

Міністерство освіти і науки України
Департамент науки і освіти Харківської облдержадміністрації
Комунальний заклад «Харківська обласна Мала академія наук
Харківської обласної ради»

Відділення математики
Секція: статистика

ЕКОНОМІКО-СТАТИСТИЧНЕ МОДЕЛЮВАННЯ ЗАБРУДНЮЮЧИХ ВИКИДІВ У ПОВІТРЯ В УКРАЇНІ

Роботу виконала:

Липенко Аліса Вячеславівна, учениця
10 класу Комунального закладу «Харківський
ліцей № 89 Харківської міської ради»,
вихованка Комунального закладу «Харківська
обласна Мала академія наук Харківської
міської ради Харківської області»

Наукові керівники:

Якунін Анатолій Вікторович, доцент кафедри
вищої математики і математичного
моделювання Харківського національного
університету міського господарства імені
О.М. Бекетова, кандидат технічних наук,
доцент;

Бондаренко Ірина Станіславівна, вчитель
математики Комунального закладу
«Харківський ліцей № 89 Харківської міської
ради», спеціаліст вищої категорії, учитель-
методист

Харків – 2024

ЕКОНОМІКО-СТАТИСТИЧНЕ МОДЕЛЮВАННЯ ЗАБРУДНЮЮЧИХ ВИКИДІВ У ПОВІТРЯ В УКРАЇНІ

Липенко Аліса Вячеславівна; Харківське територіальне відділення МАН України, Комунальний заклад «Харківська обласна Мала академія наук Харківської обласної ради», Комунальний заклад «Харківський ліцей № 89 Харківської міської ради», 10 клас, м. Харків;

Якунін Анатолій Вікторович, доцент кафедри вищої математики і математичного моделювання Харківського національного університету міського господарства імені О.М. Бекетова, кандидат технічних наук, доцент;

Бондаренко Ірина Станіславівна, вчитель математики Комунального закладу «Харківський ліцей № 89 Харківської міської ради», спеціаліст вищої категорії, учитель-методист

Актуальність роботи обумовлена необхідністю ґрунтовних наукових досліджень щодо вивчення і кількісного відображення статистичних зв'язків рівнів викидів забруднюючих речовин у повітря в Україні з основними показниками суспільно-економічного розвитку.

Мета роботи – створити регресійну модель інтенсивності викидів забруднюючих речовин в атмосферне повітря в Україні та перевірити характеристики її якості.

Відповідно до поставленої мети в роботі вирішуються такі основні завдання: провести збір і попередню обробку статистичних даних щодо забруднюючих викидів у повітря в Україні та, орієнтуючись на експертні оцінки, гіпотетично пов'язаних з ним загально-національних показників розвитку України; проаналізувати сучасні досягнення у статистичному аналізі часових рядів та їх моделюванні і вибрати перспективні підходи з урахуванням специфіки протікання та статистичного відображення соціально-економічних процесів, наявні комп'ютерні засоби обчислення кількісних статистичних показників, методи вибору структури математичних моделей та оптимального оцінювання їхніх параметрів; запропонувати

обґрунтовані неформальні побудови комплексних еколого-економічних показників для опису забруднюючих викидів у повітря в Україні; розробити обчислювальні процедури для розрахунків найпоширеніших статистичних показників ряду динаміки забруднюючих викидів; провести кореляційний аналіз часового ряду забруднюючих викидів у повітря в Україні та, орієнтуючись на експертні оцінки, гіпотетично пов'язаних з ним загальних фізичних, техніко-економічних і соціальних характеристик; запропонувати лінійну регресійну модель інтенсивності забруднюючих викидів в атмосферне повітря в Україні з оптимізованою кількістю основних екзогенних факторів; здійснити параметричну ідентифікацію синтезованої моделі; дослідити адекватність поданої моделі та значущість її коефіцієнтів.

Синтезована регресійна модель інтенсивності викидів забруднюючих речовин в атмосферне повітря в Україні має апроксимаційний характер і потребує постійного уточнення відповідно до корекції наявних даних і появи нової статистичної інформації. Перспективність моделі обумовлена можливістю її модернізації додаванням нових неврахованих факторів чи іншою трансформацією предикторних змінних, введенням зовнішніх обмежень у рамках відомих процедур лінійного регресійного аналізу.

Ключові слова: часовий ряд, статистичні показники, дисперсійний аналіз, кореляційний аналіз, регресійний аналіз, метод найменших квадратів, адекватність, значущість.

ЗМІСТ

ВСТУП	5
РОЗДІЛ 1. ЧАСОВІ РЯДИ ЕКОНОМЕТРИКИ: СТАТИСТИЧНИЙ АНАЛІЗ І МОДЕЛЮВАННЯ	8
Висновки до першого розділу	12
РОЗДІЛ 2. СТАТИСТИЧНІ ХАРАКТЕРИСТИКИ ДИНАМІКИ ЗАБРУДНЮЮЧИХ ВИКИДІВ І ГІПОТЕТИЧНО ПОВ'ЯЗАНИХ СОЦІАЛЬНО-ЕКОНОМІЧНИХ ПОКАЗНИКІВ	13
2.1. Показники положення, варіації та внутрішніх взаємозв'язків між рівнями часового ряду інтенсивності викидів	13
2.2. Дослідження тісноти взаємозв'язків між інтенсивністю забруднюючих викидів і гіпотетично пояснюючими фізичними та соціально-економічними показниками	17
Висновки до другого розділу	23
РОЗДІЛ 3. МОДЕЛЮВАННЯ ДИНАМІКИ ІНТЕНСИВНОСТІ ЗАБРУДНЮЮЧИХ ВИКИДІВ У ПОВІТРЯ В УКРАЇНІ	24
3.1. Однофакторна модель для експрес-розрахунків інтенсивності забруднюючих викидів	24
3.2. Багатофакторна модель інтенсивності забруднюючих викидів	27
3.3. Дослідження залишків багатофакторної моделі, її адекватності та значущості коефіцієнтів	31
Висновки до третього розділу	36
ВИСНОВКИ	38
СПИСОК ВИКОРИСТАНИХ ДЖЕРЕЛ	41
ДОДАТКИ	44

ВСТУП

В умовах глобальних викликів одним з пріоритетних завдань екологічної політики є ефективний захист довкілля з урахуванням національних особливостей і традицій природокористування. Це передбачає постійний екологічний моніторинг і проведення науково-технічних, соціально-економічних і організаційних заходів з природокористування і охорони навколишнього середовища, екологічний стан якого суттєво визначається викидами забруднюючих речовин у повітря [13, 17, 18].

Матеріально-технічною базою таких досліджень слугує широке застосування комп'ютеризованих інформаційних систем, що дозволяють опрацьовувати великі масиви статистичних даних за допомогою сучасного математичного апарату [28]. Взаємозв'язок соціально-економічних факторів і екологічних показників досліджено у працях багатьох науковців [13, 17, 18], що постійно удосконалюють принципи побудови економетричних моделей і процедури оперування з ними.

Важливою спільною особливістю статистичних показників соціально-економічних і технологічних процесів та їх екологічних наслідків, зокрема утворення забруднюючих викидів, є змінювання з бігом часу, що призводить до формування часових рядів. Тому темою цього дослідження обрано статистичний аналіз часових рядів – викидів забруднюючих речовин в атмосферне повітря в Україні та основних фізичних, техніко-економічних і соціальних показників, що гіпотетично (згідно експертних оцінок) можуть впливати на обсяги та динаміку забруднюючих викидів.

Застосування засобів і методів комп'ютерного економіко-статистичного моделювання [2, 12, 14, 21], що передбачає поєднання предметного якісного аналізу природи явища, обчислення його кількісних статистичних показників та їх динаміки, дозволяє конкретизувати математичний опис процесу і підвищити ефективність його дослідження та напрацювання управлінських рішень щодо проблем довкілля. Зокрема, надає можливість упорядкувати наявну інформацію щодо забруднюючих викидів та їх залежності від основних показників розвитку національної економіки.

Актуальність роботи обумовлена обмеженістю наукових досліджень щодо математичного відображення ряду динаміки забруднення повітря шкідливими

викидами та необхідністю ґрунтового вивчення залежності ряду забруднюючих викидів від динаміки основних фізичних, техніко-економічних і соціальних показників розвитку України всупереч негативному впливу війни.

Мета дослідження: синтезувати лінійну регресійну модель інтенсивності забруднюючих викидів в атмосферне повітря в Україні з оптимізованою кількістю основних екзогенних факторів та дослідити адекватність поданої моделі і значущість її коефіцієнтів.

Відповідно до поставленої мети в роботі вирішуються такі основні **завдання:**

– провести збір і попередню обробку статистичних даних щодо забруднюючих викидів у повітря в Україні та, орієнтуючись на експертні оцінки, гіпотетично пов'язаних з ним загально-національних показників розвитку України;

– проаналізувати сучасні досягнення у статистичному аналізі часових рядів та їх моделюванні і вибрати перспективні підходи з урахуванням специфіки протікання та статистичного відображення соціально-економічних процесів, наявні комп'ютерні засоби обчислення кількісних статистичних показників, методи вибору структури математичних моделей та оптимального оцінювання їхніх параметрів;

– запропонувати обґрунтовані неформальні побудови комплексних еколого-економічних показників для опису забруднюючих викидів у повітря в Україні;

– розробити обчислювальні процедури для розрахунків найпоширеніших статистичних показників ряду динаміки забруднюючих викидів;

– провести кореляційний аналіз часового ряду забруднюючих викидів у повітря в Україні та, орієнтуючись на експертні оцінки, гіпотетично пов'язаних з ним загальних фізичних, техніко-економічних і соціальних характеристик;

– запропонувати лінійну регресійну модель інтенсивності забруднюючих викидів з оптимізованою кількістю основних екзогенних факторів;

– здійснити параметричну ідентифікацію синтезованої моделі;

– дослідити адекватність поданої моделі та значущість її коефіцієнтів.

Об'єкт дослідження: економіко-статистичний аналіз і моделювання залежності часового ряду забруднюючих викидів у повітря в Україні від основних загально-національних показників розвитку як компоненти багатовимірного

нестационарного стохастичного процесу, що супроводжується дією неконтрольованих факторів різної природи.

Предмет дослідження: розрахунки основних статистичних показників положення, інтенсивності та варіації часових рядів забруднюючих викидів у повітря України та пов'язаних з ним загально-національних показників розвитку України; ідентифікація лінійної регресійної моделі інтенсивності забруднюючих викидів в атмосферне повітря в Україні з оптимізованою кількістю основних екзогенних факторів.

Для вирішення поставлених завдань використовуються **методи:** теорії ймовірностей і математичної статистики, економіко-експертного оцінювання, математичного моделювання, числові процедури дисперсійного, кореляційного та регресійного аналізу і математичного програмування, комп'ютерного обчислювального експерименту.

Гіпотеза: загально-національні показники розвитку України як складної соціально-економічної системи відображають різні сторони єдиного багатовимірного нестационарного стохастичного процесу, в якому, спираючись на загальні закономірності часових рядів з урахуванням фізичних, технологічних, техніко-економічних і соціальних впливів, можна виділити окремі компоненти та виявити взаємозв'язки між ними. Виходячи з обмеженості інформації щодо внутрішньої структури і закономірностей функціонування цієї складної системи, на даному етапі досліджень з позицій економетрики досить розглянути лише типові статистичні характеристики об'єктів і апроксимаційні підходи до їх математичного подання.

Запропоновані процедури обчислення статистичних показників легко переносяться на розширені масиви вхідних даних. Синтезована регресійна модель інтенсивності викидів забруднюючих речовин в атмосферне повітря в Україні має апроксимаційний характер і потребує постійного уточнення відповідно до корекції наявних даних і появи нової статистичної інформації. Перспективність моделі обумовлена можливістю трансформації масиву предикторних змінних, введення зовнішніх обмежень у рамках відомих процедур лінійного регресійного аналізу.

РОЗДІЛ 1

ЧАСОВІ РЯДИ ЕКОНОМЕТРИКИ: СТАТИСТИЧНИЙ АНАЛІЗ І МОДЕЛЮВАННЯ

Часові ряди використовуються для опису процесів, що протікають у часі та відображаються в послідовні дискретні часові моменти чи інтервали. Часовий ряд – це послідовність точок даних, проіндексованих у хронологічному порядку. Далі часовий ряд розглядається як вибіркова реалізація з нескінченної генеральної сукупності, що генерується стохастичним процесом, на який впливають багато факторів різної природи [1, 2, 6, 11, 12, 14]. Велика різноманітність наявних індикаторів тенденції, мінливості та взаємозв'язків рівнів окремих часових рядів та їх систем, напрацьованих методів і засобів дослідження і недостатня обізнаність прикладних фахівців про особливості їх використання потребують уваги щодо створення практичних процедур вибору і застосування відповідного математичного апарату для аналізу часових рядів з урахуванням їх специфіки.

У задачах моделювання та прогнозування часових рядів найбільш широко застосовується статистичний підхід [16], який ґрунтується на відновленні за конкретним часовим рядом наближеної моделі, що статистично відображає поведінку реального процесу. При цьому необхідною умовою проведення аналізу часового ряду є однорідність вхідних даних за структурою, тобто вони мають бути представлені однаковими показниками в єдиному масштабі та відображати однорідні за величиною та інтенсивністю реальні процеси. За наявності пропущених чи аномальних спостережень відповідні значення повинні бути відновлені чи скореговані.

Популярність статистичного моделювання пояснюється досить високим рівнем його ефективності: економність за кількістю параметрів, що потребують оцінювання; висока продуктивність процедур ідентифікації; малі витрати на реалізацію та можливість використання на комп'ютерах обмеженої потужності; практично прийнятні результати прогнозування різноманітних процесів; математична обґрунтованість теоретичних положень; адекватність при вирішенні багатьох завдань теорії та практики оперування з рядами динаміки різної природи.

Процес побудови моделі складається з декількох основних стадій [6, 12, 14, 21, 24, 28]. Спочатку визначають тип і клас моделі – проводять структурну ідентифікацію, вивчаючи вимоги і можливості застосування стосовно конкретного випадку. Далі уточнюють перелік пояснюючих змінних, форми зв'язку між ними і набір параметрів, які підлягають оцінюванню. На наступному етапі суто математичними засобами досліджують загальні властивості моделі: існування та кількість розв'язків, їх стійкість, а також межі та тенденції зміни основних величин і параметрів. Практичне використання моделі передбачає відповідну підготовку вхідної інформації: агрегування та групування даних, оцінювання вірогідності даних та їх корегування тощо. Етап комп'ютерної реалізації передбачає розробку алгоритмів числового розв'язування задачі, підготовку програм і безпосереднє виконання розрахунків [2, 6, 15]. Комп'ютерні обчислювальні експерименти істотно доповнюють результати аналітичних досліджень. При цьому уточнюється питання щодо правильності та повноти результатів моделювання, їх предметної інтерпретації, аналізуються можливості та виробляються рекомендації щодо практичного використання. Також окреслюються перспективні напрямки подальшого вдосконалення моделі та розширення сфер її застосування.

Практично значиме відображення часового ряду може ґрунтуватись лише на правдоподібній моделі. Основною характеристикою точності моделювання слугує середня відносна оцінка відхилень модельних значень результуючої змінної від даних її спостереження. Адекватність математичного подання визначається величиною і характером поведінки випадкової складової – залишків моделі як різниці модельних та експериментальних значень залежної змінної. Аналіз залишків проводиться з метою перевірки якості підгонки моделі до даних спостережень. Залишки (похибки, відхилення, збурення) – це різниця між фактичними значеннями залежної змінної та її відповідними значеннями, обчисленими за моделлю. Якщо модель підігнана добре, ряд залишків має нормальний розподіл з нульовим математичним сподіванням, сталою дисперсією і відсутністю кореляції (білий шум) [12, 14, 21, 24, 28].

Для графічної перевірки нормальності розподілу відхилень можна побудувати гістограму залишків: якщо залишки мають нормальний розподіл, то гістограма

повинна добре «лягати» на теоретичну криву розподілу Гауса з відповідними параметрами положення і розсіювання.

Коефіцієнт детермінації – це частка загальної дисперсії результуючої змінної, що пояснюється моделлю. Коефіцієнт детермінації, який є квадратом множинного коефіцієнта кореляції, відображає ступінь придатності моделі. Його значення коливається між 0 і 1. Чим ближче до 1, тим більша відповідність моделі реальним даним. І навпаки, чим ближче до 0, тим менш надійно і повно модель відображає взаємозв'язки у масиві вхідних даних [21, 25]. Більш точно значущість коефіцієнта детермінації визначається за статистикою Фішера [21].

При знаходженні оптимальних оцінок параметрів моделі значні проблеми породжує мультиколінеарність масиву пояснюючих змінних. Мультиколінеарність – це сильна лінійна залежність між більш ніж двома предикторами при множинній регресії. Мультиколінеарність спостерігається, коли між деякими предикторами існує суттєва лінійна залежність [6, 12, 14, 21, 24, 28]. Для визначення ступеня значущості оцінок коефіцієнтів моделі використовується статистика Стюдента.

Неспівмірність за масштабом і різний ступінь мінливості пояснюючих змінних не дозволяє безпосередньо порівнювати між собою значення коефіцієнтів моделі. Для усунення цих недоліків при порівнянні предикторів за їхнім впливом на залежну змінну використовують коефіцієнти моделі у стандартизованому вигляді (бета-коефіцієнти) [14, 24].

Застосування апроксимаційних моделей часового ряду для екстраполяції за межами часового діапазону наявних спостережень у більшості випадків є мало обґрунтованим, оскільки такі моделі безпосередньо не відображають причинно-наслідкові зв'язки. На практиці за регресійною моделлю можна будувати лише короткотерміновий прогноз. Для покращення прогнозних властивостей моделі рекомендується для параметричної ідентифікації використовувати узагальнений МНК, де останнім за часом спостереженням надається більша вага.

Створення моделей багатовимірних еколого-економічних часових рядів відкриває можливість аналізувати поточний стан і прогнозні характеристики відповідної складної системи, досліджувати варіанти реакції системи на зовнішні

впливи і управлінські рішення [1, 2, 11, 28]. Для побудови залежності результуючого екологічного показника $y(t)$, де t – час, від декількох пояснюючих змінних пропонується [17] використовувати гладкі двічі диференційовані функції $f(\cdot)$, де аргументами служать комплексні (інтегровані) соціально-економічні показники трьох типів: $x_1(t)$ – фактор, який відображає економічний розвиток і зазвичай провокує негативні наслідки на довкілля (такими показниками виступають індекс промислового виробництва, валовий внутрішній продукт, кінцеве енергоспоживання тощо); $x_2(t)$ – фактор, який відображає цілеспрямовану діяльність щодо підвищення якості навколишнього середовища і зазвичай забезпечує позитивний вплив на навколишнє середовище (до таких показників відносяться витрати на охорону атмосферного повітря і проблеми зміни клімату, інвестиції в природоохоронні проекти тощо); $x_3(t)$ – фактор, що віддзеркалює структурні зміни в економіці та зазвичай позитивно впливає на довкілля (такими показниками можна вважати долю альтернативної енергетики, інвестиції в інноваційну модернізацію виробництва тощо). При цьому еколого-економічні міркування дозволяють висунути припущення: числові значення предикторних змінних невід’ємні $x_1(t) \geq 0$ і $x_2(t) \geq 0$, а збільшення величини кожного пояснюючого фактору призводить до зниження ступеня його впливу на результуючий показник [17], тобто моделююча функція $f(\cdot)$ відповідає процесам з насиченням.

Комплексним дослідженням взаємозв’язків екологічних і соціально-економічних показників з урахуванням регіональної та національної специфіки присвячені публікації [13, 18], з яких випливає важливість імплементації міжнародно визнаних індикаторів і механізмів у практику природоохоронної діяльності та створення доступних для розуміння і практичного використання моделей, які описують взаємозв’язок соціально-економічних і екологічних показників.

Узагальнюючи проведені дослідження підходів до статистичного аналізу та моделювання взаємозв’язаних часових рядів і спираючись на експертні оцінки і наявний масив спостережень статистичних показників системи «суспільство – економіка – довкілля», далі буде проведено статистичний аналіз рядів динаміки

результуючого екологічного показника, що характеризує забруднюючі викиди в атмосферне повітря в Україні, і вибраних фізичних і соціально-економічних показників, які гіпотетично прямо чи опосередковано впливають на результуючу змінну, та буде синтезовано лінійну регресійну модель, яка відображає взаємозв'язки між цими величинами.

Висновки до першого розділу

У розділі досліджено сучасні напрямки статистичного аналізу часових рядів, їхні основні усереднені показники положення, інтенсивності, варіації та взаємозв'язку; висвітлено підходи до вибору класу і структури математичних моделей ряду динаміки як реалізації випадкового процесу з дискретним часом; вказано на метод найменших квадратів як основний інструмент знаходження оптимальних оцінок параметрів моделі часового ряду та його модифікації; висвітлено апарат дослідження адекватності моделі та значущості її коефіцієнтів і рекомендації щодо покращення прогнозу за моделлю.

Розглянуто специфічні підходи до моделювання економетричних часових рядів. Сформульовано задачі статистичного аналізу вибраних показників динаміки системи «суспільство – економіка – довкілля» та синтезу лінійної регресійної моделі, яка відображає в розрізі часової динаміки вплив соціально-економічних показників розвитку нашої країни на викиди забруднюючих речовин в атмосферне повітря в Україні.

РОЗДІЛ 2

СТАТИСТИЧНІ ХАРАКТЕРИСТИКИ ДИНАМІКИ ЗАБРУДНЮЮЧИХ ВИКИДІВ І ГІПОТЕТИЧНО ПОВ'ЯЗАНИХ СОЦІАЛЬНО-ЕКОНОМІЧНИХ ПОКАЗНИКІВ

2.1. Показники положення, варіації та внутрішніх взаємозв'язків між рівнями часового ряду інтенсивності викидів

Аналіз первинної статистичної інформації від Державної служби статистики України щодо обсягів забруднюючих викидів у повітря в Україні за період 2001 – 2021 рр. [19] показує суттєву відмінність даних за 2014 – 2021 рр. від попереднього часового проміжку, оскільки вони не включають звітність за тимчасово окуповані території. Спираючись на думки експертів щодо характеру розподілу та обсягів забруднюючих викидів на цих землях та їх наближену долю у загальнонаціональному масштабі, пропонується за результуючий екологічний показник, як базу дослідження, що більш стійка до територіальних змін, взяти інтенсивність викидів забруднюючих речовин у повітря в Україні на одиницю площі території, т / кв. км. Значення інтенсивності викидів знаходиться шляхом ділення обсягу викидів на площу території збору статистичної інформації та вимірюється в т / кв. км. Обчислені автором дані щодо інтенсивності викидів за 2001 – 2021 рр. відображені в табл. А.1 (дод. А) та розглядаються як інтервальний часовий ряд:

$$y_e(1), y_e(2), \dots, y_e(t), \dots, y_e(T-1), y_e(T),$$

де $y_e(t)$ – реальна інтенсивність викидів, що відповідає дискретному часу t , $t = \overline{1, T}$; T – кінцевий момент дискретного часу спостережень, $T = 21$.

Для візуально-графічного аналізу інформації статистичні таблиці даних доповнюються графіками [6, 12, 14, 21, 28]. При дослідженні динаміки одного показника досить використати одновимірну лінійну діаграму як на рис. 2.1, де відображено побудований у табличному процесорі MS Excel згладжений графік зміни в часі інтенсивності загальних щорічних забруднюючих викидів згідно табл. А.1 (дод. А).

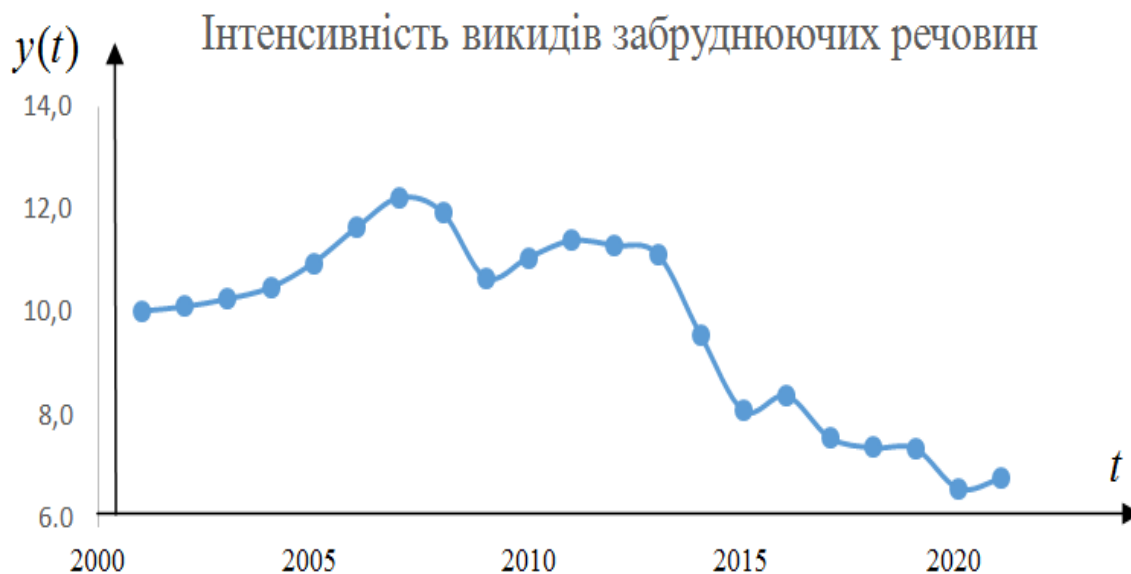


Рис. 2.1 Інтенсивність загальних щорічних забруднюючих викидів (спостереження)

Аналіз графіка демонструє лінійний тренд з помітними коливаннями різної частоти, що відповідає наявності економічних циклів різного періоду [2, 14, 21], та показує дуже повільне поступове зниження рівня цього показника відповідно до загальносвітової тенденції руху до покращення екологічної ситуації. При цьому стохастичні збурення рівнів часового ряду викидів візуально здаються незначними.

Часовий ряд є реалізацією випадкового процесу з дискретним цілочисельним параметром t . При вивченні масових стохастичних явищ значну роль відіграють їх середні характеристики, що показують положення та інтенсивність окремих проявів, і показники варіації, що відображають ступінь відмінності чи близькості індивідуальних особливостей окремих реалізацій [1, 2, 6, 11, 21]. Результати розрахунків у системі MS Excel середніх величин і показників варіації (мінливості чи стійкості) часового ряду інтенсивності забруднюючих викидів наведені в табл. 2.1.

Стійкість положення рівнів ряду інтенсивності забруднюючих викидів підтверджують близькі між собою значення двох оцінок математичного сподівання – середнього арифметичного значення $y_{ср}$ і медіани Me . Значну ступінь мінливості ряду викидів демонструють відносно великі величини дисперсії s^2_e , коефіцієнта варіації v_e та інших показників розкиду рівнів інтенсивності викидів.

Значення показників положення, інтенсивності та варіації часового ряду інтенсивності викидів забруднюючих речовин в атмосферне повітря України за 2001 – 2021 рр.

Показник	Значення	Показник	Значення
Середнє значення $y_{ср}$	9,75	Максимальне значення y_{max}	12,2
Медіана Me	1,03	Розмах варіації R_{ve}	5,66
Оцінка дисперсії s_e^2	3,33	Коефіцієнт варіації v_e	18,7 %
Середнє квадратичне відхилення s_e	1,82	Середній абсолютний приріст $\Delta_{у,ср}$	-0,163
Середнє лінійне відхилення d_e	1,60	Середній темп приросту $T_{лср}$	-1,9 %
Мінімальне значення y_{min}	6,57	Розмах темпу приросту PR_l	15,1 %

Для дослідження характеру мінливості рівнів часового ряду та тісноти взаємозв'язків між ними використовується велике розмаїття показників. При цьому можуть спостерігатись стійкий тренд і різні типи коливань часового ряду: короткоперіодичні маятникові, довгоперіодичні циклічні, випадково розподілені, детерміновано хаотичні. Результати розрахунків у середовищі MS Excel рангових показників [1, 2, 6, 11, 12] тісноти взаємозв'язків між рівнями часового ряду інтенсивності викидів і відповідно критичне значення критерію Стьюдента для обґрунтування їх значущості наведені в табл. 2.2. За рівень значущості показників прийнято $\alpha = 0,05$.

Таблиця 2.2

Значення рангових показників тісноти взаємозв'язків між рівнями часового ряду інтенсивності викидів і відповідні значення критерію Стьюдента ($\alpha = 0,05$)

Показник	Значення
Коефіцієнт рангової кореляції Спірмена ρ	0,88
Коефіцієнт рангової кореляції Фехнера K_ϕ	0,90
Критичне значення критерія Стьюдента $t_{кр}$	2,09
Фактичне значення критерія Стьюдента для коефіцієнта Спірмена t_ϕ	8,13
Фактичне значення критерія Стьюдента для коефіцієнта Фехнера t_ϕ	9,00

Статистично значимі ($t_{\phi} > t_{кр}$) обчислені величини рангових коефіцієнтів кореляції між зміщеними на один часовий лаг рівнями ряду динаміки викидів підтверджують наявність типової для часових рядів автокореляції, яка характеризує не тільки взаємозалежність рівнів одного і того ж ряду, що відповідають різним моментам спостережень, а й ступінь сталості розвитку процесу у часі.

Для більш детального вивчення автокореляції використовується автокореляційна функція як послідовність коефіцієнтів автокореляції рівнів першого, другого, третього та наступних порядків [1, 2, 6, 11, 14]. Результати розрахунків у системі MS Excel коефіцієнтів автокореляції r_k , $k = \overline{1, 20}$ між зміщеними на k у часі рівнями ряду динаміки інтенсивності викидів наведені в табл. А.2 (дод. А). За діапазон значущості коефіцієнтів автокореляції приймається $(r_n; r_e)$, де $r_n = -1/\sqrt{T} = -1/\sqrt{21}$ і $r_e = 1/\sqrt{T} = 1/\sqrt{21}$ – відповідно його нижня і верхня межі.

Аналіз корелограми, побудованої за табл. А.2 (дод. А) і поданої на рис. 2.2, дозволяє висунути гіпотезу про наявність у часовому ряді інтенсивності викидів лінійного тренду та періодичних коливань, що відповідають циклам Жюгляра і Кузнеця як вищим гармонікам Кондратьєвської хвилі [2].

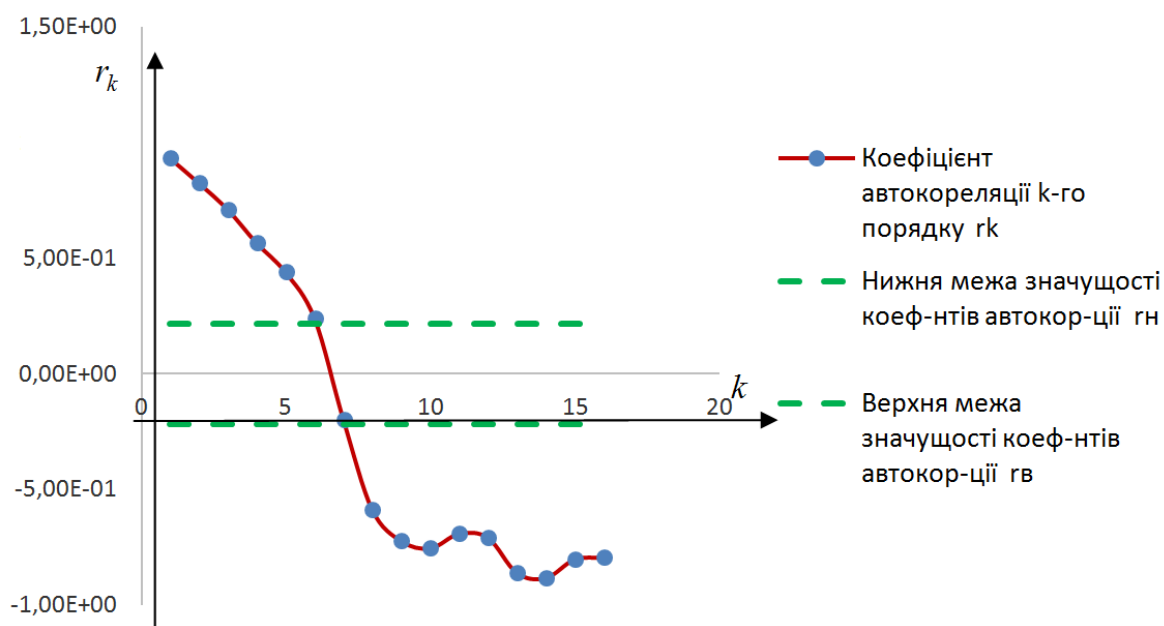


Рис.2.2 Корелограма ряду інтенсивності викидів (спостереження)

2.2. Дослідження тісноти взаємозв'язків між інтенсивністю забруднюючих викидів і гіпотетично пояснюючими фізичними та соціально-економічними показниками

Конструктивне дослідження динаміки системи «суспільство – економіка – довкілля» передбачає побудову еколого-економетричної моделі, в якій результуючою змінною $y(t)$ вибрана інтенсивність викидів забруднюючих речовин в атмосферне повітря в Україні, а для формування оптимального набору пояснюючих змінних на першому етапі необхідно, спираючись на експертні оцінки, статистично проаналізувати широкий масив фізичних і соціально-економічних показників як гіпотетичних кандидатів.

Проведене попереднє вивчення предметної області та джерел вільно доступної статистичної інформації [3–5, 7–10, 19, 20, 22, 23, 26, 27] і використання думки експертів дозволяють звужити коло претендентів на роль пояснюючих змінних щодо інтенсивності забруднюючих викидів. Як вхідна зведена інформація для обробки, у табл. А.3 і А.4 (дод. А) відображені статистичні дані щодо попередньо відібраних фізичних і соціально-економічних показників.

При складанні табл. А.3 і А.4 (дод. А) здійснено певну обробку початкових статистичних даних:

– «Чисельність населення $x_1(t)$, млн. осіб» скоригована на основі додаткової інформації про число жителів на неконтрольованих територіях. Зокрема, враховано оцінки ООН щодо чисельності населення ОРДЛО.

– «Індекс промислової продукції $x_2(t)$, % до 1990 р.» одержано автором зі звітних статистичних даних перерахуванням у відсотки до рівня 1990 р., що надає більш повну картину стану промислового виробництва, якщо порівнювати його з останнім роком, коли Україна була у складі СРСР. Більшість гігантів промисловості в Україні є застарілими і без суттєвих оновлень з часів існування СРСР, а тому виробництво є технологічно відсталим і екологічно ризикованим.

– «Валовий внутрішній продукт (ВВП) на душу населення за паритетом купівельної спроможності (ПКС) $x_3(t)$ у доларах США» є найбільш об'єктивною

оцінкою добробуту громадян і якісного рівня економіки країни, порівняно з іншими показниками щодо ВВП.

– Показник «кінцеве енергоспоживання на душу населення $x_4(t)$, тне» сформовано автором шляхом ділення обсягу кінцевого енергоспоживання на кількість населення. Його використання підвищує об'єктивність відображення рівня енергоефективності економіки країни.

– Показник «частка відновлюваних джерел енергії у загальному постачанні первинної енергії $x_5(t)$, % від загального постачання» при перспективних дослідженнях може бути доповнений включенням енергії від атомних електростанцій. Він відображає технологічний стан розвитку економіки в енергетичному аспекті, а його динаміка показує зміну рівня уваги в суспільстві до екологічних проблем.

– Показник «накопичені прямі іноземні інвестиції $x_6(t)$, млн. дол. США» сформовано автором шляхом підсумовування прямих іноземних інвестицій від першого (2001 р.) до поточного року спостережень. Він відображає кумулятивний ефект впливу капітальних інвестицій та технологічну затримку у часі реакції на них економіки. Прямі іноземні інвестиції – найпривабливіша форма капіталовкладень, які дозволяють реалізовувати великі проекти і впроваджувати нові технології та практики корпоративного управління, що змінює вплив на довкілля.

– «Частка валових приватних інвестиційних витрат у структурі ВВП $x_7(t)$, %» визначає відсоток коштів, які спрямовуються на розвиток і розширення виробництва та його матеріально-технічної бази. Чим більша ця частка у структурі ВВП, тим вищими є темпи розвитку економіки країни.

– Показник «капітальні інвестиції на охорону атмосферного повітря і проблеми зміни клімату $x_8(t)$, млн. дол. США» сформовано автором шляхом перерахунку звітних даних у фактичних цінах у тис. грн. для усунення впливу на показник змін курсу гривні. На жаль, у статистичних звітах відсутні значення цього показника за період 2001 – 2005 рр. і за 2021 р.

– «Поточні витрати на охорону атмосферного повітря і проблеми зміни клімату $x_9(t)$, млн. дол. США» сформовано автором шляхом перерахунку звітних даних у фактичних цінах у тис. грн. для усунення впливу на показник змін курсу гривні. На жаль, у статистичних звітах відсутні значення цього показника за 2021 р.

– Показник «Середня температура повітря за рік (м. Київ) $x_{10}(t)$, °C» включено до масиву предикторів, що гіпотетично впливають на результуючу екологічну змінну, для відображення дії фізичних факторів зовнішнього середовища. Використано дані щодо температури лише у столиці, оскільки для інших регіонів відсутня доступна інформація.

Для дослідження тісноти взаємозв'язків між інтенсивністю забруднюючих викидів, як результуючою змінною $y(t)$, і гіпотетично пояснюючими фізичними та соціально-економічними показниками x_i , $i = \overline{1,10}$ застосовується інструментарій кореляційного аналізу.

У табл. 2.3 відображені парні коефіцієнти кореляції r_{xy} між вихідною змінною $y(t)$ і пояснюючими змінними x_i , $i = \overline{1,10}$ та якісна оцінка тісноти взаємозв'язку за шкалою Чеддока. Значення парних коефіцієнтів кореляції демонструють тісний статистичний зв'язок. Однак спільна економетрична природа усіх показників, окрім останнього x_{10} , породжує гіпотезу про наявність мультиколінеарності всередині усього масиву змінних величин, що розглядаються.

Проведення аналізу на мультиколінеарність ускладнюється різною кількістю спостережень окремих змінних. Для усіх показників маємо лише $T = 15$ спільних за часом спостережень у період 2006 – 2020 рр., що при значному числі $m = 10$ предикторних змінних у випадку мультиколінеарності породжує обчислювальні проблеми, які призводять до нестійкості отриманих результатів. Тому висновки з аналізу мультиколінеарності при малому числі спостережень можна розглядати лише як рекомендації для подальших досліджень.

Парні коефіцієнти кореляції r_{xy} між вихідною змінною $y(t)$ і пояснюючими змінними x_i , $i = \overline{1,10}$ та якісна оцінка тісноти взаємозв'язку за шкалою Чеддока

№ з/п	Предикторна змінна	Кількість спостережень	Коефіцієнт кореляції r_{xy}	Ступінь зв'язку
1	Чисельність населення $x_1(t)$, млн. осіб	21	0,81	тісний
2	Індекс промислової продукції $x_2(t)$, % до 1990 р.	21	0,92	дуже тісний
3	ВВП на душу населення за ПКС $x_3(t)$, дол. США	21	-0,66	помітний
4	Кінцеве енергоспоживання на душу населення $x_4(t)$, тне	21	0,95	дуже тісний
5	Частка постачання енергії від відновлюваних джерел $x_5(t)$, %	21	-0,85	тісний
6	Накопичені прямі іноземні інвестиції $x_6(t)$, млн. дол. США	21	-0,65	помітний
7	Частка приватних інвестиційних витрат у ВВП $x_7(t)$, %	21	0,66	помітний
8	Капітальні інвестиції на охорону повітря $x_8(t)$, млн дол.США	15	0,63	помітний
9	Поточні витрати на охорону повітря $x_9(t)$, млн. дол. США	20	0,73	тісний
10	Середня температура повітря за рік $x_{10}(t)$, °С	21	-0,54	помітний

Згідно табл. 2.3 найбільше значення, що близьке до одиниці, має частинний коефіцієнт кореляції результуючого показника $y(t)$ з предиктором $x_4(t)$ (кінцеве енергоспоживання на душу населення). Далі це буде використано для побудови простої однофакторної моделі для експрес-розрахунків інтенсивності викидів.

Дослідження мультиколінеарності масиву пояснюючих змінних x_i , $i = \overline{1,10}$ здійснюється за алгоритмом Фаррара – Глобера при довірчій ймовірності $P = 95\%$ і, відповідно, рівні значущості $\alpha = 1 - P = 5\%$.

Критичне значення критерія Пірсона $\chi^2_{кр} = 61,7$, а фактичне значення критерія Пірсона $\chi^2_{ф} = 192,5$. Оскільки $\chi^2_{ф} > \chi^2_{кр}$, то за критерієм хі-квадрат Пірсона у масиві пояснюючих змінних існує мультиколінеарність.

У табл. А.5 (дод. А) при критичному значенні критерія Фішера $F_{кр} = 4,8$ для кожного пояснюючого показника наведено фактичне значення критерія Фішера $F_{ф}$ і встановлено наявність ($M = 1$ при $F_{ф} > F_{кр}$) чи відсутність ($M = 0$ при $F_{ф} < F_{кр}$) мультиколінеарності з іншими предикторними змінними.

Аналіз табл. А.5 (дод. А) показує, що лише для двох пояснюючих змінних x_8 і x_{10} за критерієм Фішера не підтверджується наявність мультиколінеарності.

У табл. А.6 (дод. А) відображені частинні коефіцієнти кореляції $r_{ij}^{(p)}$ масиву пояснюючих змінних x_i , $i = \overline{1,10}$, а в табл. А.7 (дод. А) наведено відповідні фактичні значення критерія Стюдента t_Φ для частинних коефіцієнтів кореляції, значення яких порівнюються з критичним значенням критерію Стюдента $t_{кр} = 2,57$.

Далі з порівняння випливає висновок про наявність ($M = 1$ при $t_\Phi > t_{кр}$) чи відсутність ($M = 0$ при $t_\Phi < t_{кр}$) мультиколінеарності, що показано в табл. А.8 (дод. А). У табл. А.9 (дод. А) відображені частинні коефіцієнти кореляції $r_{yx}^{(p)}$ пояснюючих змінних $x_i(t)$, $i = \overline{1,10}$ з інтенсивністю викидів $y(t)$ як результуючим показником.

Враховуючи високий ступінь мультиколінеарності масиву пояснюючих змінних $x_i(t)$, $i = \overline{1,10}$, можна вилучити з розгляду деякі предиктори. Спочатку можна вилучити змінні $x_8(t)$ (капітальні інвестиції на охорону атмосферного повітря і проблеми зміни клімату) і $x_9(t)$ (поточні витрати на охорону атмосферного повітря і проблеми зміни клімату), що дозволить розширити діапазон наявних спостережень до максимального значення $T = 21$. Далі можна провести аналогічне дослідження мультиколінеарності звуженого масиву пояснюючих змінних на розширеному діапазоні спостережень за алгоритмом Фаррара – Глобера.

Загальновідомо про відсутність упродовж багатьох років перепису населення в Україні, отже відповідні звітні статистичні дані насправді самі є деякими статистичними оцінками, а не результатами безпосередніх вимірювань. Тому показник $x_1(t)$ (чисельність населення), який пов'язаний з іншими згідно проведеного кореляційного аналізу, також можна видалити з розгляду. Беручи до уваги думку експертів, наявність мультиколінеарності предикторів і відносно малі значення парного і частинного коефіцієнтів кореляції з результуючим показником, можна вилучити з розгляду змінні $x_3(t)$ (ВВП на душу населення за ПКС) і $x_{10}(t)$ (середня температура повітря за рік). Далі здійснено перепозначення залишених пояснюючих змінних і проведено дослідження мультиколінеарності нового масиву предикторів $x_i(t)$, $i = \overline{1,5}$ за алгоритмом Фаррара – Глобера. Результати відображені у

табл. А.10 – А.15 (дод. А). При цьому критичне значення критерія Пірсона $\chi^2_{кр}=18,3$, критичне значення критерія Фішера $F_{кр}=3,0$, а критичне значення критерія Стюдента $t_{кр}= 2,12$.

Оскільки фактичне значення критерію Пірсона $\chi^2_{ф} =106,3$ перевищує його критичне значення $\chi^2_{кр}=18,3$, то за критерієм хі-квадрат Пірсона у масиві пояснюючих змінних $x_i(t)$, $i = \overline{1,5}$ існує мультиколінеарність. Згідно табл. А.11 (дод. А) за критерієм Фішера підтверджується наявність мультиколінеарності для кожного предиктора. Згідно табл. А.14 (дод. А) предиктор $x_4(t)$ (накопичені прямі іноземні інвестиції), за критерієм Стюдента має кореляційний зв'язок з усіма іншими пояснюючими змінними, а тому надалі буде видалений з розгляду. Можна звернути увагу, що згідно табл. А.15 (дод. А) найбільше значення має частинний коефіцієнт кореляції результуючого показника $y(t)$ з предиктором $x_3(t)$ (частка відновлюваних джерел енергії у загальному постачанні первинної енергії).

Далі здійснено перепозначення залишених пояснюючих змінних і проведено дослідження мультиколінеарності нового звуженого масиву предикторів $x_i(t)$, $i = \overline{1,4}$ за алгоритмом Фаррара – Глобера. Результати відображені у табл. А.16 – А.21 (дод. А). При цьому критичне значення критерію Пірсона $\chi^2_{кр}=12,6$, критичне значення критерія Фішера $F_{кр}=3,2$, а критичне значення критерія Стюдента $t_{кр}= 2,11$.

Оскільки фактичне значення критерію Пірсона $\chi^2_{ф} =72,5$ перевищує його критичне значення $\chi^2_{кр}=12,6$, то за критерієм хі-квадрат Пірсона у масиві пояснюючих змінних $x_i(t)$, $i = \overline{1,4}$ існує мультиколінеарність.

Згідно табл. А.17 (дод. А) за критерієм Фішера підтверджується наявність мультиколінеарності для кожного предиктора. Згідно табл. А.20 (дод. А) змінні $x_1(t)$ (індекс промислової продукції) і $x_2(t)$ (кінцеве енергоспоживання на душу населення), за критерієм Стюдента мають кореляційний зв'язок між собою і з предиктором $x_3(t)$ (частка відновлюваних джерел енергії у загальному постачанні первинної енергії), тому один з показників $x_1(t)$ і $x_2(t)$ є кандидатом на видалення з розгляду. При цьому пріоритет має змінна $x_2(t)$, оскільки згідно табл. А.21 (дод. А) вона має менше

значення частинного коефіцієнта кореляції з результируючим показником, ніж предиктор $x_2(t)$. Далі це буде використано для побудови регресійної моделі з оптимальним набором предикторів. Відповідні цілеспрямовані дослідження різних підмножин звуженого масиву пояснюючих змінних висвітлені у наступному розділі.

Висновки до другого розділу

Вибрано основні статистичні показники для опису положення, інтенсивності та варіації часових рядів викидів забруднюючих речовин в атмосферне повітря в Україні та пов'язаних фізичних і соціально-економічних змінних. Розроблено обчислювальні процедури для розрахунків основних статистичних показників і проведено їх обчислення засобами табличного процесора MS Excel.

Запропоновано комбіновані показники – «інтенсивність забруднюючих викидів», «кінцеве енергоспоживання на душу населення» і «накопичені прямі іноземні інвестиції», які краще відповідають структурі звітних статистичних даних і ефективніше віддзеркалюють кореляційні зв'язки.

Досліджено тісноту взаємозв'язків між рівнями часового ряду інтенсивності забруднюючих викидів на базі обчислення рангових коефіцієнтів кореляції Спірмена і Фехнера та встановлення їх значущості за критерієм Стьюдента.

Вивчено автокореляційну функцію часового ряду інтенсивності викидів забруднюючих речовин. Спираючись на загальні закономірності економічних процесів і аналіз корелограми, обґрунтовано припущення про наявність у часовому ряді викидів лінійного тренду і коливань різної частоти та випадкової компоненти.

Сформовано початковий набір пояснюючих фізичних і соціально-економічних змінних, що гіпотетично впливають на динаміку інтенсивності забруднюючих викидів. Проведено кореляційний аналіз вхідного масиву предикторів та їх зв'язків з результируючим показником – інтенсивністю забруднюючих викидів. Підтверджено наявність високого ступеня мультиколінеарності. Відкиданням «проблемних» змінних досягнуто зниження рівня мультиколінеарності звуженого масиву пояснюючих змінних. Намічено шляхи побудови простої однофакторної регресійної моделі для експрес-розрахунків і більш досконалої багатфакторної регресійної моделі з оптимальним набором предикторів.

РОЗДІЛ 3

МОДЕЛЮВАННЯ ДИНАМІКИ ІНТЕНСИВНОСТІ ЗАБРУДНЮЮЧИХ ВИКИДІВ У ПОВІТРЯ В УКРАЇНІ

3.1. Однофакторна модель для експрес-розрахунків інтенсивності забруднюючих викидів

Процес утворення та викидання забруднюючих речовин в атмосферне повітря відбувається на різних об'єктах народно-господарчого комплексу України і його показники та загальні тенденції відображають стан і напрямок розвитку як нашої країни, так і Світу в цілому [13, 17]. Заходи щодо послаблення негативного впливу на навколишнє середовище спрямовані на поступове зменшення забруднюючих викидів, що дозволяє очікувати слабкий нейтральний екологічний прогрес $a(t)$ як функцію часу t , який відображає вплив структурних зрушень і технологічної модернізації [2].

Виявлений при статистичному аналізі дуже тісний кореляційний зв'язок результуючого показника – інтенсивності забруднюючих викидів – з кінцевим енергоспоживанням на душу населення як предиктором дозволяє синтезувати просту однофакторну регресійну модель для експрес-розрахунків у двох формах – адитивній та мультиплікативній.

Найпростіша лінійна, як за параметрами, так і за фактором, регресійна модель (в адитивній формі) інтенсивності забруднюючих викидів $y(t)$ має вигляд:

$$y(t) = \theta_0 + \theta_1 \cdot x(t) + \varepsilon(t), \quad t = 1, 2, \dots, \quad (3.1)$$

де $x(t)$ – значення показника «кінцеве енергоспоживання на душу населення»; θ_0 і θ_1 – параметри, що підлягають оцінюванню; $\varepsilon(t)$ – похибка (залишок, відхилення, нев'язка) моделі в момент часу t , що розглядається як випадкова величина, яка відображає вплив на інтенсивність забруднюючих викидів непередбачуваних і неконтрольованих факторів різноманітного генезису.

В умовах обмеженого об'єму наявних спостережень і відсутності теоретико-ймовірнісних обґрунтувань основним критерієм правдоподібності результатів

моделювання слугує їх відповідність реальним даним. Знайдені за методом найменших квадратів (МНК) оцінки коефіцієнтів $\theta_0 = -0,9905$ і $\theta_1 = 7,246$ забезпечують для адитивної моделі (3.1) середню відносну похибку $\delta_0 = 4,9\%$, що цілком задовільно при експрес-розрахунках

На рис. 3.1 подані суміщені згладжені графіки інтенсивності викидів забруднюючих речовин в атмосферне повітря в Україні за даними спостережень і за моделлю (3.1), що демонструють прийнятний для оціночних розрахунків ступінь апроксимації упродовж всього часового проміжку, що досліджується.

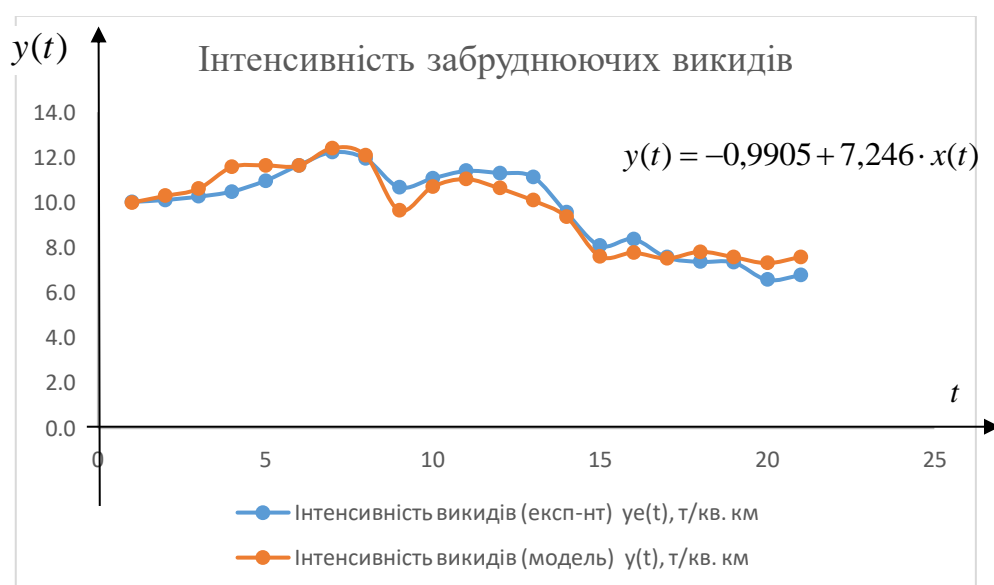


Рис. 3.1 Інтенсивність забруднюючих викидів
(однофакторна адитивна модель і спостереження)

Адитивна модель (3.1) характеризується досить високими значеннями коефіцієнта детермінації $R^2 = 0,898$ і нормованого коефіцієнта детермінації $R_n^2 = 0,893$, однак суттєвим недоліком є автокореляція залишків, що вказує на необхідність включення додаткових пояснюючих змінних чи фактору часу.

Нелінійна, як за параметрами, так і за фактором, регресійна модель (у мультиплікативній формі) інтенсивності забруднюючих викидів $y(t)$ має вигляд:

$$y(t) = \theta_0 \cdot x^{\theta_1}(t) + \varepsilon(t), \quad t = 1, 2, \dots, \quad (3.2)$$

де позначення відповідають моделі (3.1).

Мультиплікативна модель допускає лінеаризацію за параметрами за допомогою логарифмування обох частин співвідношення (3.2), що дозволяє знайти «гарні» початкові оцінки коефіцієнтів θ_0 і θ_1 для наступного ітераційного уточнення.

Знайдені МНК-оцінки коефіцієнтів $\theta_0 = 6,342$ і $\theta_1 = 1,090$ забезпечують для мультиплікативної моделі (3.2) середню відносну похибку $\delta_0 = 5,0\%$, що дещо гірше адитивної моделі (3.1), але цілком задовільно при експрес-розрахунках.

На рис. 3.2 подані суміщені згладжені графіки інтенсивності викидів забруднюючих речовин в атмосферне повітря в Україні за даними спостережень і за мультиплікативною моделлю (3.2), що демонструють прийнятний для оціночних розрахунків ступінь апроксимації упродовж всього часу дослідження.

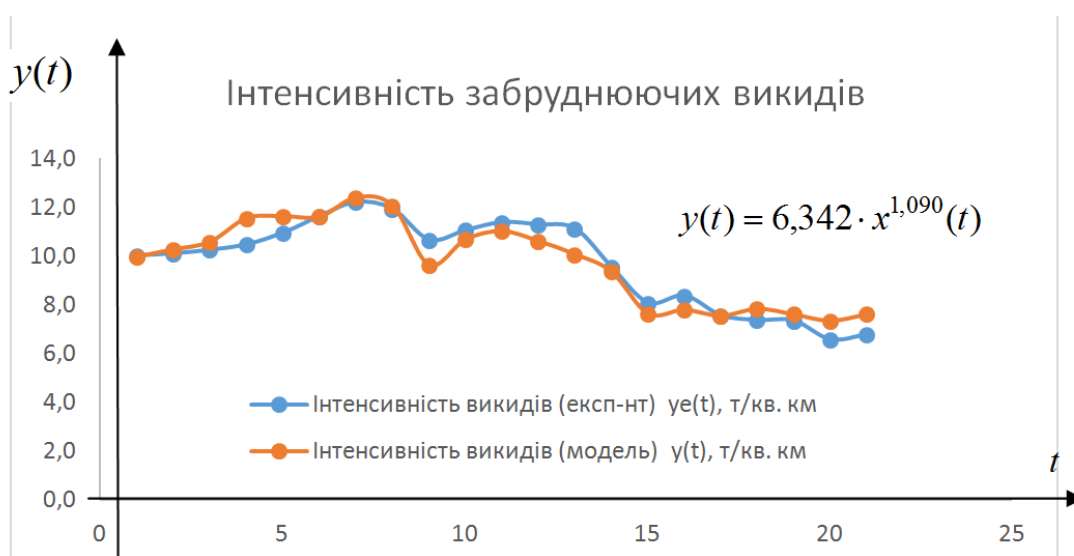


Рис. 3.2 Інтенсивність забруднюючих викидів
(однофакторна мультиплікативна модель і спостереження)

Мультиплікативна модель (3.2) має близькі до моделі (3.1) і досить високі значення коефіцієнта детермінації $R^2 = 0,896$ і нормованого коефіцієнта детермінації $R_n^2 = 0,891$, але також не позбавлена того самого недоліку – автокореляція залишків.

3.2. Багатофакторна модель інтенсивності забруднюючих викидів

Використання багатофакторних моделей дозволяє враховувати різну спрямованість впливу складових економічного розвитку на екологічні показники, зокрема наявність економічних чинників, пов'язаних зі зміною економічної структури, які позитивно впливають на екологію – впровадження інноваційних еколого-безпечних технологій, модернізація застарілого виробництва тощо.

У рамках множинної лінійної регресії

$$y(t) = \theta_0 + \theta_1 \cdot x_1(t) + \theta_2 \cdot x_2(t) + \dots + \theta_m \cdot x_m(t) + \varepsilon(t), \quad t = 1, 2, \dots, \quad (3.3)$$

потрібно здійснити оптимальний щодо якості моделювання вибір числа m і складу $x_i(t)$, $i = \overline{1, m}$ пояснюючих змінних і оцінок параметрів θ_i , $i = \overline{0, m}$, виходячи з наявного масиву предикторів-претендентів, що звужений за результатами досліджень у попередньому розділі до чотирьох соціально-економічних показників: $x_1(t)$ – «індекс промислової продукції, % до 1990 р.», $x_2(t)$ – «кінцеве енергоспоживання на душу населення, тне», $x_3(t)$ – «частка відновлюваних джерел енергії у загальному постачанні первинної енергії, % від загального постачання», $x_4(t)$ – «частка валових приватних інвестиційних витрат у структурі ВВП, %».

Векторна оптимізація структури моделі (3.3) передбачає максимізацію нормованого коефіцієнта детермінації та ступеня значущості за критерієм Стюдента усіх параметрів регресії при виконанні умови нормального розподілу залишків моделі [12, 14, 21, 24, 28]. Параметрична ідентифікація регресії (3.3), враховуючи мале число спостережень і відсутність обґрунтованих теоретико-ймовірнісних передумов та наявність мультиколінеарності, здійснюється на основі узагальненого МНК з регуляризацією. Проведені порівняльні обчислювальні експерименти дозволяють рекомендувати до використання наступну трифакторну лінійну регресійну модель:

$$y(t) = \theta_0 + \theta_1 \cdot x_1(t) + \theta_2 \cdot x_2(t) + \theta_3 \cdot x_3(t) + \varepsilon(t), \quad t = 1, 2, \dots \quad (3.4)$$

З оптимальним набором пояснюючих змінних x_i , $i = \overline{1, m}$, де $m = 3$ – кількість відібраних предикторів: $x_1(t)$ – «індекс промислової продукції, % до 1990 р.»; $x_2(t)$ – «частка відновлюваних джерел енергії у загальному постачанні первинної

енергії, % від загального постачання»; $x_3(t)$ – «частка валових приватних інвестиційних витрат у структурі ВВП, %».

Одним з обґрунтувань застосування процедур зваженого МНК для оцінки регресійних параметрів служить статистичний аналіз оптимального набору пояснюючих змінних x_i , $i = \overline{1,3}$ за алгоритмом Фаррара – Глобера. Результати відображені у табл. Б.1 – Б.6 (дод. Б). При цьому критичне значення критерію Пірсона $\chi^2_{кр} = 7,8$, критичне значення критерію Фішера $F_{кр} = 3,6$, а критичне значення критерію Стьюдента $t_{кр} = 2,10$.

Оскільки фактичне значення критерію Пірсона $\chi^2_{ф} = 27,0$ перевищує його критичне значення $\chi^2_{кр} = 7,8$, то за критерієм хі-квадрат Пірсона у масиві пояснюючих змінних $x_i(t)$, $i = \overline{1,4}$ існує мультиколінеарність. Згідно табл. Б.2 (дод. Б) за критерієм Фішера підтверджується наявність мультиколінеарності для кожного предиктора.

Згідно табл. Б.5 (дод. Б) змінні $x_1(t)$ (індекс промислової продукції) і $x_3(t)$ (частка валових приватних інвестиційних витрат у структурі ВВП) за критерієм Стьюдента мають кореляційний зв'язок між собою і згідно табл. Б.6 (дод. Б) регресор $x_1(t)$ характеризується високим додатним значенням частинного коефіцієнта кореляції з результируючим показником. Предиктор $x_2(t)$ (частка відновлюваних джерел енергії у загальному постачанні первинної енергії) за критерієм Стьюдента не має помітної мультиколінеарності з іншими пояснюючими змінними і йому відповідає високе від'ємне значення частинного коефіцієнта кореляції з результируючим показником.

Висуваючи гіпотезу, що буде перевірена далі, про нормальний розподіл випадкових збурень $\varepsilon(t)$, $t = 1, 2, \dots$ як білого шуму з нульовим математичним сподіванням, сталою дисперсією та відсутністю автокореляції, параметрична ідентифікація здійснюється узагальненим МНК з експоненціальним зважуванням для кращої адаптації моделі до нових за часом даних шляхом врахування старіння вхідної інформації. Оптимальні значення параметрів моделі знаходяться з умови мінімуму регуляризованої зваженої суми квадратів нев'язок:

$$S_w = \sum_{t=1}^T w^2(t) \cdot s^2(t) + \varepsilon_p \cdot \sum_{i=0}^m (\theta_i - \theta_i^{(0)})^2. \quad (3.5)$$

Тут $s(t) = y_e(t) - y(t)$ – відхилення (нев'язка) модельного значення залежної змінної $y(t)$ в точці t від її відповідного значення $y_e(t)$, що спостерігається на практиці; $w(t)$ – додатний ваговий коефіцієнт для спостереження в точці t , що обчислюється за формулою: $w(t) = e^{-\lambda(T-t)}$, де λ – додатний параметр старіння спостережень; ε_p – додатний параметр регуляризації (прийнято $\varepsilon_p = 0,000001$); $\theta_i^{(0)}$, $i = \overline{0, m}$ – апіорні оцінки параметрів (на початку прийнято $\theta_i^{(0)} = 0, i = \overline{0, m}$).

Значення λ вибирається з суперечливих міркувань: з одного боку, треба збільшувати λ для зростання відносної ваги «свіжих» спостережень, а з іншого боку – зменшувати λ для згладжування випадкових флуктуацій. Методом підбору визначено $\lambda = 0,001$.

Проблема знаходження оптимальних оцінок параметрів θ_i , $i = \overline{1, m}$ моделі (3.4), які мінімізують критерій МНК (3.5), є задачею опуклої оптимізації та має єдиний розв'язок. Оптимізаційна задача (3.4), (3.5) розв'язується в числовій формі в програмному середовищі MS Excel за допомогою процедури «пошук розв'язку». У результаті одержано наступні оптимальні значення параметрів:

$$\theta_0 = 4,208; \theta_1 = 0,1221; \theta_2 = -0,4851; \theta_3 = -0,07715 \quad (3.6)$$

Для усунення відмінностей у вимірі та ступені варіації різних змінних (без фіктивного предиктора $x_0(t) \equiv 1$) використовуються значення параметрів моделі у стандартизованому вигляді $\theta_i^{(s)}$, $i = \overline{1, m}$ (бета-коефіцієнти), що обчислюються за формулою:

$$\theta_i^{(s)} = \theta_i s_i / s_y, \quad i = \overline{1, m}, \quad (3.7)$$

де s_i і s_y – одержані за модельними значеннями оцінки середнього квадратичного відхилення відповідно i -го регресора $x_i(t)$, $i = \overline{1, m}$ і залежної змінної $y(t)$.

Отримані значення:

$$\theta_1 = 0,7597; \theta_2 = -0,4893; \theta_3 = -0,2004 \quad (3.8)$$

показують, що додатний параметр $\theta_1^{(s)}$, який відповідає пояснюючій змінній $x_1(t)$, за модулем помітно перевищує від'ємні коефіцієнти $\theta_2^{(s)}$ і $\theta_3^{(s)}$, що вказує на переважаючий позитивний вплив предиктора $x_1(t)$ (індекс промислової продукції, % до 1990 р.) на залежну змінну $y(t)$ (інтенсивність забруднюючих викидів) порівняно з негативним впливом факторів $x_2(t)$ (частка відновлюваних джерел енергії у загальному постачанні первинної енергії, % від загального постачання) і $x_3(t)$ (частка валових приватних інвестиційних витрат у структурі ВВП, %). Отже, для зменшення викидів забруднюючих речовин треба розвивати альтернативну енергетику і стимулювати приватні інвестиції, причому пріоритет щодо ефективності має розвиток відновлюваних джерел енергії, оскільки $|\theta_2^{(s)}| > |\theta_3^{(s)}|$.

На рисунку 3.3 подані суміщені згладжені графіки інтенсивності викидів забруднюючих речовин в атмосферне повітря в Україні за даними спостережень і за трифакторною моделлю (3.4), що демонструють ефект згладжування та високий ступінь апроксимації упродовж всього часового проміжку спостережень.

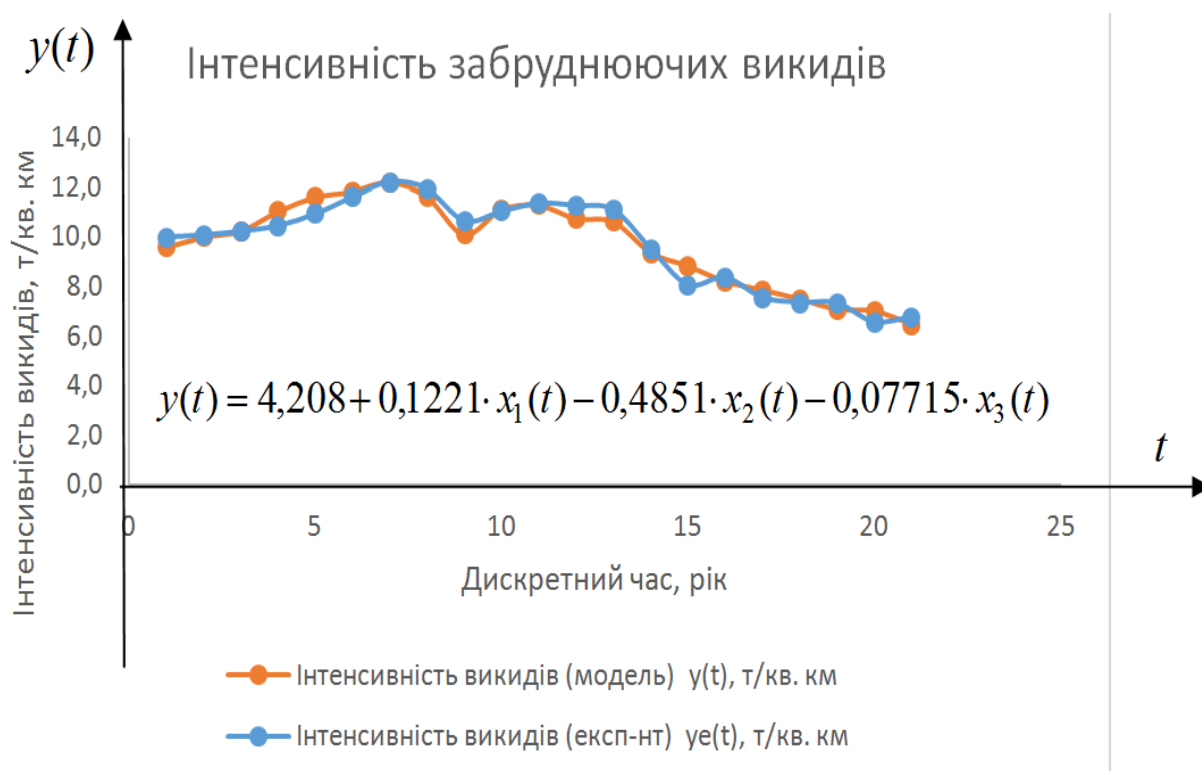


Рис. 3.3 Інтенсивність забруднюючих викидів (трифакторна регресійна модель і спостереження)

3.3. Дослідження залишків багатфакторної моделі, її адекватності та значущості коефіцієнтів

Коректне застосування МНК для оцінювання параметрів трифакторної моделі (3.4) інтенсивності забруднюючих викидів та дослідження її характеристик за загальноприйнятими статистиками Фішера і Стьюдента передбачає виконання наступних умов щодо нормального розподілу випадкових залишків моделі $\varepsilon(t)$, $t = 1, 2, \dots$ [1, 2, 6, 11, 14, 21, 24]: рівність нулю математичного сподівання кожного рівня часового ряду нев'язок $\varepsilon(t)$, $t = 1, 2, \dots$; сталість дисперсії ряду відхилень $\varepsilon(t)$, $t = 1, 2, \dots$; відсутність автокореляції в ряді збурень $\varepsilon(t)$, $t = 1, 2, \dots$. Результати розрахунків у середовищі MS Excel середніх показників щодо нормальності розподілу ряду відхилень $\varepsilon(t)$, $t = 1, 2, \dots$ та відповідних значень статистичних критеріїв наведені в табл. 3.1.

Таблиця 3.1

Середні статистичні характеристики часового ряду залишків моделі та відповідні значення статистичних критеріїв

Показник	Значення
Середнє значення ε_{cp}	$-5,49 \cdot 10^{-5}$
Медіана Me	$7,34 \cdot 10^{-2}$
Оцінка дисперсії s^2	$1,52 \cdot 10^{-1}$
Середнє квадратичне відхилення s	$3,90 \cdot 10^{-1}$
Середнє лінійне відхилення d	$3,12 \cdot 10^{-1}$
Мінімальне значення ε_{min}	$-7,67 \cdot 10^{-1}$
Максимальне значення ε_{max}	$5,48 \cdot 10^{-1}$
Розмах варіації $R = \varepsilon_{max} - \varepsilon_{min}$	1,31
Коефіцієнт асиметрії A_ε	-0,52
Коефіцієнт ексцесу E_ε	-0,58

Середнє квадратичне відхилення для коефіцієнта асиметрії S_A	0,46
Середнє квадратичне відхилення для коефіцієнта ексцесу S_E	0,76
Критичне значення критерія Стьюдента для середнього значення $t_{кр}$	2,09
Фактичне значення критерія Стьюдента для середнього значення $t_{ф}$	$6,45 \cdot 10^{-4}$

На основі аналізу табл. 3.1 можна зробити загальний висновок щодо прийнятності гіпотези про нормальний розподіл ряду відхилень $\varepsilon(t)$, $t = 1, 2, \dots$. Більш детально: оскільки $t_{ф} < t_{кр}$, то за середнім значенням розподіл ряду залишків відповідає нормальному закону; оскільки $|A_{\varepsilon}| < 1,5S_A$, то за коефіцієнтом асиметрії розподіл ряду збурень відповідає нормальному закону; оскільки $|E_{\varepsilon}| < 1,5S_E$, то за коефіцієнтом ексцесу розподіл ряду нев'язок відповідає нормальному закону.

Для дослідження автокореляції ряду залишків моделі $\varepsilon(t)$, $t = 1, 2, \dots$ використовуються: двосторонній тест Дарбіна – Уотсона; критерій фон Неймана; циклічний коефіцієнт автокореляції [2, 6, 11, 12, 21]. Результати розрахунків у табличному процесорі MS Excel показників автокореляції ряду відхилень $\varepsilon(t)$, $t = 1, 2, \dots$ та відповідних значень статистичних критеріїв наведені в табл. 3.2.

Таблиця 3.2

**Показники автокореляції ряду залишків
і відповідні значення статистичних критеріїв**

Показник	Значення
Ліве критичне значення критерія Дарбіна – Уотсона $d_1(\alpha, m, T)$	1,03
Праве критичне значення критерія Дарбіна – Уотсона $d_2(\alpha, m, T)$	1,67
Фактичне значення критерія Дарбіна – Уотсона d	1,56
Критичне значення критерія фон Неймана при додатній автокореляції $Q_{кр1}(\alpha, T)$	1,38
Критичне значення критерія фон Неймана при від'ємній автокореляції $Q_{кр2}(\alpha, T)$	3,10
Фактичне значення критерія фон Неймана Q	1,63
Ліве критичне значення циклічного коефіцієнта автокореляції $r^0_{кр1}(\alpha, T)$	-0,391
Праве критичне значення циклічного коефіцієнта автокореляції $r^0_{кр2}(\alpha, T)$	0,295
Фактичне значення скорегованого циклічного коефіцієнта автокореляції r^0	0,219

Аналізуючи табл. 3.2, можна зробити наступні висновки: оскільки $d_1 < d < d_2$, то згідно статистики Дарбіна – Уотсона спостерігається зона невизначеності щодо автокореляції; оскільки $Q_{кр1} < Q < Q_{кр2}$, то за тестом фон Неймана автокореляція залишків відсутня; оскільки $r^0_{кр1} < r^0 < r^0_{кр2}$, то за статистикою циклічного коефіцієнта автокореляції автокореляція залишків відсутня.

Додатково для дослідження автокореляції ряду залишків моделі $\varepsilon(t)$, $t = 1, 2, \dots$ можна розглянути коефіцієнт рангової кореляції Спірмена ρ та коефіцієнт знакової кореляції Фехнера K_Φ [2, 6]. Обчислені значення цих коефіцієнтів і відповідної статистики Стюдента наведені в табл. 3.3.

Таблиця 3.3

**Коефіцієнти Спірмена і Фехнера для автокореляції ряду залишків
і відповідні значення статистичних критеріїв**

Показник	Значення
Коефіцієнт рангової кореляції Спірмена ρ	0,16
Фактичне значення критерія Стюдента для коефіцієнта Спірмена t_Φ	0,72
Коефіцієнт знакової кореляції Фехнера K_Φ	0,10
Фактичне значення критерія Стюдента для коефіцієнта Фехнера t_Φ	0,44
Критичне значення критерія Стюдента $t_{кр}$	2,09

Аналізуючи таблицю 3.3, можна зробити наступні висновки: оскільки $t_\Phi < t_{кр}$, то за статистикою коефіцієнта кореляції Спірмена автокореляція залишків відсутня; оскільки $t_\Phi < t_{кр}$, то за статистикою коефіцієнта кореляції Фехнера автокореляція залишків відсутня.

За недостатності апріорної інформації основними показниками точності та адекватності моделі слугують сума квадратів нев'язок S , стандартне (середнє квадратичне) відхилення s_0 та середня відносна похибка апроксимації δ_0 [14, 21, 24]:

$$S = \sum_{t=1}^T s^2(t); \quad s_0 = \sqrt{S/(T-m-1)}; \quad \delta_0 = \frac{1}{T} \sum_{t=1}^T \left| \frac{s(t)}{y(t)} \right| \cdot 100\%. \quad (3.9)$$

Маємо: $S = 3,05$; $s_0 = 0,423$; $\delta_0 = 3,3\%$, що вказує на досить високу точність запропонованої трифакторної регресійної моделі.

Згладжений графік відхилень $s(t) = y_e(t) - y(t)$ даних спостережень від модельних значень інтенсивності забруднюючих викидів зображено на рис. 3.4. З графіка видно, що вздовж усього часового діапазону спостережень відхилення за абсолютною величиною розподілені досить рівномірно, дещо посилюючись ближче до середини діапазону.



Рис. 3.4 Відхилення між модельними значеннями інтенсивності викидів і даними спостережень

Помірну величину та рівномірність розподілу залишків моделі додатково демонструють графіки квадратів відхилень на рис. Б.1 (дод. Б) і відносних відхилень на рис. Б.2 (дод. Б).

Ступінь відображення моделлю тісноти зв'язків між вихідною змінною та предикторами описується коефіцієнтом детермінації (квадратом множинного коефіцієнта кореляції) R^2 та нормованим коефіцієнтом детермінації R_n^2 [14, 21, 24]:

$$R^2 = 1 - S/S_T; \quad R_n^2 = 1 - (1 - R^2) \cdot \frac{T-1}{T-m-1}, \quad (3.10)$$

де S_T – загальна сума квадратів відхилень даних спостережень від їхнього середнього значення y_{ecp} :

$$S_T = \sum_{t=1}^T (y(t) - y_{ecp})^2; \quad y_{ecp} = \frac{1}{T} \sum_{t=1}^T y_e(t). \quad (3.11)$$

Отримані значення $R^2 = 0,954$ і $R_n^2 = 0,946$ показують високу долю поясненої моделлю загальної суми квадратів відхилень.

Оцінка адекватності моделі здійснюється за критерієм Фішера. Для цього обчислюється фактичне значення F -статистики [14, 21, 24]:

$$F_\phi = \frac{R^2}{1-R^2} \cdot \frac{T-m-1}{m}. \quad (3.12)$$

Правдоподібність моделі підтверджується виконанням нерівності:

$$F_\phi > F_{кр}(\alpha, m, T-m-1), \quad (3.13)$$

де α – заданий рівень значущості, $\alpha = 1 - P$; P – довірна ймовірність; $F_{кр}(\alpha, n_1, n_2)$ – критична точка для рівня значущості α розподілу Фішера зі ступенями вільності n_1 і n_2 .

Порівняння одержаного фактичного значення $F_\phi = 118,1$ з критичним $F_{кр} = 3,2$ для вибраної довірчої ймовірності $P = 95\%$ демонструє виконання умови адекватності (3.13).

Перевірка значущості параметрів моделі проводиться за t -критерієм Стюдента [14, 21, 24]. До одержаного ранжування значущості оцінок параметрів потрібно підходити з обережністю, враховуючи наявність мультиколінеарності регресорів та інші неконтрольовані порушення необхідних статистичних передумов.

Для всіх параметрів θ_i , $i = \overline{0, m}$ обчислюються фактичні значення t -статистики:

$$t_\phi(i) = \frac{|\theta_i|}{s_0 \sqrt{d_{ii}}}, \quad i = \overline{0, m}, \quad (3.14)$$

де d_{ii} – i -й діагональний елемент оберненої матриці $X^T X^{-1}$, $d_{ii} = (X^T X)^{-1}_{ii}$; X – матриця спостережень значень регресорів розміру $T \times m$; X^T – транспонована матриця спостережень.

Відносний ступінь значущості оцінок параметрів θ_i , $i = \overline{0, m}$ визначається порівнянням отриманих фактичних значень $t_\phi(i)$, $i = \overline{0, m}$ статистики Стюдента між собою та відповідним критичним значенням

$$t_{кр}(\alpha, T-m-1) \quad (3.15)$$

для рівня значущості α зі ступенем вільності $n = T - m - 1$.

Результати розрахунків для вибраної довірчої ймовірності $P = 95\%$ наведені в табл. 3.4.

Таблиця 3.4

Значення статистики Стьюдента для параметрів моделі

Параметр	θ_0	θ_1	θ_2	θ_3
Фактичне значення критерія Стьюдента $t_{\phi}(i)$	4,45	9,01	6,50	2,41
Критичне значення критерія Стьюдента $t_{кр}$	2,11			

Порівняння одержаних фактичних значень t -статистики з її критичним значенням для вибраної довірчої ймовірності $P = 95\%$ показує значущість усіх параметрів моделі, при цьому найбільш значущою є оцінка коефіцієнта θ_1 , який відображає вплив на інтенсивність викидів предиктора $x_1(t)$ (індекс промислової продукції, % до 1990 р.).

Висновки до третього розділу

У рамках регресійного аналізу здійснено структурну та параметричну ідентифікацію регресійної моделі інтенсивності викидів забруднюючих речовин. Запропоновано просту однофакторну модель у адитивній та мультиплікативній формах для оціночних експрес-розрахунків. На основі обчислювальних експериментів синтезовано оптимальну за структурою трифакторну лінійну регресійну модель інтенсивності викидів та знайдено МНК-оцінки її коефіцієнтів з використанням експоненціального зважування і регуляризації для нейтралізації впливу мультиколінеарності регресорів.

Проведений аналіз трифакторної регресійної моделі за статистичними критеріями підтверджує її адекватність і значущість коефіцієнтів. Адекватність рівняння регресії перевірена за допомогою коефіцієнта детермінації та критерію Фішера. Значущість коефіцієнтів визначена за критерієм Стьюдента та доповнена

аналізом значень параметрів моделі у стандартизованому вигляді. Статистичне тестування в рамках кореляційного аналізу підтверджує відсутність автокореляції залишків. Проте наявність мультиколінеарності регресорів вимагає з обережністю ставитись до отриманих результатів. Незважаючи на високу якість апроксимації реальних даних щодо часового ряду інтенсивності забруднюючих викидів, можливість правдоподібної екстраполяції синтезованої моделі за межі часового проміжку входних спостережень потребує подальших досліджень. Регресійна модель є адаптаційною та потребує постійного корегування при включенні до розгляду нових статистичних даних.

Дослідження складових трифакторної регресійної моделі показує, що найбільший прямий вплив на інтенсивність забруднюючих викидів здійснює показник «індекс промислової продукції, % до 1990 р.», а найбільший зворотній вплив – показник «частка відновлюваних джерел енергії у загальному постачанні первинної енергії, % від загального постачання». Модель загалом можна використовувати в експертних системах для поточного аналізу та короткотермінового прогнозування інтенсивності викидів забруднюючих речовин у повітря в Україні та визначення напрямків розподілу інвестицій та інших управлінських рішень для забезпечення належного рівня екологічної безпеки. Також модель може слугувати підвищенню екологічної грамотності населення та надавати статистичні аргументи громадським активістам у боротьбі за охорону навколишнього середовища.

ВИСНОВКИ

У роботі вказано на актуальність і складність проблеми дослідження і моделювання взаємозв'язків часових рядів, що описують динаміку системи «суспільство – економіка – довкілля». Досліджено сучасні напрямки статистичного аналізу часових рядів, їхні основні усереднені показники положення, інтенсивності, варіації та взаємозв'язку; висвітлено підходи до вибору класу і структури математичних моделей ряду динаміки як реалізації випадкового процесу з дискретним часом; вказано на метод найменших квадратів як основний інструмент знаходження оптимальних оцінок параметрів моделі часового ряду та його модифікації; висвітлено апарат дослідження адекватності моделі та значущості її коефіцієнтів і рекомендації щодо покращення прогнозу за моделлю.

Розглянуто специфічні підходи до моделювання економетричних часових рядів. Сформульовано задачі статистичного аналізу вибраних показників динаміки системи «суспільство – економіка – довкілля» та синтезу лінійної регресійної моделі, яка відображає в розрізі часової динаміки вплив соціально-економічних показників розвитку нашої країни на викиди забруднюючих речовин в атмосферне повітря в Україні.

Вибрано основні статистичні показники для опису положення, інтенсивності та варіації часових рядів викидів забруднюючих речовин в атмосферне повітря в Україні та пов'язаних фізичних і соціально-економічних змінних. Розроблено обчислювальні процедури для розрахунків основних статистичних показників і проведено їх обчислення засобами табличного процесора MS Excel.

Запропоновано комбіновані показники – «інтенсивність забруднюючих викидів», «кінцеве енергоспоживання на душу населення» і «накопичені прямі іноземні інвестиції», які краще відповідають структурі звітних статистичних даних і ефективніше віддзеркалюють кореляційні зв'язки.

Досліджено тісноту взаємозв'язків між рівнями часового ряду інтенсивності забруднюючих викидів на базі обчислення рангових коефіцієнтів кореляції Спірмена і Фехнера та встановлення їх значущості за критерієм Стьюдента.

Вивчено автокореляційну функцію часового ряду інтенсивності викидів забруднюючих речовин. Спираючись на загальні закономірності економічних процесів і аналіз корелограми, обґрунтовано припущення про наявність у часовому ряді викидів лінійного тренду і коливань різної частоти та випадкової компоненти.

Сформовано початковий набір пояснюючих фізичних і соціально-економічних змінних, що гіпотетично впливають на динаміку інтенсивності забруднюючих викидів. Проведено кореляційний аналіз вхідного масиву предикторів та їх зв'язків з результируючим показником – інтенсивністю забруднюючих викидів. Підтверджено наявність високого ступеня мультиколінеарності. Відкиданням «проблемних» змінних досягнуто зниження рівня мультиколінеарності звуженого масиву пояснюючих змінних.

У рамках регресійного аналізу здійснено структурну та параметричну ідентифікацію регресійної моделі інтенсивності викидів забруднюючих речовин. Запропоновано просту однофакторну модель у адитивній та мультиплікативній формах для оціночних експрес-розрахунків. На основі обчислювальних експериментів синтезовано оптимальну за структурою трифакторну лінійну регресійну модель інтенсивності викидів та знайдено МНК-оцінки її коефіцієнтів з використанням експоненціального зважування і регуляризації для нейтралізації впливу мультиколінеарності регресорів.

Проведений аналіз трифакторної регресійної моделі за статистичними критеріями підтверджує її адекватність і значущість коефіцієнтів. Адекватність рівняння регресії перевірена за допомогою коефіцієнта детермінації та критерію Фішера. Значущість коефіцієнтів визначена за критерієм Стьюдента та доповнена аналізом значень параметрів моделі у стандартизованому вигляді. Статистичне тестування у рамках кореляційного аналізу підтверджує відсутність автокореляції залишків. Проте наявність мультиколінеарності регресорів вимагає з обережністю ставитись до отриманих результатів. Незважаючи на високу якість апроксимації реальних даних щодо часового ряду інтенсивності забруднюючих викидів, можливість правдоподібної екстраполяції синтезованої моделі за межі часового проміжку вхідних спостережень потребує подальших досліджень. Регресійна модель

є адаптаційною та потребує постійного корегування при включенні до розгляду нових статистичних даних.

Дослідження складових трифакторної регресійної моделі показує, що найбільший прямий вплив на інтенсивність забруднюючих викидів здійснює показник «індекс промислової продукції, % до 1990 р.», а найбільший зворотній вплив – показник «частка відновлюваних джерел енергії у загальному постачанні первинної енергії, % від загального постачання». Модель загалом можна використовувати в експертних системах для поточного аналізу та короткотермінового прогнозування інтенсивності викидів забруднюючих речовин у повітря в Україні та визначення напрямків розподілу інвестицій та інших управлінських рішень для забезпечення належного рівня екологічної безпеки. Також модель може слугувати підвищенню екологічної грамотності населення та надавати статистичні аргументи громадським активістам у боротьбі за охорону навколишнього середовища.

У даній роботі реалізовано об'єктно-орієнтований неформальний підхід до статистичного аналізу та моделювання з урахуванням економетричної специфіки та думки експертів. Спроби його поширення на інші масиви статистичних даних вимагають відповідних зусиль дослідників із зануренням у предметну проблематику і поданням отриманих результатів досліджень на рівні сприймання практикуючих спеціалістів.

СПИСОК ВИКОРИСТАНИХ ДЖЕРЕЛ

1. Box G. E. P. Time Series Analysis: Forecasting and Control / G. E. P. Box, G. M. Jenkins, G. C. Reinsel, G. M. Ljung. – Hoboken : Prentice Hall, 2016. – 705 p.
2. Enders W. Applied Econometrics Time Series / W. Enders. – London : John Wiley & Sons, Inc., 2010. – 517 p.
3. Eurostat. Official site. – URL:<https://ec.europa.eu/eurostat> (дата звернення: 19.08.2023).
4. International Energy Agency. Official site. – URL: <https://www.iea.org> (дата звернення: 27.08.2023).
5. International Monetary Fund. Official site. – URL: <https://www.imf.org/en/Home> (дата звернення: 15.07.2023).
6. Nielsen A. Practical Time Series Analysis: Prediction with Statistics and Machine Learning / A. Nielsen. – Sebastopol, California: O'Reilly Media, Inc., 2020. 504 p.
7. The Worldwide Governance Indicators (WGI). The World Bank Group. 2019. – URL: <http://info.worldbank.org/governance/wgi/#home> (дата звернення: 26.07.2023).
8. Transparency International's Global Corruption. The Global Corruption Barometer 2013 is the biggest ever survey tracking worldwide public opinion on corruption. – URL: <https://www.transparency.org/gcb2013/report> (дата звернення: 04.07.2023).
9. World Bank. Official site. – URL: <https://databank.worldbank.org/> (дата звернення: 21.08.2023).
10. Yahoo Finance. – URL: <https://finance.yahoo.com> (дата звернення: 07.08.2023).
11. Бідюк П. Аналіз часових рядів / П. Бідюк, В. Романенко, О. Тимошук. – Київ : Політехніка, 2010. – 317 с.
12. Бредюк В. І. Економіко-математичне моделювання в середовищі табличного процесора MS Excel / В. І. Бредюк, О. І. Джоші. – Рівне : НУВГП, 2015. – 241 с.
13. Визначення рівня енергетичної безпеки України : аналіт. доп. /

[Суходоля О. М., Харазішвілі Ю. М., Бобро Д. Г., Рябцев Г. Л., Завгородня С. П.] ; за заг. ред. О. М. Суходолі. – Київ : НІСД, 2021. – 71 с. <https://doi.org/10.53679/NISS-analytrep.2022.02> (дата звернення: 11.09.2023).

14. Геєць В. М. Моделі і методи соціально-економічного прогнозування / В. М. Геєць, Т. С. Клебанова, О. І. Чкерняк. – Харків : ВД «ІНЖЕК», 2008. – 396 с.

15. Глебов І. Прогнозування часових рядів та інтегрування торгових сигналів за допомогою PYTHON / І. Глебов, В. Кабачій // НТКП ВНТУ. Факультет інтелектуальних інформаційних технологій та автоматизації, Вінниця, 10 бер. 2021 р. – Вінниця, 2021.– С. 203–254

16. Горбійчук М. І. Порівняльний аналіз методів фільтрації при експериментальному дослідженні динамічних властивостей об'єктів / М. І. Горбійчук, М. І. Когутяк, В. С. Борин // Методи та прилади контролю якості ISSN – 2020. – № 2. – С. 53–64.

17. Горошкова Л. А. Економіко-статистичне моделювання детермінант динаміки забруднення довкілля України / Л. А. Горошкова, Є. В. Хлобистов, В. О. Трофимчук // Економіка і організація управління. – 2019. – № 2 (34). –С.46–55. – URL: <http://jeou.donnu.edu.ua/article/view/7350> (дата звернення: 25.09.2023).

18. Грицаєнко Г. І. Екологічні інвестиції як пріоритетний напрямок розвитку аграрної сфери економіки / Г. І. Грицаєнко, М. І. Грицаєнко // Науковий вісник ТДАТУ. – 2021. – Т. 1. – Вип. 11. – Режим доступу: <https://oj.tsatu.edu.ua/index.php/visnik/issue/view/19/%D0%92%D0%B8%D0%B4%D0%B0%D0%BD%D0%BD%D1%8F.%20PDF> (дата звернення: 12.09.2023).

19. Державна служба статистики України. Офіційний сайт. – Режим доступу: <https://ukrstat.gov.ua/> (дата звернення: 06.08.2023).

20. Державне агентство з енергоефективності та енергозбереження України. Офіційний сайт. – Режим доступу: <http://saee.gov.ua/> (дата звернення: 03.07.2023).

21. Козьменко О. В. Економіко-математичні методи та моделі (економетрика) / О. В. Козьменко, О. В. Кузьменко. – Суми : Унів. кн., 2014. – 405 с.

22. Міністерство захисту довкілля та природних ресурсів України. Офіційний сайт. – Режим доступу: <https://merp.gov.ua/> (дата звернення: 15.07.2023).

23. Міністерство фінансів України. Офіційний сайт. – Режим доступу: <http://www.minfin.gov.ua/> (дата звернення: 09.07.2023).
24. Мокін Б. І. Математичні методи ідентифікації динамічних систем / Б. І. Мокін, В. Б. Мокін, О. Б. Мокін. – Вінниця : ВНТУ, 2010. – 260 с.
25. Опенько І. А. Кореляційний аналіз впливу існуючої системи державного управління на використання земель лісогосподарського призначення в Україні / І. А. Опенько // Економіка та держава. – 2019. – № 7. – С. 55–62.
26. Портал Європейської Кліматичної Асоціації (European Climate Assessment) при KNMI. – Режим доступу: <http://eca.knmi.nl> (дата звернення: 25.07.2023).
27. Центральна геофізична обсерваторія ім. Б. Срезневського. Офіційний сайт. – Режим доступу: <http://cgo-sreznevskyi.kyiv.ua/uk/diialnist/klimatolohichna/klimatychni-dani-po-kyievu> (дата звернення: 04.07.2023).
28. Черняк О. І. Інтелектуальний аналіз даних / О. І. Черняк, П. В. Захарченко. – Київ : Знання, 2014. – 599 с.

ДОДАТКИ

Додаток А

**Вхідні статистичні дані, статистичні показники автокореляції
та мультиколінеарності**

Таблиця А.1

Інтенсивність щорічних викидів забруднюючих речовин
в атмосферне повітря в Україні за 2001–2021 рр.

Календарний рік спостереження t (рік)	1	2001	2002	2003	2004
Дискретний час спостереження t (рік)	2	1	2	3	4
Інтенсивність викидів $y_e(t)$ (т / кв. км)	3	10,0	10,1	10,3	10,5

Продовж. табл. А.1

1	2005	2006	2007	2008	2009	2010	2011	2012	2013
2	5	6	7	8	9	10	11	12	13
3	11,0	11,6	12,2	11,9	10,7	11,1	11,4	11,3	11,1

Продовж. табл. А.1

1	2014	2015	2016	2017	2018	2019	2020	2021
2	14	15	16	17	18	19	20	21
3	9,6	8,1	8,4	7,6	7,4	7,3	6,6	6,8

Джерело: складено автором за даними Державної служби статистики України [19].

Таблиця А.2

Значення коефіцієнтів автокореляції ряду інтенсивності забруднюючих викидів

Лаг k	1	1	2	3	4	5	6
Коефіцієнт автокореляції r_k	2	0,93	0,83	0,71	0,57	0,44	0,24

Продовж. табл. А.2

1	7	8	9	10	11	12	13	14	15	16
2	-0,20	-0,59	-0,72	-0,75	-0,69	-0,71	-0,86	-0,88	-0,80	-0,80

Основні показники (частина 1), що гіпотетично впливають на рівень викидів забруднюючих речовин у повітря в Україні (без урахування тимчасово окупованих територій) за 2001–2021 рр.

Рік t	Чисельність населення $x_1(t)$, млн. осіб	Індекс промислової продукції $x_2(t)$, % до 1990 р.	ВВП на душу населення за ПКС $x_3(t)$, дол. США	Кінцеве енергоспоживання на душу населення $x_4(t)$, тне	Частка відновлюваних джерел енергії у загальному постачанні $x_5(t)$, % від загального постачання
2001	48,9	62,1	4789	1,52	1,0
2002	48,2	63,4	5173	1,56	0,8
2003	47,8	66,8	5824	1,60	0,8
2004	47,4	73,0	6736	1,73	0,9
2005	47,1	79,0	7214	1,74	0,9
2006	46,8	84,1	8054	1,74	1,4
2007	46,5	90,1	9004	1,85	1,7
2008	46,2	85,6	9434	1,80	1,9
2009	46,0	67,9	8093	1,47	2,2
2010	45,9	76,2	8559	1,61	2,0
2011	45,7	78,8	9246	1,66	2,0
2012	45,6	74,4	9705	1,60	2,0
2013	45,5	73,7	11111	1,53	2,7
2014	43,0	61,0	10743	1,43	2,6
2015	42,8	60,0	10164	1,19	3,0
2016	42,7	61,9	11148	1,21	3,8
2017	42,5	60,1	11860	1,17	4,4
2018	42,3	57,3	12631	1,22	4,6
2019	42,0	52,5	13345	1,18	4,9
2020	41,7	54,9	13087	1,15	6,6
2021	41,2	53,7	14219	1,18	6,6

Джерело: складено автором за даними [4, 9, 10, 19, 20, 23]

Основні показники (частина 2), що гіпотетично впливають на рівень викидів забруднюючих речовин у повітря в Україні (без урахування тимчасово окупованих територій) за 2001–2021 рр.

Рік t	Накопичені прямі іноземні інвестиції $x_6(t)$, млн. дол. США	Частка валових приватних інвестиційних витрат у структурі ВВП $x_7(t)$, %	Капітальні інвестиції на охорону атмосферного повітря і проблеми зміни клімату $x_8(t)$, млн. дол. США	Поточні витрати на охорону атмосферного повітря і проблеми зміни клімату $x_9(t)$, млн. дол. США	Середня температура повітря за рік (м. Київ) $x_{10}(t)$, °C
2001	792	22	–	56,8	8,8
2002	1485	20	–	78,1	9,3
2003	2909	22	–	83,4	8,2
2004	4624	21	–	114,0	8,7
2005	12432	23	–	171,2	8,7
2006	18036	25	151,0	163,7	8,4
2007	27927	28	273,1	226,1	9,9
2008	38840	27	280,3	256,3	9,6
2009	43656	17	163,5	132,9	9,4
2010	50151	18	143,6	165,7	9,4
2011	57358	20	318,2	185,2	9,2
2012	65759	20	308,2	167,9	9,0
2013	70258	16	301,8	177,1	9,4
2014	70668	13	161,1	104,2	9,4
2015	70210	16	65,1	69,6	10,5
2016	74020	22	98,0	68,9	9,5
2017	77712	20	98,1	79,1	9,8
2018	82167	19	128,9	106,5	9,5
2019	88027	15	165,5	114,7	10,6
2020	87159	9	207,5	88,1	10,9
2021	93846	14	–	–	9,2

Джерело: складено автором за даними [10, 19, 23, 27].

Таблиця А.5

Фактичне значення критерія Фішера F_ϕ для пояснюючих змінних.

Наявність ($M = 1$) чи відсутність ($M = 0$) мультиколінеарності
вибраної пояснюючої змінної з іншими

№ з/п	Предикторна змінна	F_ϕ	$F_{кр}$	M
1	Чисельність населення $x_1(t)$, млн. осіб	33,4	4,8	1
2	Індекс промислової продукції $x_2(t)$, % до 1990 р.	31,2	4,8	1
3	ВВП на душу населення за ПКС $x_3(t)$, дол. США	12,8	4,8	1
4	Кінцеве енергоспоживання на душу населення $x_4(t)$, тне	84,6	4,8	1
5	Частка постачання енергії від відновлюваних джерел $x_5(t)$, %	9,5	4,8	1
6	Накопичені прямі іноземні інвестиції $x_6(t)$, млн. дол. США	18,3	4,8	1
7	Частка приватних інвестиційних витрат у ВВП $x_7(t)$, %	5,0	4,8	1
8	Капітальні інвестиції на охорону $x_8(t)$, млн. дол. США	4,6	4,8	0
9	Поточні витрати на охорону $x_9(t)$, млн. дол. США	25,5	4,8	1
10	Середня температура повітря за рік $x_{10}(t)$, °С	3,0	4,8	0

Таблиця А.6

Частинні коефіцієнти кореляції $r_{ij}^{(p)}$ масиву пояснюючих змінних x_i , $i = \overline{1,10}$

	x_1	x_2	x_3	x_4	x_5	x_6	x_7	x_8	x_9	x_{10}
x_1	-1,0	0,6	-0,5	-0,7	-0,3	-0,6	-0,6	0,5	0,8	-0,7
x_2	0,6	-1,0	0,0	0,7	0,4	0,3	0,8	-0,1	-0,5	0,5
x_3	-0,5	0,0	-1,0	-0,1	0,5	0,2	0,0	0,1	0,4	-0,3
x_4	-0,7	0,7	-0,1	-1,0	-0,6	-0,7	-0,7	0,5	0,8	-0,7
x_5	-0,3	0,4	0,5	-0,6	-1,0	-0,6	-0,5	0,4	0,3	-0,2
x_6	-0,6	0,3	0,2	-0,7	-0,6	-1,0	-0,6	0,7	0,5	-0,5
x_7	-0,6	0,8	0,0	-0,7	-0,5	-0,6	-1,0	0,1	0,7	-0,6
x_8	0,5	-0,1	0,1	0,5	0,4	0,7	0,1	-1,0	-0,2	0,3
x_9	0,8	-0,5	0,4	0,8	0,3	0,5	0,7	-0,2	-1,0	0,8
x_{10}	-0,7	0,5	-0,3	-0,7	-0,2	-0,5	-0,6	0,3	0,8	-1,0

Частинні коефіцієнти кореляції $r_{yx}^{(p)}$ пояснюючих змінних
з результуючим показником

	x_1	x_2	x_3	x_4	x_5	x_6	x_7	x_8	x_9	x_{10}
y	0,62	0,03	0,21	0,44	-0,71	0,30	0,03	0,31	-0,28	0,19

Таблиця А.10

Парні коефіцієнти кореляції r_{xy} між вихідною змінною $y(t)$ і пояснюючими змінними x_i , $i = \overline{1,5}$ та якісна оцінка тісноти взаємозв'язку за шкалою Чеддока

№ з/п	Предикторна змінна	Кількість спостережень	Коефіцієнт кореляції r_{xy}	Ступінь зв'язку
1	Індекс промислової продукції $x_1(t)$, % до 1990 р.	21	0,92	дуже тісний
2	Кінцеве енергоспоживання на душу населення $x_2(t)$, тне	21	0,95	дуже тісний
3	Частка постачання енергії від відновлюваних джерел $x_3(t)$, %	21	-0,85	тісний
4	Накопичені прямі іноземні інвестиції $x_4(t)$, млн. дол. США	21	-0,65	помітний
5	Частка приватних інвестиційних витрат у ВВП $x_5(t)$, %	21	0,66	помітний

Таблиця А.11

Фактичне значення критерія Фішера F_ϕ для пояснюючих змінних.

Наявність ($M = 1$) чи відсутність ($M = 0$) мультиколінеарності
вибраної пояснюючої змінної з іншими

№ з/п	Предикторна змінна	F_ϕ	$F_{кр}$	M
1	Індекс промислової продукції $x_1(t)$, % до 1990 р.	56,4	3,0	1
2	Кінцеве енергоспоживання на душу населення $x_2(t)$, тне	76,9	3,0	1
3	Частка постачання енергії від відновлюваних джерел $x_3(t)$, %	20,0	3,0	1
4	Накопичені прямі іноземні інвестиції $x_4(t)$, млн. дол. США	25,9	3,0	1
5	Частка приватних інвестиційних витрат у ВВП $x_5(t)$, %	9,2	3,0	1

Таблиця А.12

Частинні коефіцієнти кореляції $r_{ij}^{(p)}$ масиву пояснюючих змінних x_i , $i = \overline{1,5}$

	x_1	x_2	x_3	x_4	x_5
x_1	-1,00	0,90	-0,02	0,65	0,60
x_2	0,90	-1,00	-0,20	-0,59	-0,41
x_3	-0,02	-0,20	-1,00	0,52	0,00
x_4	0,65	-0,59	0,52	-1,00	-0,49
x_5	0,60	-0,41	0,00	-0,49	-1,00

Таблиця А.13

Фактичне значення критерія Стюдента t_f для частинних коефіцієнтів кореляції

	x_2	x_3	x_4	x_5
x_1	8,31	0,10	3,41	3,00
x_2		0,80	2,92	1,78
x_3			2,46	0,01
x_4				2,23

Таблиця А.14

Наявність ($M = 1$) чи відсутність ($M = 0$)

мультиколінеарності пояснюючих змінних за критерієм Стюдента

	x_2	x_3	x_4	x_5
x_1	1	0	1	1
x_2		0	1	0
x_3			1	0
x_4				1

Таблиця А.15

Частинні коефіцієнти кореляції $r_{yx}^{(p)}$ пояснюючих змінних
з результуючим показником

	x_1	x_2	x_3	x_4	x_5
y	0,47	0,47	-0,78	0,50	-0,27

Таблиця А.16

Парні коефіцієнти кореляції r_{xy} між вихідною змінною $y(t)$ і пояснюючими змінними x_i , $i = \overline{1,4}$ та якісна оцінка тісноти взаємозв'язку за шкалою Чеддока

№ з/п	Предикторна змінна	Кількість спостережень	Коефіцієнт кореляції r_{xy}	Ступінь зв'язку
1	Індекс промислової пропорції $x_1(t)$, % до 1990 р.	21	0,92	дуже тісний
2	Кінцеве енергоспоживання на душу населення $x_2(t)$, тне	21	0,95	дуже тісний
3	Частка постачання енергії від відновлюваних джерел $x_3(t)$, %	21	-0,85	тісний
4	Частка приватних інвестиційних витрат у ВВП $x_5(t)$, %	21	0,66	помітний

Таблиця А.17

Фактичне значення критерія Фішера F_ϕ для пояснюючих змінних.

Наявність ($M=1$) чи відсутність ($M=0$) мультиколінеарності
вибраної пояснюючої змінної з іншими

№ з/п	Предикторна змінна	F_ϕ	$F_{кр}$	M
1	Індекс промислової пропорції $x_1(t)$, % до 1990 р.	43,9	3,2	1
2	Кінцеве енергоспоживання на душу населення $x_2(t)$, тне	69,2	3,2	1
3	Частка постачання енергії від відновлюваних джерел $x_3(t)$, %	19,0	3,2	1
4	Частка приватних інвестиційних витрат у ВВП $x_5(t)$, %	8,6	3,2	1

Таблиця А.18

Частинні коефіцієнти кореляції $r_{ij}^{(p)}$ масиву пояснюючих змінних x_i , $i = \overline{1,4}$

	x_1	x_2	x_3	x_4
x_1	-1,00	0,84	0,49	0,43
x_2	0,84	-1,00	-0,73	-0,17
x_3	0,49	-0,73	-1,00	-0,34
x_4	0,43	-0,17	-0,34	-1,00

Таблиця А.19

Фактичне значення критерія Стюдента t_f для частинних коефіцієнтів кореляції

	x_2	x_3	x_4
x_1	6,48	2,30	1,94
x_2		4,45	0,71
x_3			1,50

Таблиця А.20

Наявність (M=1) чи відсутність (M=0) мультиколінеарності
пояснюючих змінних за критерієм Стюдента

	x_2	x_3	x_4
x_1	1	1	0
x_2		1	0
x_3			0

Таблиця А.21

Частинні коефіцієнти кореляції $r_{yx}^{(p)}$ пояснюючих змінних
з результуючим показником

	x_1	x_2	x_3	x_4
y	0,72	0,20	-0,69	-0,48

**Статистичні показники мультиколінеарності пояснюючих змінних
та візуалізація залишків синтезованої трифакторної моделі**

Таблиця Б.1

Парні коефіцієнти кореляції r_{xy} між вихідною змінною $y(t)$ і пояснюючими змінними x_i , $i = \overline{1,3}$ та якісна оцінка тісноти взаємозв'язку за шкалою Чеддока

№ з/п	Предикторна змінна	Кількість спостережень	Коефіцієнт кореляції r_{xy}	Ступінь зв'язку
1	Індекс промислової продукції $x_1(t)$, % до 1990 р.	21	0,92	дуже тісний
2	Частка постачання енергії від відновлюваних джерел $x_2(t)$, %	21	-0,85	тісний
3	Частка приватних інвестиційних витрат у ВВП $x_3(t)$, %	21	0,66	помітний

Таблиця Б.2

Фактичне значення критерія Фішера F_ϕ для пояснюючих змінних.

Наявність ($M = 1$) чи відсутність ($M = 0$) мультиколінеарності
вибраної пояснюючої змінної з іншими

№ з/п	Предикторна змінна	F_ϕ	$F_{кр}$	M
1	Індекс промислової продукції $x_1(t)$, % до 1990 р.	13,7	3,6	1
2	Частка постачання енергії від відновлюваних джерел $x_2(t)$, %	9,1	3,6	1
3	Частка приватних інвестиційних витрат у ВВП $x_3(t)$, %	13,0	3,6	1

Таблиця Б.

Частинні коефіцієнти кореляції $r_{ij}^{(p)}$ масиву пояснюючих змінних x_i , $i = \overline{1,3}$

	x_1	x_2	x_3
x_1	-1,00	-0,36	0,53
x_2	-0,36	-1,00	-0,32
x_3	0,53	-0,32	-1,00

Таблиця Б.4

Фактичне значення критерія Стюдента t_{ϕ} для частинних коефіцієнтів кореляції

	x_2	x_3
x_1	1,65	2,68
x_2		1,45

Таблиця Б.5

Наявність (M=1) чи відсутність (M=0) мультиколінеарності
пояснюючих змінних за критерієм Стюдента

	x_2	x_3
x_1	0	1
x_2		0

Таблиця Б.6

Частинні коефіцієнти кореляції $r_{yx}^{(p)}$ пояснюючих змінних
з результуючим показником

	x_1	x_2	x_3
y	0,91	-0,84	-0,50



Рис. Б.1 Квадрат відхилення між модельними значеннями
інтенсивності викидів і даними спостережень



Рис. Б.2 Відносне відхилення між модельними значеннями
інтенсивності викидів і даними спостережень