

УДК 519.237.5

С.М. ЛАПАС*

КОНФЛІКТ ВИМОГ ДО ТОЧНОСТІ КОРИСТУВАЧА І ПОКАЗНИКІВ РЕГРЕСІЙНОЇ МОДЕЛІ

*Національний технічний університет України «КПІ ім. Ігоря Сікорського», м. Київ, Україна

Анотація. У роботі виконується дослідження взаємозв'язку між різними групами показників якості регресійної моделі. Розглядаються три групи показників: статистичні, показники точності апроксимації відповідно до вимог технічних застосувань, теоретичні відносно коефіцієнтів регресії. Статистичні показники використовуються тільки ті, що найбільш поширені в прикладних дослідженнях (хоч їх замало для обґрунтованої оцінки якості регресійного рівняння), а саме: залишкова дисперсія і множинний коефіцієнт кореляції. Як оцінки точності апроксимації використовуються ті, що найчастіше висуваються як критерії адекватності в технічних застосуваннях: середнє відносне відхилення моделі від експериментальних даних, максимальне відносне відхилення, середнє абсолютне відхилення, максимальне абсолютне відхилення. Крім того, розглядається різниця між значеннями коефіцієнтів кореляції і їх оцінками. Для охоплення можливих варіантів, в яких виконується побудова моделі, дослідження проводились у таких умовах: а) відповідність усім передумовам регресійного аналізу; б) відповідність усім передумовам регресійного аналізу при наявності великого розсіяння (великої дисперсії відтворюваності); в) наявність гетероскедастичності; г) наявність викидів (із різним розміщенням і різною кількістю в навчальній вибірці). Встановлено, що при виконанні передумов і припущень регресійного аналізу різниця при застосуванні різних методів несуттєва з прикладної точки зору для показників усіх груп. При великому розсіянні і гетероскедастичності методи з метою отримання найкращих відносних показників дають значне покращення своїх характеристик при значному погіршенні всіх інших характеристик. Розроблені рекомендації для вибору методів у залежності від особливостей даних і вимог до характеристик моделі.

Ключові слова: регресійний аналіз, точність апроксимації, метод найменших квадратів, зважений метод найменших квадратів, непараметрична регресія Тейла, коефіцієнти регресії, гетероскедастичність, викиди, велике розсіяння, залишкова дисперсія, множинний коефіцієнт кореляції, середнє відносне відхилення моделі від експериментальних даних, максимальне відносне відхилення, середнє абсолютне відхилення, максимальне абсолютне відхилення.

Abstract. The paper investigates the relationship between different groups of quality indicators of the regression model. Three groups of indicators are considered: statistical indicators, indicators of approximation accuracy in accordance with the requirements of technical applications, and theoretical indicators in relation to regression coefficients. The statistical indicators are those that are the most common in applied research (although they are not enough for a reasonable assessment of the regression equation quality), namely the residual variance and the multiple correlation coefficient. Approximation accuracy estimates are those that are most often put forward as adequacy criteria in technical applications: average relative deviation of the model from experimental data, maximum relative deviation, average absolute deviation, and maximum absolute deviation. In addition, the difference between the values of the correlation coefficients and their estimates is considered. In order to cover the possible variants of the conditions under which the model is built, the research was conducted under the following conditions: a) compliance with all the prerequisites of the regression analysis; b) compliance with all the prerequisites of the regression analysis in the presence of large dispersion (large dispersion of reproducibility); c) the presence of heteroskedasticity; d) the presence of outliers (with different locations and different numbers in the training sample). It was established that when the prerequisites and assumptions of the regression analysis are met, the difference when using different methods is insignificant from an applied point of view for

the indicators of all groups. With large dispersion and heteroskedasticity, methods with the aim of obtaining the best relative indicators give a significant improvement of “their” characteristics with a significant deterioration of all other characteristics. Recommendations for the selection of methods depending on the characteristics of the data and requirements for model characteristics have been developed.

Keywords: regression analysis, accuracy of approximation, least squares, weighted least squares, non-parametric Theil regression, regression coefficients, heteroskedasticity, outliers, large dispersion, residual variance, multiple correlation coefficient, mean relative deviation of the model from experimental data, maximum relative deviation, mean absolute deviation, maximum absolute deviation.

DOI: 10.34121/1028-9763-2025-1-91-102

1. Вступ

Проблема

Конфлікт статистичної адекватності регресійної моделі і вимог адекватності з точки зору цілей математичного моделювання, сформульованих замовником.

Стан проблеми

У регресійному аналізі (РА), крім внутрішніх протиріч між його частинами [1], існують і протиріччя зовнішні — це протиріччя між статистичними методами і математичним моделюванням. А саме: при моделюванні деякого процесу, виходячи з мети моделювання, встановлюються вимоги до моделі, згідно з якими модель буде описувати процес [2]. При цьому модель може бути адекватна з точки зору регресійного аналізу і неадекватна з точки зору мети моделювання або навпаки. Найбільш частим є протиріччя в точності опису. Відносна точність апроксимації не є статистичним показником, але є поширеною вимогою в прикладних роботах. Модель може бути адекватною з точки зору регресійного аналізу, але при цьому середня точність опису не відповідає вимогам, поставленим при побудові моделі. Є інша, також дуже поширена ситуація, при якій середня відносна точність відповідає меті, але наявні окремі відхилення, що перевищують граничні вимоги. Наприклад, середня похибка апроксимації менше 3 %, але при цьому в окремих точках відхилення досягає 10–15 %, що не дозволяє, на думку прикладників, використовувати модель. Якщо фізика процесу допускає такі відхилення, наприклад, у дослідженнях міцності, то цю вимогу можна ігнорувати, але не завжди це можливо.

Статистичні показники, як правило, пересічним прикладником не сприймаються: для нього більш природне детерміноване сприйняття світу. Ніби вираз $A \pm a$ а в нього весь час перед очима, але сприймається тільки як технічні умови, а не довірчий інтервал. Тобто ніякого відношення цей запис у головах більшості інженерів не має до математичної статистики. Вихід за інтервал у певних межах із певною ймовірністю сприймається не як норма, хоч є стандарт контролю якості, в якому це фактично прописано, і хоч є процеси, в яких певна норма браку — норма, а не виняток.

Тому для інженерів природно задавати вимоги в середніх і максимальних процентах відхилення або ж в середніх і максимальних абсолютних відхиленнях.

Необхідність урахувати ці вимоги розглядалась ще з 80-х років минулого століття в рамках обговорення проблеми в літературі і на конференціях, але ніяких конкретних робіт такого плану не було. Слід звернути увагу, що в машинному навчанні популярною характеристикою є якраз точність прогнозу на контрольній послідовності дослідів [3–6]. Сама ідея використання двох вибірок для побудови регресійної моделі не нова, вона використовувалась школою Івахненка О.Г. [7–9]. У МГУА побудова моделі виконується за двома вибірками, а за третьою — модель перевіряється. Там же відбувається автоматична генерація видів моделей. Автоматизація побудови регресійної моделі, включаючи формування її структури, використовувалась в роботах Лапача С.М. [10–12] (існували програмні засоби автоматичної побудови моделі і для інших багатофакторних статистичних методів).

Разом із тим роботи по аналізу і порівнянню вказаного критерію адекватності за прикладною метою по відношенню до прийнятих статистичних характеристик моделі не виконувались.

У зв'язку з цим існує необхідність визначити, як співвідносяться вимоги користувача зі статистичними для прийняття обґрунтованих рішень про їх застосування.

Мета статті: встановити відповідність і взаємозв'язок між класами статистичних і технічних характеристик регресійної моделі для прийняття обґрунтованих рішень щодо їх застосування.

Задачі

1. Встановити взаємовплив оцінок коефіцієнтів регресії, статистичних характеристик моделі і оцінок точності моделі.
2. Визначити вплив методів і особливостей вихідних даних на вказані вище характеристики.
3. Розробити рекомендації щодо вирішення проблеми конфлікту між статистичними і модельними показниками якості.

2. Показники і методи

Статистичні

Хоча для обґрунтованої оцінки моделі необхідний комплекс статистичних показників [10, 11], будемо використовувати тільки коефіцієнт множинної кореляції і залишкову дисперсію як такі, які найчастіше зустрічаються в прикладних дослідженнях.

Коефіцієнти регресії

Для штучних прикладів визначається деформація істинних значень коефіцієнтів регресії B_0 і B_1 при отриманні їх оцінок. Якщо ми за моделлю будемо виконувати прогноз або аналіз зв'язків, то це вкрай необхідно.

Точність апроксимації

Найбільш часто висуваються такі характеристики моделі для прийняття адекватності моделі від замовника (перелічені цілі побудови):

- 1) мінімізація найбільшого за абсолютною величиною відносного відхилення
$$\min \max_i \left| \frac{y_i - \hat{y}_i}{y_i} \right| \text{ (ММВВ);}$$
- 2) мінімізація найбільшого за абсолютною величиною абсолютного відхилення
$$\min \max_i |y_i - \hat{y}_i| \text{ (ММАВ);}$$
- 3) мінімізація середнього абсолютного відхилення $\min_i \text{Aver} |y_i - \hat{y}_i| \text{ (МСАВ);}$
- 4) мінімізація середнього за абсолютною величиною відносного відхилення
$$\min_i \text{Aver} \left| \frac{y_i - \hat{y}_i}{y_i} \right| \text{ (МСВВ).}$$

Побудова моделей за вказаними цілями виконувалась методами непараметричної регресії.

Методи

Для порівняння використовувалися такі методи ідентифікації регресійних рівнянь:

- стандартний метод найменших квадратів (МНК) [13];

- зважений метод найменших квадратів (ЗМНК) із використанням як ваги обернених величин відносних похибок ($1/\varepsilon_i^2$);
- непараметрична регресія Тейла [13, 14].

Обрахування моделей виконувалось за програмами надбудови «Аналіз даних» («Регресія») Excel і макросами в середовищі Excel з PRIAM-Ex (One_Factor_RA).

3. Вибір умов застосування і методів для тестування

Для отримання повноти картини виконується порівняння як стандартних регресійних методів, так і різних варіантів вимог прикладної галузі застосування моделі. Також ці порівняння виконуються в різних умовах застосування методів.

Для порівняння вибрані такі варіанти тестування.

1. Ідеальний варіант 1: близько до прямої з невеликим розсіянням (рис. 1).

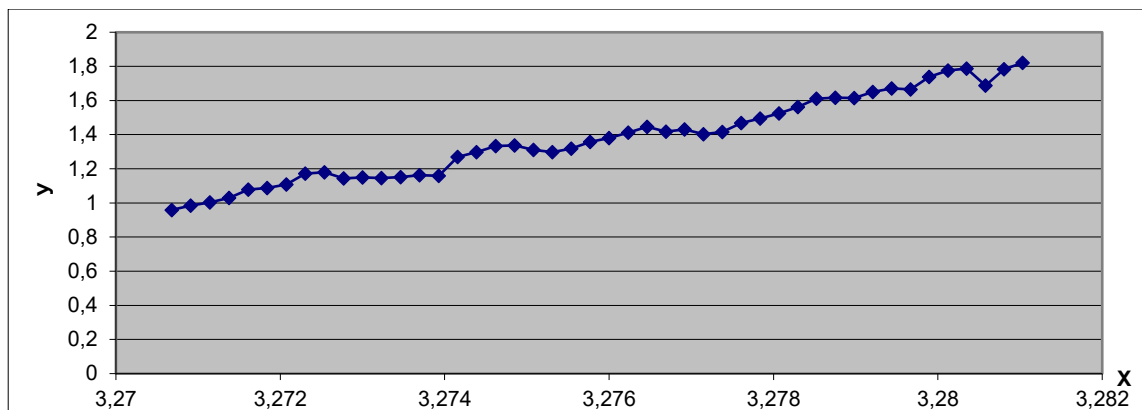


Рисунок 1 — Вихідні дані з виконанням передумов РА

2. Ідеальний варіант 2: близько до прямої з великим розсіянням. Деяким реальним процесам, не тільки біологічним, а й технічним, притаманне велике розсіяння [15] (рис. 2).

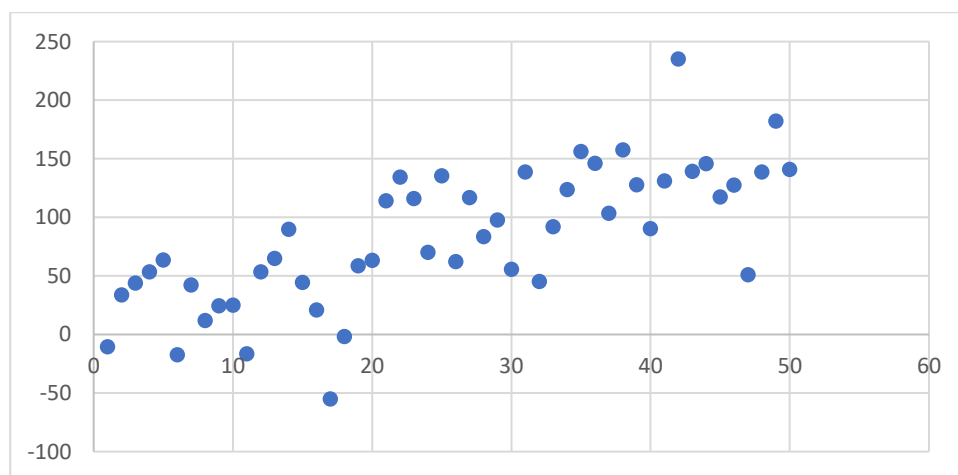


Рисунок 2 — Виконання умов РА з великим розсіянням

3. Наявність гетероскедастичності: збільшення дисперсії при збільшенні значення відгуку (рис. 3).

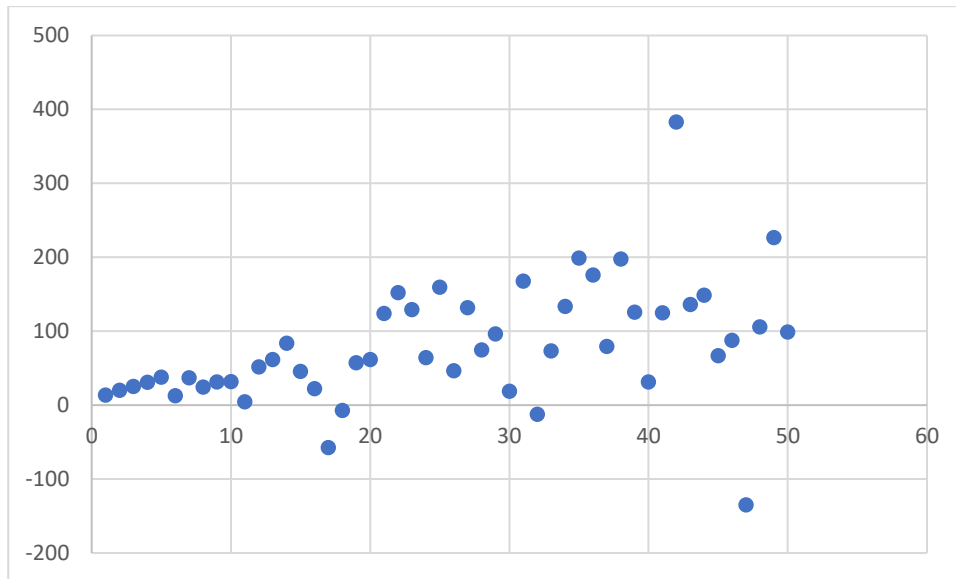


Рисунок 3 — Вихідні дані з наявністю гетероскедастичності

4. Наявність викидів. Розглядалися кілька варіантів розміщення викидів і їх кількості (рис. 4).

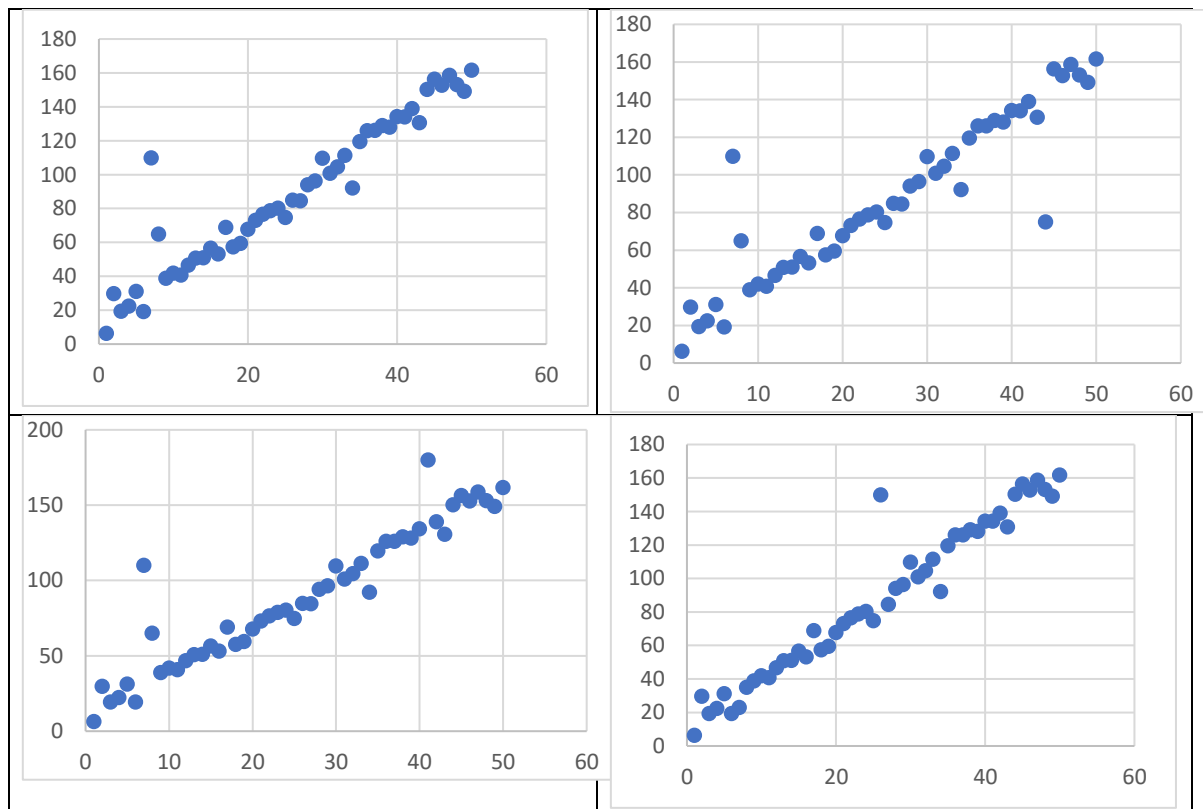


Рисунок 4 — Варіанти розміщення і кількості викидів

5. Сумнівні випадки, коли зрозуміти, чи є певні значення викидами, чи частиною вибірки, важко (рис. 5, 6). Дані запозичені з [13], аналіз на викиди виконано в [17].

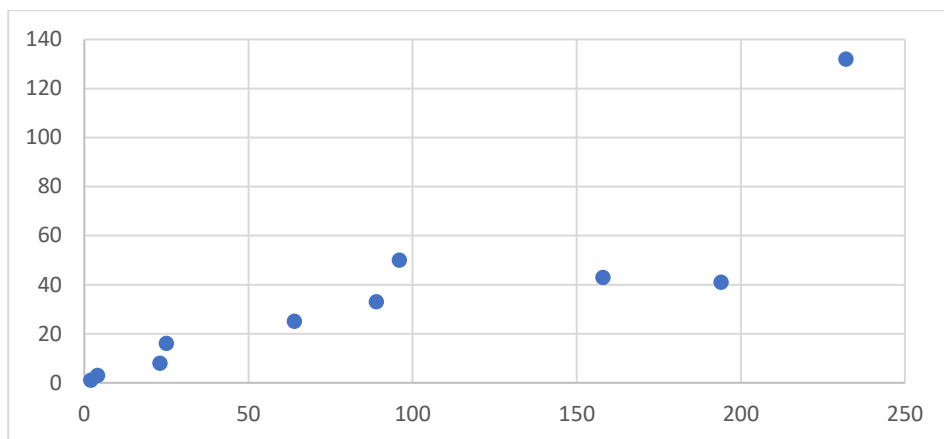


Рисунок 5 — Сумнівний варіант 1

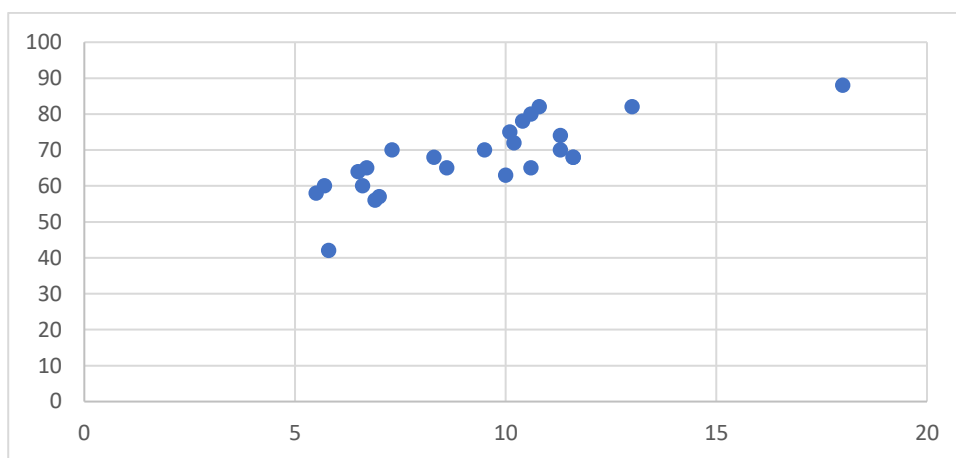


Рисунок 6 — Сумнівний варіант 2

4. Результати досліджень

У випадку виконання всіх передумов регресійного аналізу і функції, яка не має різких або швидких змін, статистичні характеристики моделей і коефіцієнти регресії, розраховані різними методами, розрізняються тільки в третьому значущому знаку. Характеристики точності наводяться в табл. 1.

Таблиця 1 — Ідеальний варіант 1

Вид моделі	Відхилення похибки від найкращого значення (%)			
	% середній	% максимальний	середнє абсолютне відхилення	максимальне абсолютне відхилення
МНК	0,90	5,73	0,80	6,45
ЗМНК	0,72	7,99	0,88	2,30
Непараметрична Тейла	0,00	15,45	0,19	3,19
МСВВ	0,00	14,24	0,15	2,15
ММВВ	11,71	0,00	14,79	35,59
МСАВ	0,00	13,37	0,00	5,76
ММАВ	0,00	11,81	0,69	0,00

Зрозуміло, що для мінімізації середніх значень як в абсолютному, так і в відносному обчисленні, вибір методу практично не впливає на результат. Винятком є мінімізація максимального абсолютного відхилення, яке значно гірше всіх інших.

Для мінімізації максимального абсолютного відхилення можна вважати рівнозначними ЗМНК, непараметричну регресію Тейла, мінімізацію максимального відносного відхилення і мінімізацію середнього відносного відхилення.

Що стосується максимального відносного відхилення, то умовно придатними, крім спеціального, є МНК і ЗМНК.

Отже, можна зробити висновок, що при виконанні передумов РА і відсутності особливостей функції відгуку найкращим варіантом є використання МНК.

5. Велике розсіяння

Для великого розсіяння найкращі статистичні показники у МНК і ЗМНК. Незначно відстають спеціальні по мінімізації абсолютних відхилень. Всі інші суттєво гірші.

Як видно з табл. 2, оцінки коефіцієнта B_0 мають занадто велике відхилення для всіх методів. І якщо для МНК і ЗМНК їх можна вважати наближенням до істинних, то всі інші — ні в якому разі. Для B_1 прийнятними можемо вважати оцінки МНК, ЗМНК і регресії Тейла.

Таблиця 2 — Велике розсіяння. Відхилення (в %) коефіцієнтів від істинного значення

Істинні значення		МНК	ЗМНК	Регресія Тейла	Мін. середн. відхилення в %	Мін. максим. % відхилення	Мін. середнього абсол. відхилення	Мін. максим. абсол. відхилення
B_0	12	35,08	72,83	308,76	203,94	193,25	111,21	218,41
B_1	3	1,36	3,67	3,72	39,86	82,64	7,54	18,00

Порівняння точності показує, що для мінімізації відносного відхилення як середнього, так і максимального, найкращі спеціальні методи (табл. 3), з якими всі інші ніяк не можуть змагатися. Але зауважимо, що ці методи, забезпечуючи найкращу користувацьку апроксимацію, абсолютно не придатні для прогнозування через великої похибки в оцінці коефіцієнтів регресії. Для мінімізації абсолютних відхилень найкращі результати мають МНК і ЗМНК, не рахуючи спеціальних.

Таблиця 3 — Велике розсіяння. Порівняння оцінок точності опису. Відхилення від найкращого значення (в %)

Вид моделі	Відхилення похибки від найкращого значення (%)			
	% середній	% максимальний	середнє абсолютне відхилення	максимальне абсолютне відхилення
МНК	39,11	54,10	0,05	0,00
ЗМНК	34,80	22,60	0,02	9,31
Непараметрична Тейла	142,21	350,47	45,66	68,15
МСВВ	0,00	10,89	85,74	87,01
ММВВ	6,70	0,00	179,04	192,72
МСАВ	29,16	2373,68	0,00	89,87
ММАВ	16,37	2068,76	3,65	7,71

Гетероскедастичність

Статистичні характеристики для МНК, ЗМНК, Тейла і МСАВ розрізняються несуттєво. Для всіх інших вони значно гірші.

Будь-які методи незадовільні щодо відтворення істинних коефіцієнтів регресії.

З аналізу табл. 4 і 5 випливає, що найкращим методом при гетероскедастичності буде МСАВ. Він дає найкращі або співставні з іншими всі показники, за винятком мінімізації максимального відносного відхилення.

Таблиця 4 — Гетероскедастичність. Відхилення від істинних значень коефіцієнтів

Істинні значення		МНК	ЗМНК	Регресія Тейла	Мін. середн. відхилення в %	Мін. максим. % відхилення	Мін. середн. абс. відхилення	Мін. максим. абс. відхилення
B0	12	31,48	26,75	68,65	7,21	9,72	11,07	1056,72
B1	3	16,11	14,20	12,49	80,17	126,70	7,06	111,24

Таблиця 5 — Гетероскедастичність. Точність апроксимації

Вид моделі	Відхилення похибки від найкращого значення (%)			
	% середній	% максимальний	середнє абсолютне відхилення	максимальне абсолютне відхилення
МНК	32,61	207,84	0,46	6,73
ЗМНК	33,44	212,72	0,40	2,85
Непараметрична Тейла	46,15	707,58	0,00	0,00
МСВВ	0,00	179,77	44,55	80,62
ММВВ	19,33	0,00	104,30	59,86
МСАВ	32,67	608,46	0,33	9,76
ММАВ	263,70	629,92	68,79	2,32

Відхилення коефіцієнта B_1 для різних методів і різних умов застосування представлено на рис. 7.

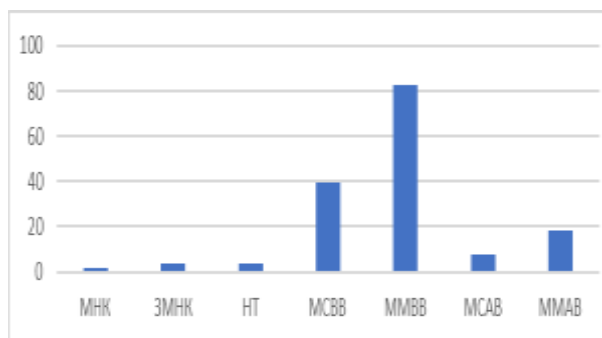


Рисунок 7.1 — B_1 при великому розсіянні

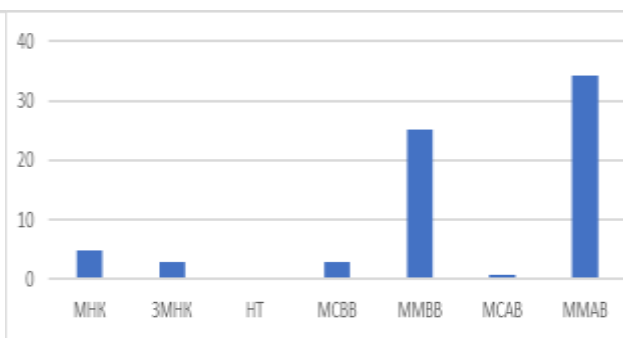


Рисунок 7.2 — B_1 при великому викиді

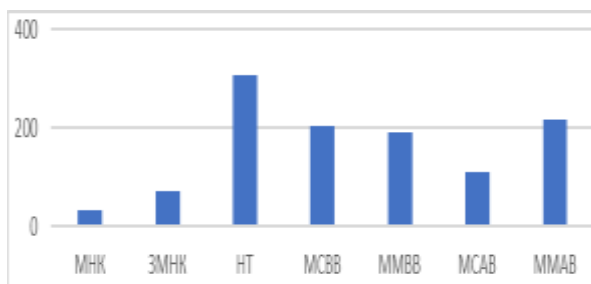


Рисунок 7.3 — В0 при великому розсіянні

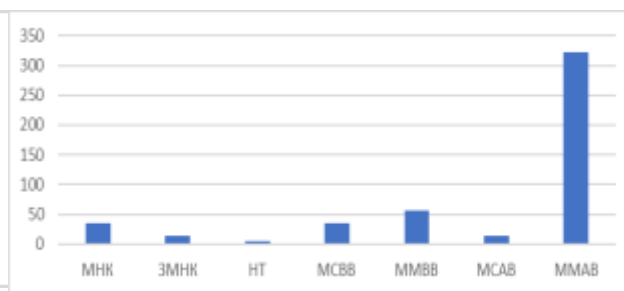


Рисунок 7.4 — В0 при великому викиді

Рисунок 7 — Відносне зміщення коефіцієнтів регресії

Результати точності апроксимації для сумнівних (реальних) варіантів представлені в табл. 6 і 7.

Таблиця 6 — Реальний варіант 1

Вид моделі	Відхилення похибки від найкращого значення (%)			
	% середній	% макси- мальний	середнє абсолютне відхилення	максимальне абсолютне відхилення
МНК	31,33	38,68	3,28	4,02
ЗМНК	6,72	102,82	1,38	11,00
Непараметрична Тейла	0,11	2,41	0,60	22,43
МСВВ	0,00	0,00	4,40	43,35
ММВВ	12,18	33,72	13,89	82,27
МСАВ	32,47	214,41	0,00	28,05
ММАВ	88,78	314,80	5,35	0,00

Таблиця 7 — Реальний варіант 2

Вид моделі	Відхилення похибки від найкращого значення (%)			
	% середній	% макси- мальний	середнє абсолютне відхилення	максимальне абсолютне відхилення
МНК	1,49	86,69	2,78	110,18
ЗМНК	25,95	20,30	32,91	32,64
Непараметрична Тейла	7,43	116,22	5,56	0,00
МСВВ	0,00	93,35	1,50	118,28
ММВВ	48,65	0,00	57,91	65,14
МСАВ	1,22	111,02	0,00	138,25
ММАВ	21,35	18,60	27,99	54,18

6. Висновки

1. При виконанні передумов і припущень регресійного аналізу різниця при застосуванні різних методів із прикладної точки зору несуттєва для показників всіх груп.
2. При великому розсіянні % точки дають значне покращення «своїх» характеристик при значному погіршенні всіх інших характеристик.
3. При гетероскедастичності % точки дають значне покращення своїх характеристик при значному погіршенні всіх інших характеристик.
4. При великому розсіянні найкраще МНК, потім ЗМНК, окрім % характеристик.

Порівняння для вибору найкращого методу побудови моделі в залежності від вимог до моделі і особливостей вихідних даних узагальнено в табл. 8–10.

Для V_1 приймалось допустимим відхилення до 6 %, а для V_0 — до 12 %.

Таблиця 8 — Вибір методів при наявності викидів

Група характеристик	Назва	Великий викид на малих	Великий викид в середині	Великий викид на великих	Баланс викидів	Два викиди на кінцях	Три викиди рівно
Статистичні	V_0	Жоден	Жоден	Жоден	НТ	НТ	Жоден
	V_1	Крім ММВВ, ММВВ, ММВВ	Крім ММВВ	Крім ММВВ, ММВВ, ММВВ	НТ, МСВВ, ММВВ	Крім МСВВ, ММВВ, ММВВ, ММВВ	Крім ММВВ, ММВВ
Моделльні коефіцієнти	R	Всі	Всі	Всі	Всі	Крім ММВВ	Крім ММВВ
	$S^2_{\text{зал}}$	Крім ММВВ	Крім ММВВ	Крім ММВВ	Крім ММВВ	Крім ММВВ, ММВВ	Крім ММВВ, ММВВ
Користувачі	Сер. % відх.	Крім ММВВ	Крім ММВВ	Крім ММВВ	Крім ММВВ	Крім ММВВ, ММВВ, ММВВ	Крім ММВВ, ММВВ, ММВВ
	Макс. % відх.	МСВВ, ММВВ	Крім ММВВ	Крім ММВВ, НТ	МСВВ, ММВВ	МСВВ, ММВВ, ММВВ	МСВВ, ММВВ, ММВВ
	Сер. абс. відх.	Крім ММВВ	Крім ММВВ	Крім ММВВ	Крім ММВВ	Крім ММВВ, ММВВ, ММВВ	Крім ММВВ, ММВВ, ММВВ
	Сер. макс. відх.	ЗМНК, ММВВ	ММВВ	ММВВ	ММВВ	МНК, ММВВ	Крім МСАВ

Можна бачити, що намагання досягти мінімального абсолютного відхилення призводить до погіршення всіх інших характеристик. Значне відхилення V_0 неможливо компенсувати жодним методом, крім непараметричної регресії Тейла і то тільки в деяких окремих специфічних випадках. У цілому найбільш придатними виявляються стандартний МНК і непараметрична регресія Тейла. Вони непридатні тільки в 14 випадках із розглянутих 48 (кожен, тільки в різних).

Таблиця 9 — Вибір методів при різних особливостях вихідних даних

Характеристики		Групи видів даних		
Група характеристик	Назви	Ідеальна	Велике розсі- яння	Гетероскедастичність
Статистичні	R	Будь-який	Будь-який, крім Тейла	МНК, ЗМНК, Тейла, МСАВ
	$S^2_{\text{зал}}$	Будь-який	МНК, ЗМНК, МСАВ, ММВВ	МНК, ЗМНК, Тейла, МСАВ

Продовження таблиці 9

Модельні коефіцієнти	V_0	Будь-який	МНК	МСВВ
	V_1	Будь-який	МНК, ЗМНК, Тейла	МСАВ
Користувацькі	Сер. % відх.	Будь-який, крім МСВВ	МСВВ, ММВВ	МСВВ
	Макс. % відх.	ММАВ, МНК, ЗМНК	ММВВ	МСАВ
	Сер. абс. відх.	Будь-який, крім ММВВ	МНК, ЗМНК, МСАВ, ММАВ	МНК, ЗМНК, Тейла, МСАВ
	Сер. макс. відх.	Будь-який, крім ММВВ	МНК, ММАВ	МНК, ЗМНК, Тейла, ММАВ
У сукупності		МНК	МНК, крім відносних користувачьких характеристик	Не визначено

Таблиця 10 — Вибір методів при невизначених відхиленнях від передумов РА

Характеристики		Група видів даних	
Група характеристик	Назви	ДС741	ДС743
Статистичні	R	Крім ЗМНК, ММВВ	Крім ММВВ
	$S^2_{\text{зал}}$	Крім МСВВ, ММВВ	Крім ММВВ
Користувацькі	Сер. % відх.	ЗМНК, Тейла, МСВВ	МНК, МСВВ, МСАВ, Тейла
	Макс. % відх.	Тейла, МСВВ	ММВВ
	Сер. абс. відх.	Крім ММАВ	МНК, МСВВ, МСАВ, Тейла
	Сер. макс. відх.	МНК, ММАВ	ММАВ

При сумнівних даних найкраще використовувати непараметричну регресію Тейла, за винятком необхідності досягнення для моделі певних характеристик точності.

СПИСОК ДЖЕРЕЛ

1. Лапач С.М., Радченко С.Г. Основні проблеми побудови регресійних моделей. *Математичні машини і системи*. 2012. № 4. С. 125–133.
2. Лапач С.М. Регресійний аналіз. Процесний підхід. *Математичні машини і системи*. 2016. № 1. С. 129–138.
3. Kelleher J.D., Namee, B.M., D'Arcy A. Fundamentals of Machine Learning for predictive data analysis. Algorithms, worked Examples, and case studies. Cambridge: The MIT Press, 2015. 624 p.
4. Kuhn M., Johnson K. Applied Predictive Modelling. New York: Springer, 2013. 612 p.
5. Hastie T., Tibshirani R., Friedman J. The elements of Statistical Learning. Data Mining, Inference and Prediction. Second Edition. Springer, 2009. 764 p.
6. Brunton S.L., Kurtz J.N. Data-Driven Science and engineering. Machine Learning, Dynamical Systems, and Control. Cambridge: Cambridge university Press, 2022. 574 p.

7. Müller J.A., Ivachnenko A.G. Selbstorganisation von vorhersage Modellen. Berlin: Verlag Technik, 1984. 222 p.
8. URL: https://uk.wikipedia.org/wiki/Метод_групового_урахування_аргументів.
9. URL: https://en.wikipedia.org/wiki/Group_method_of_data_handling.
10. Лапач С.М. Теорія планування експериментів: Виконання розрахунково-графічної роботи: навч. посіб. для студ. спеціальності 131 «Прикладна механіка», спеціалізації «Технологія машинобудування». Київ: КПІ ім. Ігоря Сікорського, 2020. 86 с. URL: <https://ela.kpi.ua/handle/123456789/38858>.
11. Лапач С.Н., Чубенко А.В., Бабич П.Н. Статистические методы в медико-биологических исследованиях с использованием Excel. 2 изд. перераб. и доп. К.: Морион, 2001. 408 с.
12. Лапач С.М., Радченко С.Г. Проблеми визначення структури рівняння регресії в множинному регресійному аналізі. *Наукові вісті НТУУ «КПІ»*. 2007. № 1 (51). С. 150–155.
13. Draper N.R., Smith H. Applied regression Analysis. Third edition. New York: Wiley-Interscience, 1998. 736 p.
14. Hettmansperger T.P. Statistical inference based on ranks. New York: Wiley, 1984. 323 p.
15. Лапач С.М. Застосування непараметричної регресії при наявності викидів. *Теорія ймовірностей та математична статистика*: Шістнадцята міжнар. конф. ім. акад. Михайла Кравчука (м. Київ, 14–15 травня 2015 р.). К.: НТУУ «КПІ», 2015. Т. 3. С. 43–45.
16. Лапач С.М. Проблеми побудови регресійних моделей процесів різання металів. *Вісник НТУУ «КПІ». Машинобудування*. 2014. № 3 (72). С. 40–47.
17. Лапач С.М. Визначення викидів у кореляційному і одновимірному регресійному аналізі. *Математичні машини і системи*. 2019. № 4. С. 126–138.

Стаття надійшла до редакції 20.01.2025