1. Lucato, W. C., Pacchini, A. P. T., Facchini, F., & Mummolo, G. (2019). Model to evaluate the Industry 4.0 readiness degree in industrial companies. *IFAC-PapersOnLine*, 52 (13), 1808–1813. DOI: 10.1016/j.ifacol.2019.11.464
2. Einav, L., & Levin, J. (2014). Economics in the age of big data. *Science*, 346 (6210), 715–721. <https://doi.org/10.1126/science.1243089>
3. Oneshko, S., & Pashchuk, L. (2021). Industry 4.0 and creative economy (globalization challenges of the time). *Futurity Economics & Law*, 1 (4), 4–11. <https://doi.org/10.57125/FEL.2021.12.25.01>
4. Dziamulych, M. I., Shmatkovska, T. O., & Borysiuk, O. V. (2021). Big data and their role in the formation of the digital economy. *Galician Economic Bulletin*, 70 (3), 16–21. https://doi.org/10.33108/galicianvisnyk_tntu2021.03 [in Ukrainian].
5. Cong, L. W., Li, B., & Zhang, Q. T. (2021). Alternative data in fintech and business intelligence. *The Palgrave Handbook of FinTech and blockchain*. M. Pompella & R. Matousek (Eds.). (Chapter 9, pp. 217–242). DOI: 10.1007/978-3-030-66433-6_9
6. Sapotnitska, N., Ovander, N., Harkava, V., Kirieieva, K., & Orlenko, O. (2023). Using Big Data to optimize economic processes in the digital age. *Financial and Credit Activity: Problems of Theory and Practice*, 4 (51), 164–174. <https://doi.org/10.55643/fcaptop.4.51.2023.4131>
7. Ptashchenko, L., & Dobropas, A. (2025). Vykorystannia tekhnolohii Big Data pry naukovykh doslidzhenniakh v ekonomitsi [Use of big data technology in economic research]. *Grail of Science*, 59, 334–340. <https://doi.org/10.36074/grail-of-science.12.12.2025.034> [in Ukrainian].
8. Sobolevsky, S., Massaro, E., Bojic, I., Arias, J. M., & Ratti, C. (2017). Predicting regional economic indices using big data of individual bank card transactions. *Proceedings from 2017 IEEE International Conference on Big Data (Big Data)*. (pp. 1313–1318). DOI: 10.1109/BigData.2017.8258061
9. Zanin, M., Papo, D., Romance, M., Criado, R., & Moral, S. (2016). The topology of card transaction money flows. *Physica A: Statistical Mechanics and Its Applications*, 462, 134–140. <https://doi.org/10.1016/j.physa.2016.06.091>
10. See-To, E. W. K., & Ngai, E. W. T. (2019). An empirical study of payment technologies, the psychology of consumption, and spending behavior in a retailing context. *Information & Management*, 56 (3), 329–342. <https://doi.org/10.1016/j.im.2018.07.007>
11. Супруненко С., Чорновол А., & Гаврилюк В. (2024). Vykorystannia analytyky danykh dlia upravlinnia finansovymy protsesamy v tsyfrovomu seredovyshchi Ukrainy [Using Data Analytics to Manage Financial Processes in The Digital Environment of Ukraine]. *Economy and Society*, 62. <https://doi.org/10.32782/2524-0072/2024-62-51> [in Ukrainian].
12. Open data portal. (2025). *Rynok platizhnykh kartok. Nabir danykh [Payment card market. Dataset]*. Retrieved from <https://data.gov.ua/dataset/e69526a2-26bb-4278-a49b-bdc4163cb662> [in Ukrainian].
13. Ukrstat. (2025). *Index of physical volume of retail turnover of Ukraine*. Data Bank. Retrieved March 25, 2025 from https://stat.gov.ua/en/explorer?urn=SSSU%3ADF_SALE_AND_STOCKS_OF_GOODS_RETAIL_A_Q%28~%29&filter=IDX_TRNVR_RTL_ENTRP_VOL

Надійшла до редакції / Received on: 23.01.2026

Прорецензована / Reviewed on: 12.02.2026

Підписана до друку / Signed for printing on: 25.02.2026

Оприлюднена / Published on: 20.04.2026

T. V. Tomashevskya,

*PhD in Technical Sciences, Associate Professor,
Associate Professor of the Department of Computer Sciences
and Information Systems,
State University of Trade and Economics,
E-mail: t.tomashevskya@knute.edu.ua*

*Researcher ID: I-8901-2018,
ORCID: <http://orcid.org/0000-0001-5001-9226>;*

N. O. Hordiiko,

*PhD in Technical Sciences, Associate Professor,
Associate Professor of the Department of Applied Physics,
National Technical University of Ukraine "Igor Sikorsky Kyiv Polytechnic Institute",
E-mail: n.hordiiko-ipt@ill.kpi.ua*

ORCID: <https://orcid.org/0000-0002-0925-5160>

Predictive Modelling for the Dynamics of Consumer Activity by Methods of Machine Learning and Big Data of Payment Systems

The article contains a feasibility analysis of using big data of payment systems for building forecasting models for the dynamics of consumer demand as a core factor of the retail turnover. The relevance of this research stems from the need for prompt and reliable estimates of economic activity in a highly uncertain and volatile economic environment, especially due to the large-scale Russian invasion of Ukraine in 2022.

It is demonstrated that using big data as an alternative information source is an expedient option to minimize time lags intrinsic in the official statistics, and to enhance the accuracy of assessment of the current economic performance. A set of relative indicators was built to measure various operative aspects of the payment infrastructure and consumer behavior: ratio of payment card activity, availability of POS terminals, intensity of using payment infrastructure, and proliferation of contactless technologies. Principal component analysis was applied to this set of indicators to build Integral index of consumer activity (IICA), summing up multidimensional information and preserving a larger part of variation in output data. It was found that the first principal component explained more than 80% of the overall variance. The results show that IICA has a high information capacity and can be interpreted as a leading indicator of change in the consumer demand.

To estimate IICA impact on the dynamics of consumer demand, an econometric model was built by least squares with account for lag dependencies. Index of Physical Volume of Retail Turnover was used as an output variable, interpreted as a proxy indicator for the effective consumer demand. The results of the modelling show statistical significance of IICA, and a high explanatory capacity of the model, thus confirming the existence of a stable correlation between payment activity indicators and macroeconomic dynamics.

The research includes a comparative analysis of the efficiency of classical econometric methods and advanced methods of machine learning, Prophet model of time series in particular. The model demonstrated high-quality approximation on a training sample but was found inefficient for forecasting under sudden economic changes. The results showed that the above methods could not ensure quality of forecasting compared with the basic econometric model. Models of machine learning demonstrated low capacity for summarization on test samples, which might be attributed to a limited data volume and the existence of structural breaks in time series.

It is concluded that a core factor enhancing the accuracy of forecasting is not so much the model's complexity as the quality and information capacity of input data. Using big data of payment systems allows one to build leading indicators measuring more adequately the current economic performance and ensuring a prompt response on changes in the consumer behavior.

The practical significance of this research is that the proposed approach is applicable for monitoring and forecasting consumer demand in the mode approximated to real time, which is especially important for managerial decision-making in the conditions of crises.

Key words: *big data, modelling of consumer demand, retail turnover, Integral index of consumer activity, methods of machine learning, Prophet model.*

Цитування:

Томашевська Т. В., Гордійко Н. О. Побудова прогностичних моделей динаміки споживчої активності на основі методів машинного навчання та Big Data платіжних систем. *Статистика України*. 2026. № 1. С. 44–54. Doi: 10.31767/su.1(112)2026.01.05

Cite this article:

Tomashevskya, T. V., & Hordiiko, N. O. (2026). Pobudova prohnoznykh modelei dynamiky spozhyvchoi aktyvnosti na osnovi metodiv mashynnoho navchannia ta Big Data platizhnykh system [Predictive Modelling for the Dynamics of Consumer Activity by Methods of Machine Learning and Big Data of Payment Systems]. *Statystyka Ukrainy – Statistics of Ukraine*, 1, 44–54. Doi: 10.31767/su.1(112)2026.01.05 [in Ukrainian].