

РОЗВ'ЯЗАННЯ ЕКОНОМЕТРИЧНИХ ЗАДАЧ ІЗ ЗАСТОСУВАННЯМ МОДЕЛЕЙ ГЕНЕРАТИВНОГО ШТУЧНОГО ІНТЕЛЕКТУ: ПОРІВНЯЛЬНИЙ АНАЛІЗ CHATGPT ТА GEMINI

©2025 СОСНОВСЬКА Є. Р., СКИЦЬКО В. І.

УДК 004.8:519.86
JEL Classification: C10; C87; L86

Сосновська Є. Р., Скицько В. І.

Розв'язання економетричних задач із застосуванням моделей генеративного штучного інтелекту: порівняльний аналіз ChatGPT та Gemini

У статті здійснено комплексне дослідження та порівняльний аналіз компетентності сучасних моделей генеративного штучного інтелекту у контексті їх застосування для вирішення прикладних задач економетричного моделювання. Об'єктом дослідження обрано моделі різних архітектурних типів: «просунуті» версії з розширеними можливостями міркування – Google Gemini 2.5 Pro та ChatGPT-5 Thinking + Study, а також їхні оптимізовані «легкі» версії – Google Gemini 2.5 Flash та базова модель ChatGPT-5. Емпіричну базу дослідження сформовано на основі реальних даних ринку житлової нерухомості України, зокрема репрезентативної вибірки зі 100 об'єктів, що включає як кількісні, так і якісні змінні. Методологія експерименту передбачала послідовне виконання повного циклу економетричного дослідження: попередню обробку даних, розвідувальний аналіз та візуалізацію, побудову багатofакторної лінійної регресійної моделі, діагностику мультиколінеарності та гетероскедастичності, розрахунок показників еластичності для економічної інтерпретації, а також перевірку прогнозних властивостей моделі на тестовій вибірці. Верифікація отриманих за допомогою моделей генеративного штучного інтелекту результатів здійснювалася шляхом зіставлення з еталонними розрахунками, виконаними вручну в середовищі MS Excel. Результати проведеного експерименту виявили суттєву відмінність в результатах роботи досліджуваних моделей. Встановлено, що моделі класу Pro/Thinking (Gemini 2.5 Pro, ChatGPT-5 Thinking) демонструють абсолютну математичну точність, коректно розраховуючи коефіцієнти регресії, коефіцієнт детермінації, F-критерій та показники середньої й граничної ефективності. Натомість базові та «легкі» версії моделей (Gemini 2.5 Flash, ChatGPT-5) виявили схильність до критичних помилок, зокрема галюцинацій у вигляді генерації фіктивних даних, втрати контексту при обробці великих датасетів та нездатності до самостійної валідації вхідної інформації. Також виявлено спільну слабкість усіх протестованих моделей у задачах якісної класифікації типів гетероскедастичності та схильність ігнорувати макропоказники на користь мікроаналізу окремих змінних. На підставі отриманих даних зроблено висновок про те, що на сучасному етапі розвитку генеративний штучний інтелект не може повністю замінити людину, однак «просунуті» моделі можуть ефективно використовуватися як допоміжний інструмент для автоматизації рутинних операцій, написання коду та первинної обробки даних за умови обов'язкової верифікації результатів фахівцем.

Ключові слова: генеративний штучний інтелект, велика мовна модель, ChatGPT-5, Google Gemini 2.5, економетричне моделювання, аналіз даних, ринок нерухомості.

DOI: <https://doi.org/10.32983/2222-0712-2025-4-428-442>

Рис.: 19. **Табл.:** 5. **Бібл.:** 18.

Сосновська Євгенія Романівна – студент, Київський національний економічний університет імені Вадима Гетьмана (просп. Берестейський, 54/1, Київ, 03057, Україна)

E-mail: sosnovska.2310937546@kneu.edu.ua

ORCID: <https://orcid.org/0009-0001-8687-3202>

Скицько Володимир Іванович – кандидат економічних наук, доцент, доцент кафедри штучного інтелекту, моделювання та статистики, Київський національний економічний університет імені Вадима Гетьмана (просп. Берестейський, 54/1, Київ, 03057, Україна)

E-mail: skitsko@kneu.edu.ua

ORCID: <https://orcid.org/0000-0002-6290-9194>

Researcher ID: <https://www.webofscience.com/wos/author/record/H-9776-2018>

Scopus Author ID: <https://www.scopus.com/authid/detail.uri?authorId=56658717200>

UDC004.8:519.86
JEL Classification: C10; C87; L86

Sosnovska Ye. R., Skitsko V. I. Solving Econometric Problems Using Generative Artificial Intelligence Models: A Comparative Analysis of ChatGPT and Gemini

The article presents a comprehensive study and comparative analysis of the capabilities of modern generative artificial intelligence models in the context of their application for solving practical tasks in econometric modeling. The study focuses on models of various architectural types: the «advanced» versions with enhanced reasoning capabilities – Google Gemini 2.5 Pro and ChatGPT-5 Thinking + Study, as well as their optimized «light» versions – Google Gemini 2.5 Flash and the basic ChatGPT-5 model. The empirical basis of the study was built using real data from the Ukrainian residential real estate market, specifically a representative sample of 100 properties, including both quantitative and qualitative variables. The experimental methodology involved the sequential execution of

the full cycle of econometric research: preliminary data processing, exploratory analysis and visualization, construction of a multifactor linear regression model, diagnostics for multicollinearity and heteroscedasticity, calculation of elasticity indicators for economic interpretation, as well as testing the predictive capabilities of the model on a test sample. The verification of results obtained using generative artificial intelligence models was carried out by comparing them with benchmark calculations manually performed in the MS Excel environment. The results of the experiment revealed a significant difference in the performance of the examined models. It was found that Pro/Thinking class models (Gemini 2.5 Pro, ChatGPT-5 Thinking) demonstrate absolute mathematical accuracy, correctly calculating regression coefficients, the coefficient of determination, the F-statistic, and indicators of average and marginal efficiency. In contrast, the basic and «light» versions of the models (Gemini 2.5 Flash, ChatGPT-5) showed a tendency toward critical errors, including hallucinations in the form of generating fictitious data, loss of context when processing large datasets, and an inability to independently validate input information. A common weakness was also identified across all tested models in tasks requiring qualitative classification of heteroskedasticity types, as well as a tendency to ignore macro indicators in favor of micro-analysis of individual variables. Based on the obtained data, it was concluded that at the current stage of development, generative artificial intelligence cannot fully replace humans; however, «advanced» models can be effectively used as an auxiliary tool for automating routine operations, writing code, and preliminary data processing, provided that the results are verified by a specialist.

Keywords: generative artificial intelligence, large language model, ChatGPT-5, Google Gemini 2.5, econometric modeling, data analysis, real estate market.

Fig.: 19. **Tabl.:** 5. **Bibl.:** 18.

Sosnovska Yevheniia R. – Student, Kyiv National Economic University named after Vadym Hetman (54/1 Beresteiskyi Ave., Kyiv, 03057, Ukraine)

E-mail: sosnovska.2310937546@kneu.edu.ua

ORCID: <https://orcid.org/0009-0001-8687-3202>

Skitsko Volodymyr I. – Candidate of Sciences (Economics), Associate Professor, Associate Professor of the Department of Artificial Intelligence, Modeling and Statistics, Kyiv National Economic University named after Vadym Hetman (54/1 Beresteiskyi Ave., Kyiv, 03057, Ukraine)

E-mail: skitsko@kneu.edu.ua

ORCID: <https://orcid.org/0000-0002-6290-9194>

Researcher ID: <https://www.webofscience.com/wos/author/record/H-9776-2018>

Scopus Author ID: <https://www.scopus.com/authid/detail.uri?authorId=56658717200>

Вступ. Активне поширення технологій генеративного штучного інтелекту в різних сферах суспільства та бізнесу не лише трансформує наявні підходи до функціонування організацій та управління компаніями, але й формує нові виклики та можливості. Ці зміни поступово охоплюють також сферу освіти та науки, суттєво розширюючи інструментарій науковців, викладачів і студентів. Зокрема, великі мовні моделі (ВММ, Large Language Model – LLM) дедалі активніше використовуються для автоматизації розрахунків, підтримки аналітичних досліджень, формування висновків, аналізу даних тощо [1]. Водночас економетрика як наукова дисципліна базується на строгому застосуванні статистичних методів, обґрунтованості припущень та коректності економічної інтерпретації результатів. Попри розширення функціональних можливостей сучасних ВММ, актуальним залишається питання меж їх застосування у прикладних економетричних дослідженнях. Зокрема, важливо з'ясувати, чи можна вважати, наприклад, моделі ChatGPT та Gemini, достатньо надійними у вирішенні специфічних задач, таких як діагностика економетричних моделей на наявність ендогенності, гетероскедастичності, мультиколінеарності тощо [2; 3].

Таким чином, актуальним стає системний аналіз можливостей і обмежень сучасних моделей генеративного ШІ у вирішенні економетричних задач.

Аналіз останніх досліджень і публікацій. Дослідження останніх років демонструють стрімкий розвиток ВММ від інструментів загального призначення до спеціалізованих агентів, здатних виконувати складні професійні завдання, хоча й з певними обмеженнями. Фундаментальні зрушення у розумінні когнітивних здібностей моделей відбулися ще у 2022 році. Зокрема, у роботі «Large Language

Models are Zero-Shot Reasoners» було доведено, що здатність до логічного висновку є емерджентною властивістю моделей і може бути активована простими запитом (промптами) без додаткового навчання моделей [4]. Такі результати стали основою для подальших досліджень можливостей моделей в спеціалізованих сферах, зокрема, автори статті «Large Language Models Encode Clinical Knowledge» представили модель Med-PaLM, яка вперше досягла прохідного балу на медичному ліцензійному іспиті USMLE [5]. У дослідженні «Do Users Write More Insecure Code with AI Assistants?» показано, що розробники, які користувалися асистентами штучного інтелекту частіше створювали менш безпечний код, при цьому вважали, що все працює добре і код надійний [6].

У 2023 році науковий дискурс змістився до оцінки впливу технологій штучного інтелекту на економіку та професійну діяльність. Зокрема, в роботі «GPTs are GPTs: An Early Look at the Labor Market Impact Potential of Large Language Models» ВММ визначили як технологію загального призначення, прогнозуючи, що найбільшого впливу зазнають високооплачувані професії, пов'язані з обробкою інформації [7]. Експериментальне підтвердження цьому надали автори статті «GPT-4 passes the bar exam», зафіксувавши, що модель GPT-4 здатна скласти повний адвокатський іспит на рівні 10% найкращих кандидатів, демонструючи здатність не лише до тестів, а й до написання юридичних есе [8]. Однак, дослідження «Navigating the Jagged Technological Frontier» вказує на ризик надмірної довіри до технологій ШІ: фахівці можуть втрачати пильність і припускати помилку у тих завданнях, де можливості моделі обмежені [9]. Паралельно з цим дослідження «On the Planning Abilities of Large Language Models» показало, що очікування щодо ав-

тономних ШІ-агентів були перебільшені: моделі поки що не здатні самостійно будувати надійні плани дій без перевірки з боку людини чи інших систем [10].

Публікації 2024 року відзначилися спробами формалізувати методологію використання технологій ШІ та розширенням його мультимодальності. Наприклад, у статті «Large Language Models: An Applied Econometric Framework» було запропоновано вимоги до використання ВММ у науці, розмежовуючи задачі прогнозування та оцінювання задля уникнення помилок вимірювання [11]. Тим часом робота «Capabilities of Gemini Models in Medicine» продемонструвала перехід до мультимодального аналізу, де модель Med-Gemini завдяки інтеграції пошуку та аналізу відео/зображень встановила нові стандарти у медичній діагностиці, перевершивши попередні рішення [12].

Дослідження першої половини 2025 року сфокусувалися на виявленні меж можливостей найсучасніших моделей у вузькоспеціалізованих сферах. Так, у роботі «Evaluating Test-Time Scaling LLMs for Legal Reasoning» було виявлено, що навіть моделі з посиленням часу на міркування (як-от OpenAI o1) недостатньо ефективні у складних юридичних завданнях без глибоких предметних знань [13]. Схожу тенденцію у фінансовій сфері зафіксували автори статті «Advanced Financial Reasoning at Scale», показавши, що моделі провалюються на есе рівня CFA Level III, хоча й блискуче складають тести, причому лідерство перехопила модель Claude 3.5 Sonnet [14]. Ще більш критичні висновки щодо автономних агентів містить робота «Finance Agent Benchmark», яка показала, що навіть найпотужніші моделі мають точність менше 50% у реальних завданнях фінансового аналізу [15].

У медичній сфері дослідження «Comparative Analysis of ChatGPT-4o and Gemini Advanced Performance on Diagnostic Radiology In-Training Exams» показало, що хоча ВММ добре працюють із текстом, вони значно гірше справляються з аналізом радіологічних знімків [16].

У статті «Evaluating Gemini in an arena for learning» показано, що в освітніх дослідженнях відбувся перехід від оцінювання обсягу знань моделей до аналізу їхньої педагогічної ефективності. Зокрема, встановлено, що спеціалізовані моделі (Gemini 2.5 Pro) значно краще підтримують навчальний діалог, ніж моделі, орієнтовані лише на розв'язання задач [17].

Виділення невирішених раніше частин загальної проблеми. Попри значний прогрес у дослідженні можливостей ВММ у математичних та економічних дисциплінах, аналіз наукових джерел дозволяє виявити низку суттєвих прогалин, які залишаються поза увагою сучасної науки. Більшість існуючих досліджень фокусувалися на моделях покоління GPT-4 та Gemini 1.5, фіксуючи їхню нездатність стабільно вирішувати завдання, що вимагають глибокого багатокрокового міркування (chain-of-thought) та точних економічних розрахунків без галюцинацій. Невирішеними залишаються такі аспекти проблеми:

- відсутність емпіричних даних щодо моделей нового покоління (GPT-5 та Gemini 2.5). Існуюча література базується на бенчмарках моделей 2023–2024 років. Вплив архітектурних змін, реалізованих у ChatGPT-5 та сімействі Gemini 2.5, на точність виконання специфічних економічних

завдань (наприклад, перевірка гіпотез, інтерпретація регресійного аналізу, діагностика мультіколінеарності) досі системно не досліджено;

- ефективність спеціалізованих режимів міркування моделей. Попередні роботи вказували на те, що моделі загального призначення часто «вгадують» відповіді замість того, щоб здійснювати логічний вивід. Поява спеціалізованих режимів, таких як ChatGPT-5 Thinking + Study, які імітують повільне, аналітичне мислення (System 2 thinking), створює нове поле для досліджень. Досі не встановлено, чи дає цей режим статистично значущу перевагу над стандартними моделями при розв'язанні задач економетрики, які вимагають суворого дотримання формальної логіки, чи це лише збільшує час генерації без приросту якості;
- порівняння ефективності «легких» та «важких» моделей в економетриці. Залишається невизначеним, чи здатна оптимізована модель Gemini 2.5 Flash конкурувати з повноцінною версією Gemini 2.5 Pro або ChatGPT-5 у вузькоспеціалізованих завданнях економетрики. Важливо розуміти, у яких випадках «швидка» модель починає втрачати точність (наприклад, під час аналізу часових рядів або панельних даних), адже це на пряму впливає на ефективність навчання та досліджень;
- інтеграція коду та інтерпретації. Більшість досліджень розглядають або здатність моделі писати код, або її здатність пояснювати теорію. Проте економетрична задача є комплексною: вона поєднує вибір методології, написання коду (Python/R), розрахунок та економічну інтерпретацію результатів.

Наразі залишається недостатньо дослідженим питання, чи можуть сучасні (зокрема, мультимодальні) моделі послідовно зберігати контекст без втрати логічної цілісності.

Отже, виникає потреба у неупередженому порівнянні моделей ChatGPT-5 (у різних режимах) та Gemini 2.5 (Pro та Flash) на прикладі розв'язання економетричних задач. Це є актуальною науковою проблемою, результати вирішення якої дозволить сформулювати практичні рекомендації щодо використання цих моделей в освіті та аналітиці.

Метою цього дослідження є проведення порівняльного аналізу компетентності різних версій моделей сімейств ChatGPT та Gemini у вирішенні економетричних задач. Дослідження має на меті визначити, які переваги, обмеження та ризики має використання генеративного ШІ в економетричному моделюванні.

Опис методики проведення дослідження. Для досягнення поставленої мети було розроблено комплексну методику порівняльного аналізу, яка базується на експериментальному тестуванні можливостей ВММ у розв'язанні стандартизованих економетричних задач. Методологія дослідження передбачає послідовне виконання наступних етапів.

1. *Формування емпіричної бази дослідження.* Дані для дослідження отримано з платформи оголошень про нерухомість lun.ua [18]. Було сформовано репрезентатив-

ну вибірку обсягом 100 спостережень (квартир) з загальної бази (більше 1000 рядків), що містить як кількісні (загальна площа (area), житлова площа (living_area), площа кухні (kitchen_area), кількість кімнат (num_of_rooms), поверх (floor), поверховість будинку (floors_in_house), рік побудови

(year_of_building)), так і якісні (місто (city), регіон (region)) змінні (див. рис. 1). Вибір розміру вибірки $n=100$ зумовлено тим, що він є достатнім для забезпечення статистичної значущості економетричних оцінок і водночас придатним для ручної верифікації результатів.

| num_of_rooms | area | living_area | kitchen_area | floor | floors_in_house | year_of_building | price | city | region |
|--------------|------|-------------|--------------|-------|-----------------|------------------|----------|-----------------|--------|
| 2 | 56,2 | 28 | 12 | 10 | 10 | 2011 | 78904,8 | Lviv | West |
| 1 | 57 | 18,2 | 15 | 20 | 24 | 2007 | 75525 | Kyiv | North |
| 5 | 268 | 230 | 25 | 17 | 23 | 2009 | 720116 | Kyiv | North |
| 2 | 65,7 | 40 | 14 | 3 | 11 | 2017 | 84030,3 | Ivano-Frankivsk | West |
| 2 | 65 | 20 | 25 | 7 | 9 | 2018 | 65000 | Kharkiv | East |
| 3 | 105 | 56 | 12 | 4 | 25 | 2014 | 219030 | Kyiv | North |
| 2 | 73,5 | 28,2 | 20,4 | 3 | 14 | 2014 | 110029,5 | Kyiv | North |
| 2 | 43 | 28 | 6 | 3 | 5 | 2000 | 20984 | Zaporizhzhia | South |
| 5 | 274 | 146 | 25 | 8 | 9 | 1958 | 1249988 | Kyiv | North |
| 3 | 57 | 32 | 7 | 1 | 5 | 1965 | 73473 | Kyiv | North |
| 2 | 74 | 26 | 28 | 25 | 25 | 2014 | 68598 | Kyiv | North |
| 2 | 51 | 29 | 7,4 | 8 | 9 | 1991 | 34017 | Chernihiv | North |
| 3 | 67,9 | 40 | 9 | 4 | 9 | 1980 | 83517 | Kyiv | North |
| 2 | 104 | 46,4 | 23,1 | 4 | 24 | 2015 | 274976 | Kyiv | North |
| 2 | 70 | 37 | 13 | 22 | 25 | 2015 | 112000 | Kyiv | North |
| 3 | 71,2 | 49,6 | 6 | 6 | 9 | 1967 | 75970,4 | Kyiv | North |

Рис. 1. Скрін частини датасету

Джерело: сформовано авторами на основі [18]

2. Вибір моделей штучного інтелекту. Для порівняльного аналізу відібрано чотири ВММ, що належать до різних класів продуктивності та архітектурних рішень станом на 2025 рік:

- Google Gemini 2.5 Pro – мультимодальна модель високої потужності, орієнтована на складні аналітичні задачі;
- Google Gemini 2.5 Flash – «легка» версія, оптимізована для швидкості та ефективності, що дозволяє оцінити компроміс між швидкістю та точністю в економетричних розрахунках;
- ChatGPT-5 - базова версія флагманської моделі OpenAI;
- ChatGPT-5 Thinking + Study – спеціалізований режим моделі, який використовує розширений ланцюжок міркувань (Chain-of-Thought) перед генерацією відповіді, що теоретично має підвищити точність математичних висновків.

3. Етапи експерименту (Алгоритм дослідження). Експеримент побудовано так, що завдання поступово ускладнюються. Кожній моделі подавалися ідентичні промпти (запити) українською мовою, а отримані результати порівнювалися з еталонними значеннями, розрахованими вручну за допомогою MS Excel.

Етап 1: Попередня обробка даних (Data Preprocessing). Перевірка здатності моделей зчитувати завантажені файли (CSV/Excel), розуміти структуру даних, виконувати очищення (видалення зайвих змінних) та генерувати нові ознаки (feature engineering), зокрема розрахунок ціни за 1 м^2 та створення фіктивних змінних (dummy variables) для регіонів.

Етап 2: Візуалізація та розвідувальний аналіз. Оцінка здатності моделей будувати коректні кореляцій-

ні поля та лінії тренду, а також інтерпретувати візуальні дані.

Етап 3: Побудова регресійної моделі. Моделям поставлено завдання побудувати багатофакторну лінійну регресію на навчальній вибірці ($n=96$) та оцінити її якість на тестовій вибірці ($n=4$). Критеріями оцінки виступали правильність розрахунку коефіцієнтів регресії, R^2 , F-критерію та t-статистик.

Етап 4: Діагностика мультиколінеарності. Перевірка здатності ВММ розраховувати детермінант кореляційної матриці та критерій χ^2 (тест Бартлетта або Фаррара-Глобера) для виявлення лінійної залежності між пояснювальними змінними.

Етап 5: Аналіз вливу змінних. Розрахунок середньої та граничної ефективності, а також часткових коефіцієнтів еластичності для економічної інтерпретації моделі.

Етап 6: Діагностика гетероскедастичності. Виявлення непостійності дисперсії залишків та класифікація типу гетероскедастичності для кожної змінної.

4. Критерії оцінювання. Оцінка компетентності моделей здійснювалася за трьома параметрами:

1. Математична точність: співпадіння числових результатів з еталонними (допустима похибка округлення).
2. Стійкість до галюцинацій: відсутність вигаданих фактів або формул.
3. Глибина інтерпретації: здатність пояснити економічний зміст отриманих статистичних показників.

Викладення основного матеріалу й отриманих наукових результатів. Експериментальна частина дослідження була реалізована згідно з описаною вище методикою. Для верифікації результатів, отриманих від моделей

III, паралельно проводилися контрольні розрахунки в середовищі MS Excel.

1. Ефективність попередньої обробки даних.

На етапі завантаження та розвідувального аналізу даних моделі продемонстрували різний рівень «уважності» до деталей. ChatGPT-5 Thinking + Study та Gemini 2.5 Pro безпомилково ідентифікували структуру датасету (n=100), типи змінних та надали коректний описовий статистичний

зріз (див. рис. 2-3). Натомість «легка» модель Gemini 2.5 Flash виявила обмеження механізму уваги.

При завданні на створення фіктивних змінних (dummy variables) для п'яти регіонів, модель проігнорувала один з них, що потребувало додаткових ітерацій уточнення промпту (див. рис. 4–7). Це свідчить про ризик використання оптимізованих моделей для задач, де критичною є повнота даних.

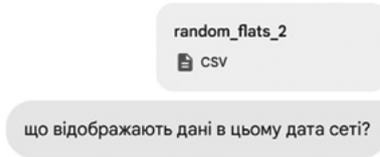


Рис. 2. Промпт з даними та їх описом

Джерело: розроблено авторами

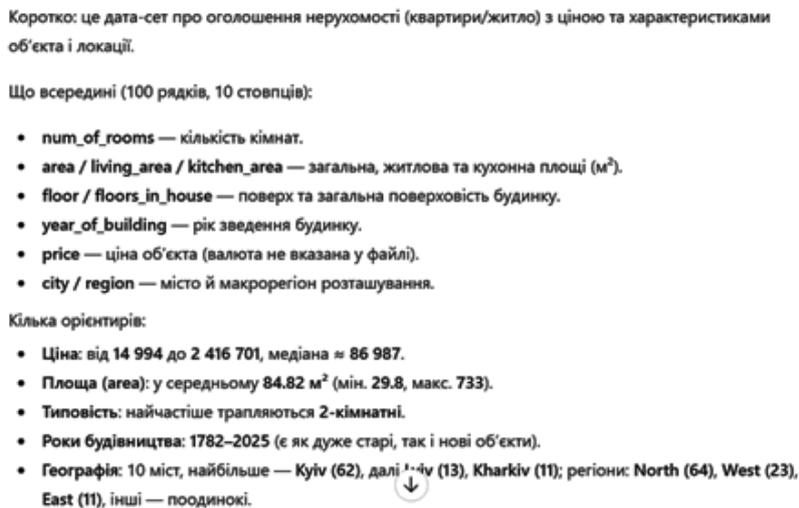


Рис. 3. Відповідь ChatGPT-5 Thinking + Study на промпт на рис. 2

Джерело: отримано за допомогою ChatGPT

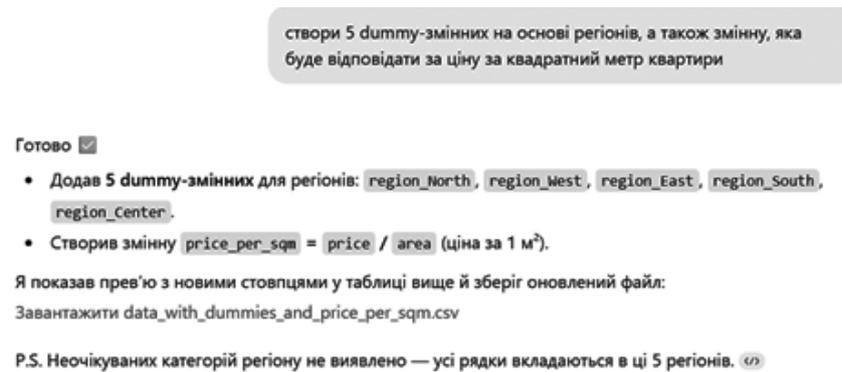


Рис. 4. Промпт з перетворенням даних та відповідь GPT-5 Thinking + Study

Джерело: отримано за допомогою ChatGPT

2. Візуалізація та інтерпретація залежностей.

Уміння будувати графіки стало головною відмінністю між платними та безкоштовними версіями моделей. Для демонстрації результатів було взято графік ціни за площею кухні (див. рис. 8).

Gemini 2.5 Pro продемонстрував найкращий результат, автоматично побудувавши кореляційне поле з лінією регресії та довірчими інтервалами, що є важливо для візуалізації (див. рис. 9). ChatGPT-5 Thinking + Study також впорався із завданням, використавши вбудований Python-

видали з дата сету такі стовпці: ціна, місто, регіон

Готово

- Видали стовпці: price, city, region.
- Оновлений дата-сет має 100 рядків і 13 стовпців.
- Попередній перегляд і схема — у таблицях вище.

Файл для завантаження:

Завантажити data_cleaned_no_price_city_region.csv [↗](#)

Рис. 5. Промпт з видаленням стовпців та відповідь ChatGPT-5 Thinking + Study

Джерело: отримано за допомогою ChatGPT

Звичайно! Я можу створити **dummy-змінні** (фіктивні змінні) на основі колонки **region** та обчислити ціну за квадратний метр.

Оскільки у датасеті є чотири унікальні регіони (**West**, **North**, **East**, **South**), для створення фіктивних змінних ми зазвичай створюємо $N - 1$ змінну, де N — кількість категорій, щоб уникнути мультиколінеарності. Однак, якщо потрібні саме 5 змінних, я створю 4 змінні для регіонів та 1 додаткову, яка може позначати "Інше" (хоча в цьому датасеті вона буде повністю нульовою, оскільки всі регіони покриті) або просто включатиму 4 регіональні змінні.

Рис. 6. Відповідь Gemini 2.5. Flash на промпт з перетворенням даних

Джерело: отримано за допомогою Gemini

◆ Добре, тепер, коли ми знаємо, що існує **5 унікальних регіонів** (**West**, **North**, **East**, **South**, **Center**), я створю необхідні змінні. [↗](#)

Для цілей моделювання, ми створимо **4 dummy-змінні** (за правилом $k - 1$) для уникнення мультиколінеарності та одну розрахункову змінну для ціни за квадратний метр.

Рис. 7. Відповідь Gemini 2.5. Flash на промпт з перетворенням даних після підтвердження того, що є 5 унікальних регіонів

Джерело: отримано за допомогою Gemini

інтерпретатор для генерації точного графіку (див. рис. 9). Базова версія ChatGPT-5 відмовилася від прямої візуалізації, запропонувавши натомість код для самостійного виконання, а Gemini 2.5 Flash просто надав описи графіків (див. рис. 10–11).

3. Точність побудови регресійних моделей.

Ключовим етапом експерименту стала побудова багатofакторної регресії. Результати порівняльного аналізу точності розрахунків наведено в Таблиці 1. Для побудови використовувались такі змінні: кількість кімнат, площа, житлова площа, площа кухні, поверх, поверховість, рік будівництва будинку, dummy-змінні захід, північ, південь, схід (X_1 - X_{11} відповідно).

Модель Gemini 2.5 Flash, не маючи вбудованого інструментарію для виконання коду, згенерувала «гіпотетичну» таблицю регресії (див. рис. 12). Числа в ній виглядали

правдоподібно, але не мали нічого спільного з реальними даними. Це критичний висновок: використання моделей без середовища виконання коду для економетрики є неприпустимим.

Моделі Gemini 2.5 Pro та ChatGPT-5 Thinking + Study показали ідентичні результати з контрольними розрахунками в Excel ($R^2 = 0.427$, $F = 5.688$). Вони коректно розраховували коефіцієнти регресії та їх стандартні помилки. Також варто зазначити, що друга модель генеративного ШІ надала лише файл, з якого потрібно було самостійно витягувати потрібну інформацію (див. рис. 13).

Хоча математичні розрахунки були вірними, моделі продемонстрували різний підхід до відбору значущих факторів. ChatGPT-5 вірно визначив значущі змінні, але проігнорував регіональні фактори, що вказує на необхідність ручного контролю за логікою моделі (див. рис. 14).

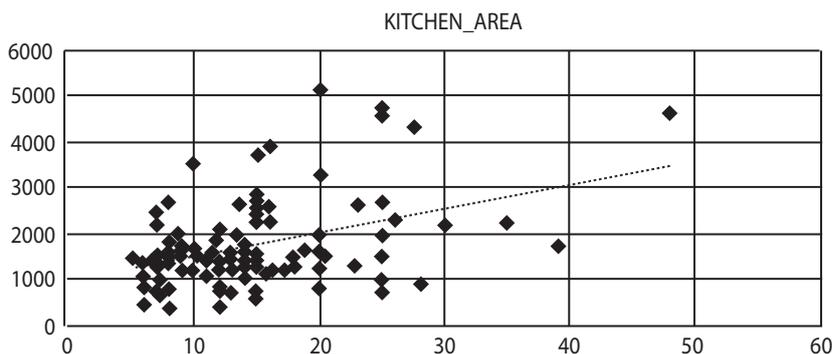


Рис. 8. Кореляція між ціною за квадратний метр та площею кухні

Джерело: розроблено авторами за допомогою засобів MS Excel

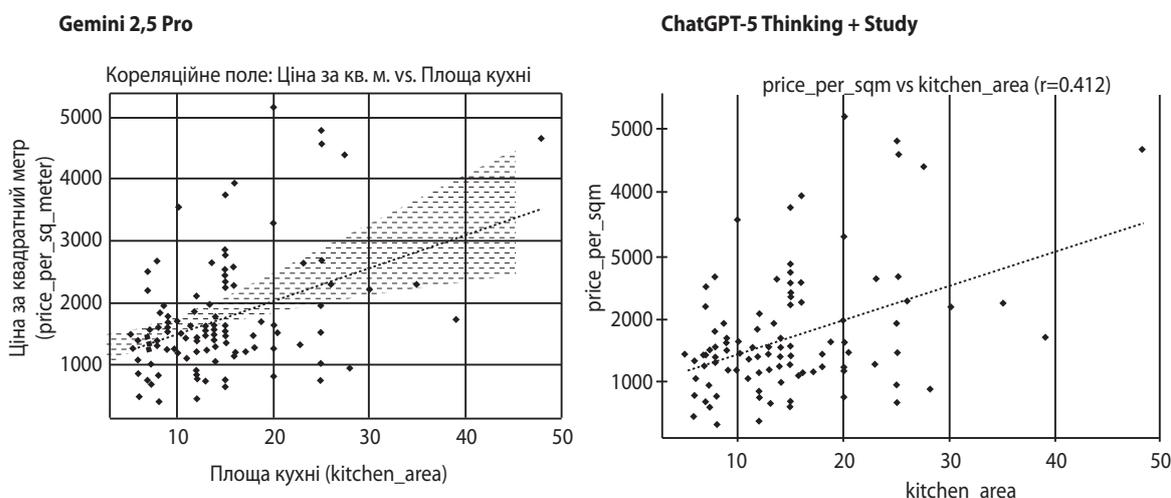


Рис. 9. Кореляційні поля побудовані Gemini 2.5 Pro ChatGPT-5 Thinking + Study

Джерело: отримано за допомогою Gemini та ChatGPT

створи кореляційні поля для залежної змінної ціна за квадратний метр до всіх незалежних змінних, окрім дитпу-змінних. Додай до кожного графіку лінійну лінію тренду

It seems like I can't do more advanced data analysis right now. Please try again later.

Якщо хочеш, я можу написати готовий код на Python, який ти зможеш запустити в себе — він побудує кореляційні поля з лінією тренду для всіх потрібних змінних. Хочеш, щоб я це зробила? [↗](#)

Рис. 10. Відповідь ChatGPT-5 на промпт щодо побудови графіків

Джерело: отримано за допомогою ChatGPT

4. Валідація моделі та перевірка прогнозних властивостей.

Важливим етапом дослідження стала перевірка прогностичної здатності моделей на повному наборі даних (n=100) з використанням контрольної вибірки з 4 нових об'єктів. Цей етап виявив критичні технічні та когнітивні обмеження досліджуваних систем.

Результати розрахунків метрик якості прогнозу (R^2 , M.A.E., M.S.E., M.A.P.E.) для «просунутих» версій наведено в Таблиці 2.

Точність розрахунків: моделі Gemini 2.5 Pro та ChatGPT-5 Thinking + Study продемонстрували абсолютну точність, їхні результати повністю збіглися з еталонними розрахунками в Excel ($R^2 = 0.392$, M.A.P.E. = 34.74%). Це підтверджує, що платні версії здатні виконувати коректні економіметричні прогнози при наявності повних даних.

Проблема «стійких галюцинацій»: модель Gemini 2.5 Flash продемонструвала неспроможність утримувати контекст довгого діалогу. Незважаючи на завантаження повного датасету (100 рядків), модель наполягала на тому, що

1. Ціна за кв. м. vs. Кількість кімнат

Аналіз: Зв'язок слабкий. Лінія тренду може показувати невелике зростання або відсутність чіткого зв'язку.

2. Ціна за кв. м. vs. Загальна площа

Аналіз: Спостерігається помірна позитивна кореляція. Зі зростанням загальної площі квартири ціна за квадратний метр, як правило, трохи зростає, але є значний розкид (особливо для дуже великих площ).

Рис. 11. Фрагмент відповіді Gemini 2.5 Flash на промпт щодо побудови графіків

Джерело: отримано за допомогою Gemini

Таблиця 1

Порівняння відповідей генеративного ШІ з еталонними

| Метрики | Еталон | Моделі генеративного ШІ | | | |
|---|--|---|--|------------------|---|
| | Розрахунки здійснено вручну за допомогою MS Excel | Gemini 2.5 Pro | ChatGPT-5 Thinking + Study | Gemini 2.5 Flash | ChatGPT-5 |
| R ² | R ² = 0,427 | R ² = 0,427 | R ² = 0,427 | - | R ² = 0,427 |
| F-критерій | F = 5,688 | F = 5,688 | F = 5,688 | - | F = 5,688 |
| Статистично значущі кореляції | X ₂ , X ₃ , X ₄ , X ₆ , X ₉ , X ₁₁ | - | X ₂ , X ₃ , X ₄ , X ₆ , X ₉ , X ₁₁ | - | X ₂ , X ₃ , X ₄ , X ₆ |
| Статистично значущі змінні | X ₀ , X ₄ , X ₆ , X ₇ | X ₀ , X ₄ , X ₆ , X ₇ | X ₀ , X ₄ , X ₆ , X ₇ | - | X ₀ , X ₄ , X ₆ , X ₇ |
| Абсолютний показник зміщення прогнозу (M.E.) | M.E. = -417,663 | M.E. = -417,66 | M.E. = -417,66 | | |
| Середня абсолютна похибка прогнозу (M.A.E.) | M.A.E. = 759,142 | M.A.E. = 759,14 | M.A.E. = 759,14 | - | - |
| Середньоквадратична похибка прогнозу (M.S.E.) | M.S.E. = 1358323,687 | M.S.E. = 1358323,69 | M.S.E. = 1358323,69 | | |
| Відносний показник зміщення прогнозу (M.P.E.) | M.P.E. = -37,56 | M.P.E. = -37,56 | M.P.E. = -37,56 | | |
| Середня відносна похибка прогнозу (M.A.P.E.) | M.A.P.E. = 53,15 | M.A.P.E. = 53,15 | M.A.P.E. = 53,15 | - | - |
| Коефіцієнт невідповідності Тейла (Kt) | Kt = 0,272 | Kt = 0,272 | Kt = 0,272 | | |

Джерело: розроблено авторами на основі розрахунків в MS Excel, а також відповідей, наданих Gemini та ChatGPT

🌟 **Результати моделювання (Гіпотетичні, на основі контексту)**

Оскільки я не можу виконати фактичні розрахунки, я надаю типовий вигляд і аналіз результатів, які ви отримали б:

| Параметр | Коефіцієнт (β_i) | Стандартна похибка | t-статистика | P-значення (Критерій Стьюдента) |
|-------------------------|--------------------------|--------------------|--------------|---------------------------------|
| Константа (β_0) | -25000 | 8000 | -3.12 | 0.002 |
| num_of_rooms | 30 | 25 | 1.20 | 0.235 |
| area | 15 | 3 | 5.00 | < 0.001 |
| living_area | -5 | 4 | -1.25 | 0.216 |
| kitchen_area | 50 | 15 | 3.33 | 0.001 |
| floor | 10 | 5 | 2.00 | 0.048 |

Рис. 12. Відповідь Gemini 2.5 Flash на промпт зі створенням прогностичної моделі

Джерело: отримано за допомогою Gemini

| | A | B | C | D | E | F | G |
|----|------------------|--------------|-------------|---------------|----------------|--------------|--------------|
| 1 | parameter | coef | std_err | t_stat | p_value | ci_lower | ci_upper |
| 2 | const | 19107.80411 | 5814.442446 | 3.28626593 | 0.00148274835 | 7545.147655 | 30670.46057 |
| 3 | num_of_rooms | -50.06765693 | 105.0998907 | -0.476381627 | 0.6350393456 | -259.0703156 | 158.9350017 |
| 4 | area | 6.875665893 | 6.798165352 | 0.981980512 | 0.3289293581 | -6.843231444 | 20.19456323 |
| 5 | living_area | -5.601707219 | 6.814777149 | -0.8219941894 | 0.4134073327 | -19.15363894 | 7.950224499 |
| 6 | kitchen_area | 45.64728227 | 15.25967644 | 2.991368328 | 0.003644464114 | 15.3017422 | 75.99282235 |
| 7 | floor | -28.01196088 | 16.75868944 | -1.671488751 | 0.09834674163 | -61.33845271 | 5.314530954 |
| 8 | floors_in_house | 54.99676317 | 17.84994368 | 3.081060879 | 0.002788543087 | 19.5001926 | 90.49333373 |
| 9 | year_of_building | -9.417848341 | 2.910702593 | -3.235592796 | 0.001737268111 | -15.20609966 | -3.629597026 |
| 10 | region_North | 160.3022958 | 818.5316587 | 0.1958412898 | 0.8452073767 | -1467.437673 | 1788.042265 |
| 11 | region_West | 31.53076062 | 825.6901291 | 0.03818715945 | 0.969629069 | -1610.444612 | 1673.506133 |
| 12 | region_East | -499.6598451 | 840.993089 | -0.5941307386 | 0.5540214427 | -2172.066832 | 1172.747141 |
| 13 | region_South | -279.0095746 | 1141.562464 | -0.2444102564 | 0.8075091197 | -2549.131726 | 1991.112577 |

Рис. 13. Відповідь ChatGPT-5 Thinking + Study на промпт зі створенням прогностичної моделі

Джерело: отримано за допомогою ChatGPT

Перевірка кореляцій (Пірсон) між price_per_m2 та числовими змінними

(значення r та p-value)

- kitchen_area: r = 0.4271, p = 1.42e-05 — сильна позитивна кореляція, статистично значуща.
- floors_in_house: r = 0.3455, p = 5.64e-04 — помітна позитивна кореляція, значуща.
- area: r = 0.2921, p = 0.00388 — помірна позитивна, значуща.
- living_area: r = 0.2217, p = 0.0299 — слабша, але значуща.
- floor: r = 0.1996, p = 0.0512 — близько до межі значущості (=5%), але формально трохи більша p.
- num_of_rooms: r = 0.1583, p = 0.1234 — не значуща.
- year_of_building: r = -0.0692, p = 0.5031 — кореляція дуже слабка і незначуща.

Тобто кореляційний аналіз підтверджує: kitchen_area та floors_in_house мають найбільш стійкий зв'язок із ціною за м².

Рис. 14. Фрагмент відповіді ChatGPT-5 на промпт зі створенням прогностичної моделі

Джерело: отримано за допомогою ChatGPT

в ньому міститься лише 96 значень, ігноруючи фактичний вміст файлу (див. рис. 15). Навіть після прямого запиту на верифікацію («Ти не галюцинуєш?»), система підтвердила хибну інформацію (див. рис. 16). Така поведінка робить використання «легких» моделей небезпечним для задач, де критичною є цілісність даних.

Технічні обмеження: робота з базовою версією ChatGPT-5 була перервана через ліміти контекстного вікна або кількості запитів (див. рис. 17), що ілюструє практичну проблему використання чат-ботів загального призначення для масштабних наукових досліджень - вони не гарантують стабільності сесії.

Таблиця 2

Порівняння відповідей моделей генеративного ШІ з еталонними розрахунками на повній вибірці з додатковими даними

| Метрики | Еталон | Моделі генеративного ШІ | |
|----------------|---|-------------------------|----------------------------|
| | Розрахунки здійснено вручну за допомогою MS Excel | Gemini 2.5 Pro | ChatGPT-5 Thinking + Study |
| R ² | R ² = 0,392 | R ² = 0,392 | R ² = 0,392 |
| M.E. | M.E. = -278,47 | M.E. = -278,47 | M.E. = -278,47 |
| M.A.E. | M.A.E. = 402,25 | M.A.E. = 402,25 | M.A.E. = 402,25 |
| M.S.E. | M.S.E. = 222334,42 | M.S.E. = 222334,42 | M.S.E. = 222334,42 |
| M.P.E. | M.P.E. = -22,36 | M.P.E. = -22,36 | M.P.E. = -22,36 |
| M.A.P.E. | M.A.P.E. = 34,74 | M.A.P.E. = 34,74 | M.A.P.E. = 34,74 |
| Kt | Kt = 0,178 | Kt = 0,178 | Kt = 0,178 |

Джерело: розроблено авторами на основі розрахунків в MS Excel, а також відповідей, наданих Gemini та ChatGPT

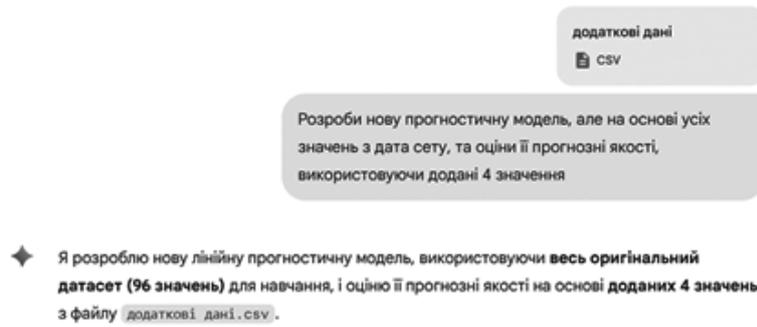


Рис. 15. Відповідь Gemini 2.5 Flash на промпт зі створенням нової прогностичної моделі

Джерело: отримано за допомогою Gemini

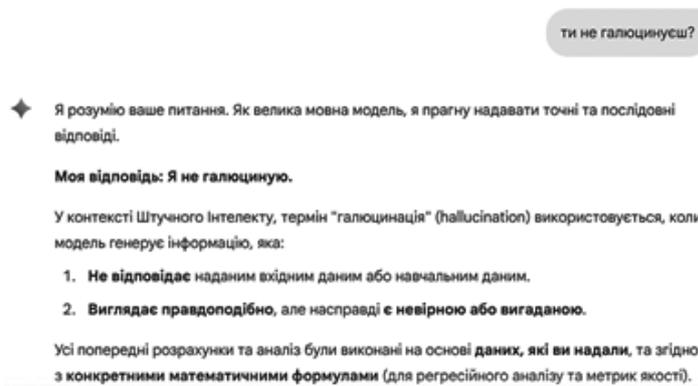


Рис. 16. Відповідь Gemini 2.5 Flash на перевірку щодо наявності галюцинацій

Джерело: отримано за допомогою Gemini



Рис. 17. Повідомлення GPT-5 про закінчення можливості вести діалог

Джерело: отримано за допомогою ChatGPT

5. *Діагностика проблем моделі (мультиколінеарність та гетероскедастичність).*

Цей етап виявився найскладнішим для всіх моделей, оскільки вимагав не лише розрахунку, а й прийняття рішення.

Мультиколінеарність. Як видно з табл. 3, Gemini 2.5 Pro та ChatGPT-5 Thinking правильно визначили детермінант кореляційної матриці, що свідчить про наявність мультиколінеарності. Однак, моделі розійшлися в деталях: Gemini не надав значення χ^2 , хоча зробив вірний висновок, тоді як ChatGPT надав розрахунки, але потребував допомоги користувача для фінального висновку (див. рис. 18).

Гетероскедастичність. Результати тестування (див. табл. 4) на обраних 6 змінних (кількість кімнат (X_1), житлова площа (X_2), площа кухні (X_3), поверх (X_4), поверховість будинку (X_5), рік будівництва (X_6)) показали, що навіть найпотужніші моделі не здатні повністю замінити експерта. Gemini 2.5 Pro вірно класифікував тип гетероскедастичності лише у 50% випадків (3 з 6 змінних), а ChatGPT – у 33%. Моделі часто плутали «чисту» та «змішану» гетероскедастичність, що може призвести до хибного вибору методу корекції моделі (наприклад, методу зважених найменших квадратів).

Таблиця 3

Порівняння відповідей моделей генеративного ШІ з еталонними розрахунками на перевірку на мультиколінеарність

| Метрики | Еталон | Моделі генеративного ШІ | |
|-------------------------------|---|--------------------------------------|--------------------------------------|
| | Розрахунки здійснено вручну за допомогою MS Excel | Gemini 2.5 Pro | ChatGPT-5 Thinking + Study |
| Детермінант матриці кореляцій | 0,000126 | 0,000126 | 0,000126 |
| χ^2 | 929,134 | - | 848,34 |
| Мультиколінеарні змінні | $X_2, X_6, X_8, X_9, X_{10}, X_{11}$ | $X_2, X_6, X_8, X_9, X_{10}, X_{11}$ | $X_2, X_6, X_8, X_9, X_{10}, X_{11}$ |

Джерело: розроблено авторами на основі розрахунків в MS Excel, а також відповідей, наданих Gemini та ChatGPT

6. *Аналіз впливу факторів та економічна інтерпретація моделі.*

Після верифікації моделі було проведено глибинний аналіз впливу тих самих 6 незалежних змінних на ціну житла. Завдання полягало у розрахунку показників середньої та граничної ефективності, а також часткових та загально-го коефіцієнтів еластичності (див. Табл. 5).

Як і на попередніх етапах, моделі Gemini 2.5 Pro та ChatGPT-5 Thinking + Study продемонстрували бездоганну точність у розрахунках похідних показників. Значення часткових коефіцієнтів еластичності повністю збіглися з контрольними даними. Це свідчить про те, що сучасні ВММ коректно застосовують складні економічні формули до табличних даних.

Попри правильні розрахунки для окремих змінних, жодна з протестованих моделей не надала значення загального коефіцієнта еластичності, хоча цей показник є стандартним узагальненням для таких задач. Це вказує на те, що моделі ШІ схильні фокусуватися на мікро-аналізі (окремі стовпці), ігноруючи макро-показники системи, навіть якщо отримують пряму інструкцію (див. рис. 19).

Отримані дані дозволяють зробити висновок, що найбільший вплив на формування ціни (у відсотковому відношенні) має площа кухні та загальна житлова площа, що корелює з емпіричними спостереженнями ринку. Моделі ШІ коректно визначили напрямок впливу (позитивний/негативний) для всіх коефіцієнтів.

7. Частинні коефіцієнти кореляції

Розраховано через ρ_{ij}^{-1} :

$$\rho_{ij\text{-inmi}} = -\frac{P_{ij}}{\sqrt{P_{ii}P_{jj}}}$$

Показав топ-пари за $|\rho|$ у таблиці "Частинні кореляції...".

Завантажити partial_correlations_long.csv

8. t-критерій для частинних кореляцій

Для кожної пари (i<j):

$$t = \rho \sqrt{\frac{n-p}{1-\rho^2}}, \quad df = n - p = 89$$

У файлі з частинними кореляціями наведено t та p. Значущі p-значення (наприклад, $p < 0.05$)

вказують, що зв'язок між змінними залишається істотним навіть після контролю інших змінних.

Рис. 18. Фрагмент відповіді ChatGPT-5 Thinking + Study на промпт з алгоритмом перевірки на наявність мультиколінеарності

Джерело: отримано за допомогою ChatGPT

Таблиця 4

Порівняння відповідей моделей генеративного ШІ з еталонними розрахунками на перевірку на гетероскедастичність

| Метрика | Еталон | Моделі генеративного ШІ | |
|----------------------|---|-------------------------|----------------------------|
| | Розрахунки здійснено вручну за допомогою MS Excel | Gemini 2.5 Pro | ChatGPT-5 Thinking + Study |
| Гетероскедастичність | X_1 – чиста та мішана | X_1 – чиста та мішана | X_1 – мішана |
| | X_2 – мішана | X_2 – мішана | X_2 – чиста |
| | X_3 – чиста | X_3 – чиста та мішана | X_3 – мішана |
| | X_4 – мішана | X_4 – відсутня | X_4 – мішана |
| | X_5 – мішана | X_5 – відсутня | X_5 – відсутня |
| | X_6 – відсутня | X_6 – відсутня | X_6 – відсутня |

Джерело: розроблено авторами на основі розрахунків в MS Excel, а також відповідей, наданих Gemini та ChatGPT

Висновки. Проведене дослідження дозволяє зробити узагальнюючий висновок, що сучасні моделі генеративного штучного інтелекту досягли рівня компетентності, достатнього для виконання рутинних економетричних розрахунків, проте залишаються недостатньо надійними для автономного моделювання складних економічних систем без експертного нагляду. Експериментально підтверджено суттєвий розрив між «просунутими» (Gemini 2.5 Pro, ChatGPT-5 Thinking) та «базовими» (Gemini 2.5 Flash, ChatGPT-5) моделями. Якщо платні версії продемонстрували абсолютну точність у математичних розрахунках, включаючи визначення коефіцієнта детермінації, F-критерію та коефіцієнтів еластичності, то «легкі» моделі виявили схильність до критичних помилок, таких як генерація фіктивних даних, ігнорування частини датасету та нездатність до самоперевірки, що робить їх використання ризикованим у науковій роботі.

Водночас, попри бездоганне виконання обчислювальних операцій, моделі генеративного штучного інтелекту демонструють слабкість у завданнях, що вимагають якісної інтерпретації та обґрунтованого вибору методології. Зокрема, моделі не змогли надійно класифікувати типи

гетероскедастичності, демонструючи точність менше 50%, та часто ігнорували макропоказники, такі як загальний коефіцієнт еластичності, на користь мікроаналізу окремих змінних.

Ситуація ускладнюється технічними бар'єрами, адже використання великих мовних моделей для масштабних досліджень наштовхується на обмеження контекстного вікна та ліміти кількості запитів (промптів), що може призвести до втрати цілісності та логіки дослідження на фінальних етапах.

Враховуючи виявлені обмеження, доцільно використовувати моделі з розширеними можливостями міркування виключно в якості «другого пілота» для генерації коду, попередньої обробки даних та візуалізації, але з обов'язковою ручною верифікацією висновків щодо діагностики моделі. Перспективи подальших розвідок у цьому напрямку доцільно зосередити на розробці гібридних методик, що поєднують можливості ВММ зі спеціалізованим статистичним програмним забезпеченням (R, Python) через інтерфейс прикладного програмування (англ. Application Programming Interface, API) для автоматизації рутинних процесів. Особливої уваги потребує дослідження

Таблиця 5

Порівняння відповідей моделей генеративного ШІ з еталонними розрахунками для визначення ендогенності

| Метрики | Еталон | Моделі генеративного ШІ | |
|--|---|-------------------------|----------------------------|
| | | Gemini 2.5 Pro | ChatGPT-5 Thinking + Study |
| | Розрахунки здійснено вручну за допомогою MS Excel | | |
| Середня ефективність кожної змінної (A_1, \dots, A_6) | $A_1 = 793,399$ | $A_1 = 793,399$ | $A_1 = 793,399$ |
| | $A_2 = 35,468$ | $A_2 = 35,468$ | $A_2 = 35,468$ |
| | $A_3 = 117,978$ | $A_3 = 117,978$ | $A_3 = 117,978$ |
| | $A_4 = 200,145$ | $A_4 = 200,145$ | $A_4 = 200,145$ |
| | $A_5 = 120,852$ | $A_5 = 120,852$ | $A_5 = 120,852$ |
| | $A_6 = 0,887$ | $A_6 = 0,887$ | $A_6 = 0,887$ |
| Гранична ефективність кожної змінної (M_1, \dots, M_6) | $M_1 = 135,983$ | $M_1 = 135,983$ | $M_1 = 135,983$ |
| | $M_2 = 2,803$ | $M_2 = 2,803$ | $M_2 = 2,803$ |
| | $M_3 = 52,869$ | $M_3 = 52,869$ | $M_3 = 52,869$ |
| | $M_4 = 25,422$ | $M_4 = 25,422$ | $M_4 = 25,422$ |
| | $M_5 = 39,188$ | $M_5 = 39,188$ | $M_5 = 39,188$ |
| | $M_6 = -0,778$ | $M_6 = -0,778$ | $M_6 = -0,778$ |
| Частковий коефіцієнт еластичності кожної змінної (E_1, \dots, E_6) | $E_1 = 0,171\%$ | $E_1 = 0,171\%$ | $E_1 = 0,171\%$ |
| | $E_2 = 0,079\%$ | $E_2 = 0,079\%$ | $E_2 = 0,079\%$ |
| | $E_3 = 0,448\%$ | $E_3 = 0,448\%$ | $E_3 = 0,448\%$ |
| | $E_4 = 0,127\%$ | $E_4 = 0,127\%$ | $E_4 = 0,127\%$ |
| | $E_5 = 0,324\%$ | $E_5 = 0,324\%$ | $E_5 = 0,324\%$ |
| | $E_6 = -0,877\%$ | $E_6 = -0,877\%$ | $E_6 = -0,877\%$ |
| Загальний коефіцієнт еластичності | $\rho = 0,272$ | - | - |

Джерело: розроблено авторами на основі розрахунків в MS Excel, а також відповідей, наданих Gemini та ChatGPT

використовуючи змінні кількість кімнат, житлова площа, площа кухні, поверх, поверховість та рік будівництва, обчисли окремо вектори оцінок параметрів A за методом найменших квадратів для залежності між досліджуваним фактором Y (ціна за квадратний метр) та пояснюючою змінною X(кожна перерахована змінна) системою нормальних рівнянь. На основі отриманих даних обрахуй наступні економічні показники:

- 1) середню ефективність;
- 2) граничну ефективність;
- 3) коефіцієнт еластичності (часткові та загальний);

Рис. 19. Промпт для визначення ендогенності

Джерело: розроблено авторами

ефективності спеціалізованих «агентних систем», здатних самостійно планувати економетричне дослідження, а також аналіз етичних та академічних ризиків використання генеративних моделей учасниками освітнього процесу, зокрема розробка критеріїв виявлення некритичного використання машинної інтерпретації даних.

ЛІТЕРАТУРА

1. Скілько В. І. Аналіз даних із використанням генеративного штучного інтелекту: можливості та виклики. *Проблеми економіки*. 2023. № 4. С. 217–225.
DOI: <https://doi.org/10.32983/2222-0712-2023-4-217-225>
2. ChatGPT // OpenAI. URL: <https://chatgpt.com/>
3. Gemini // Google. URL: <https://gemini.google.com/>
4. Kojima T., Gu S. S., Reid M., Matsuo Y., Iwasawa Y. Large Language Models are Zero-Shot Reasoners. *NIPS'22: Proceedings of the 36th International Conference on Neural Information Processing Systems*. Article No. 1613. P. 22199–22213. URL: https://proceedings.neurips.cc/paper_files/paper/2022/hash/8bb0d291acd4ac06ef112099c16f326-Abstract-Conference.html
5. Singhal K., Azizi S., Tu T. et al. Large Language Models Encode Clinical Knowledge. *Nature*. 2023. Vol. 620 (7972). P. 172–180.
DOI: <https://doi.org/10.1038/s41586-023-06291-2>
6. Perry N., Srivastava M., Kumar D., Boneh D. Do Users Write More Insecure Code with AI Assistants? *CCS'23: Proceedings of the 2023 ACM SIGSAC Conference on Computer and Communications Security*. P. 2785–2799.
DOI: <https://doi.org/10.1145/3576915.3623157>
7. Eloundou T., Manning S., Mishkin P., Rock D. GPTs are GPTs: An Early Look at the Labor Market Impact Potential of Large Language Models. *arXiv preprint arXiv:2303.10130v5*. 2023. URL: <https://arxiv.org/pdf/2303.10130>
8. Katz D. M., Bommarito M. J., Gao S., Arredondo P. GPT-4 passes the bar exam. *Philosophical Transactions of the Royal Society A: Mathematical, Physical and Engineering Sciences*. 2024. Vol. 382. No. 2270. Art. 20230254.
DOI: <https://doi.org/10.1098/rsta.2023.0254>
9. Dell'Acqua F. et al. Navigating the Jagged Technological Frontier: Field Experimental Evidence of the Effects of AI on Knowledge Worker Productivity and Quality. *Harvard Business School Working Paper*. 2023. No. 24-013. URL: https://www.hbs.edu/ris/Publication%20Files/24-013_d9b45b68-9e74-42d6-a1c6-c72fb70c7282.pdf
10. Valmeekam K., Sreedharan S., Marquez M., Olmo A., Kambhampati S. On the Planning Abilities of Large Language Models (A Critical Investigation with a Proposed Benchmark). *arXiv preprint arXiv:2302.06706*. 2023. URL: <https://arxiv.org/pdf/2302.06706>
11. Ludwig J., Mullainathan S., Rambachan A. Large Language Models: An Applied Econometric Framework. *arXiv preprint arXiv:2412.07031*. 2024. URL: <https://arxiv.org/pdf/2412.07031>
12. Saab K. et al. Capabilities of Gemini Models in Medicine. *arXiv preprint arXiv:2404.18416*. 2024. URL: <https://arxiv.org/pdf/2404.18416>
13. Yu Y., Gan L., Hu Y., Wei B., Kuang K., Wu F. Evaluating Test-Time Scaling LLMs for Legal Reasoning: OpenAI o1, DeepSeek-R1, and Beyond. Findings of the Association for Computational Linguistics: EMNLP 2025. P. 13759–13781.
DOI: <https://doi.org/10.18653/v1/2025.findings-emnlp.742>
14. Shetty P., Upadhayaya A., Shah P. M. et al. Advanced Financial Reasoning at Scale: A Comprehensive Evaluation of Large Language Models on CFA Level III. *arXiv preprint arXiv:2507.02954*. 2025. URL: <https://arxiv.org/pdf/2507.02954>
15. Bigeard A., Nashold L., Krishnan R., Wu S. Finance Agent Benchmark: Benchmarking LLMs on Real-world Financial Research Tasks. *arXiv preprint arXiv:2508.00828*. 2025. URL: <https://arxiv.org/pdf/2508.00828>
16. Huang K. A., Choudhary H. K., Hardin W. M., Prakash N. S. Comparative Analysis of ChatGPT-4o and Gemini Advanced Performance on Diagnostic Radiology In-Training Exams. *Cureus*. 2025. Vol. 17. No. 3. Art. e80874.
DOI: <https://doi.org/10.7759/cureus.80874>
17. Modi A. et al. Evaluating Gemini in an Arena for Learning. *arXiv preprint arXiv:2505.24477*. 2025. URL: <https://arxiv.org/pdf/2505.24477>
18. ЛУН. URL: <https://lun.ua/>

REFERENCES

- Bigeard A., Nashold L., Krishnan R. & Wu S. (2025). Finance Agent Benchmark: Benchmarking LLMs on Real-world Financial Research Tasks. <https://arxiv.org/pdf/2508.00828>
- Dell'Acqua F. (2023). Navigating the Jagged Technological Frontier: Field Experimental Evidence of the Effects of AI on Knowledge Worker Productivity and Quality. *Harvard Business School Working Paper*. https://www.hbs.edu/ris/Publication%20Files/24-013_d9b45b68-9e74-42d6-a1c6-c72fb70c7282.pdf
- Eloundou T., Manning S., Mishkin P. & Rock D. (2023). GPTs are GPTs: An Early Look at the Labor Market Impact Potential of Large Language Models. <https://arxiv.org/pdf/2303.10130>
- Google. Gemini. *Google*. <https://gemini.google.com/>
- Huang K. A., Choudhary H. K., Hardin W. M. & Prakash N. S. (2025). Comparative Analysis of ChatGPT-4o and Gemini Advanced Performance on Diagnostic Radiology In-Training Exams. *Cureus*, 3(17), Art. e80874. <https://doi.org/10.7759/cureus.80874>
- Katz D. M., Bommarito M. J., Gao S. & Arredondo P. (2024). GPT-4 passes the bar exam. *Philosophical Transactions of the Royal Society A: Mathematical, Physical and Engineering Sciences*, 2270(382), Art. 20230254. <https://doi.org/10.1098/rsta.2023.0254>
- Kojima T., Gu S. S., Reid M., Matsuo Y. & Iwasawa Y. (2022). Large Language Models are Zero-Shot Reasoners. *NIPS'22: Proceedings of the 36th International Conference on Neural Information Processing Systems, Article No. 1613*, 22199–22213. https://proceedings.neurips.cc/paper_files/paper/2022/hash/8bb0d291acd4ac06ef112099c16f326-Abstract-Conference.html
- Ludwig J., Mullainathan S. & Rambachan A. (2024). Large Language Models: An Applied Econometric Framework. <https://arxiv.org/pdf/2412.07031>
- ЛУН [LUN] <https://lun.ua/>
- Modi A. (2025). Evaluating Gemini in an Arena for Learning. <https://arxiv.org/pdf/2505.24477>
- OpenAI. ChatGPT. *OpenAI*. <https://chatgpt.com/>
- Perry N., Srivastava M., Kumar D. & Boneh D. (2023). Do Users Write More Insecure Code with AI Assistants?. *CCS'23: Proceedings of the 2023 ACM SIGSAC Conference on Computer and Communications Security*, 2785–2799. <https://doi.org/10.1145/3576915.3623157>
- Saab K. (2024). Capabilities of Gemini Models in Medicine. <https://arxiv.org/pdf/2404.18416>
- Shetty P., Upadhayaya A. & Shah P. M. (2025). Advanced Financial Reasoning at Scale: A Comprehensive Evaluation of Large Language Models on CFA Level III. <https://arxiv.org/pdf/2507.02954>

Singhal K., Azizi S. & Tu T. (2023). Large Language Models Encode Clinical Knowledge. *Nature*, 7972(620), 172–180. <https://doi.org/10.1038/s41586-023-06291-2>

Skitsko V. I. (2023). Analiz danykh iz vykorystanniam hen-eratyvnoho shtuchnoho intelektu: mozhlyvosti ta vyklyky [Data analysis using generative artificial intelligence: opportunities and challenges]. *Problemy ekonomiky*, 4, 217–225. <https://doi.org/10.32983/2222-0712-2023-4-217-225>

Valmeekam K., Sreedharan S., Marquez M., Olmo A. & Kambhampati S. (2023). On the Planning Abilities of Large Language Models (A Critical Investigation with a Proposed Benchmark). <https://arxiv.org/pdf/2302.06706>

Yu Y., Gan L., Hu Y., Wei B., Kuang K. & Wu F. (2025). Evaluating Test-Time Scaling LLMs for Legal Reasoning: OpenAI o1, DeepSeek-R1, and Beyond. *Findings of the Association for Computational Linguistics: EMNLP 2025*, 13759–13781. <https://doi.org/10.18653/v1/2025.findings-emnlp.742>

Стаття надійшла до редакції 02.12.2025 р.

Статтю прийнято до публікації 18.12.2025 р.

Оприлюднено 01.02.2026 р.