

Лабораторна робота № 11

ДОСЛІДЖЕННЯ НАЯВНОСТІ МУЛЬТИКОЛІНЕАРНОСТІ МОДЕЛІ ЗГІДНО АЛГОРИТМУ ФАРРАРА-ГЛОБЕРА. УСУНЕННЯ МУЛЬТИКОЛІНЕАРНОСТІ

Мета роботи – навчитися перевіряти наявність мультиколінеарності згідно алгоритму Фаррара-Глобера та усунення цього явища засобами Microsoft Excel.

Приклад виконання лабораторної роботи.

Завдання.

За допомогою табличного редактора Microsoft Excel дослідити наявність мультиколінеарності між незалежними змінними (фондовіддачею, продуктивністю праці та питомими інвестиціями), скориставшись формальними ознаками, алгоритмом Фаррара-Глобера та усунити це явище в моделі (таблиця 1).

Порядок виконання роботи:

- 1) перевірити адекватність моделі за допомогою F-критерію Фішера;
- 2) перевірити виконання умови – «високе значення коефіцієнта детермінації R^2 і не значимість t – статистики»;
- 3) перевірка значень парних коефіцієнтів кореляції на перевищення 0,85;
- 4) перевірити мультиколінеарність згідно алгоритму Фаррара-Глобера:
 - 4.1) перевірка всього масиву незалежних змінних (критерій χ^2);
 - 4.2) перевірка кожної незалежної змінної з усіма іншими (F-критерій);
 - 4.3) перевірка кожної пари незалежних змінних (t-критерій).
- 5) побудувати адекватну модель без мультиколінеарності.
- 6) зробити висновки.

Таблиця 1

Вихідні дані

Місяць	Прибуток, тис. грн.	Фондовіддача, тис.грн. / тис.грн.	Продуктивність праці, тис.грн. /люд.-год.	Питомі інвестиції, %
	Y	x1	x2	x3
1	40	12	5	15
2	45	17	7	18
3	40	13	6	16
4	43	14	7	17
5	48	16	6	20
6	39	15	5	15
7	42	14	6	16
8	45	17	9	18
9	38	12	5	19
10	48	18	10	20
11	50	20	11	22
12	48	17	10	21
13	49	18	12	21
14	45	19	8	20
15	49	20	9	22
16	52	22	14	23
17	54	24	15	24

18	51	21	13	20
19	55	25	16	24
20	56	27	18	25

РОЗВ'ЯЗАННЯ.

1) перевірити адекватність моделі за допомогою F-критерію Фішера;

Виконуємо специфікацію економетричної моделі.

На етапі специфікації економетричної моделі виконуємо ідентифікацію змінних моделі та вводимо наступні умовні позначення для залежної (пояснювальної) змінної та незалежних (пояснюючих) змінних моделі:

y – прибуток, тис. грн.;

x_1 – фондівіддача, тис.грн. / тис.грн.;

x_2 – продуктивність праці, тис.грн. /люд.-год.;

x_3 – питомі інвестиції, %.

За допомогою надбудови «Аналіз даних / Data Analytics» інструменту аналізу «Регресія / Regression» проведемо попередній економіко-математичний аналіз моделі. В результаті отримали наступні значення (рис. 1):

SUMMARY OUTPUT																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																											</
----------------	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	----

Рис. 1. Результати застосування інструменту «Регресія / Regression»

За результатами розрахунків (рис. 1) було отримане таке оцінене вибіркове рівняння регресії

$$\hat{y} = 23,3 + 0,485x_1 + 0,410 x_2 + 0,551 x_3 \quad (1)$$

Перевіримо цю модель на адекватність. Оскільки розраховане значення F-критерію Фішера (67,65) є більшим критичне (3,239), отже виконується умова $F_p > F_{\alpha,k1,k2}$ робимо висновок про адекватність моделі.

Крім того, значення «Significance F»= 0,000000002625, що значно менше ніж 0,05, що також є свідченням адекватності отриманої моделі (1).

2) перевірити виконання умови – «високе значення коефіцієнта детермінації R^2 і незначущість t – статистики»;

З рис. 1, видно, коефіцієнт детермінації $R^2=0,9269$. Оскільки воно дуже близьке до 1, то таке значення є високим.

Але значення t –статистики для всіх змінних, крім вільного члена a_0 , (на рис. 1 виокремлено фіолетовим) менше за t -критичне=2,12. Це означає, що всі отримані оцінки параметрів a_i в моделі (1) є статистичні незначущими і їх не варто використовувати для економічної інтерпретації та аналізу.

Таким чином одночасно виконуються дві умови – високе значення коефіцієнта детермінації R^2 (що є свідченням якості моделі в цілому) та незначущість t –статистики (що вказує на незначущість отриманих оцінок параметрів a_i в моделі (1)). З чого робимо висновок про наявність мультиколінеарності в моделі.

3) перевірка значень парних коефіцієнтів кореляції на перевищення 0,8;

Для побудови кореляційної матриці використаємо інструмент «Кореляція /Correlation».

	Y	x1	x2	x3
Y	1			
x1	0,943461	1		
x2	0,933945	0,941919	1	
x3	0,915723	0,891652	0,875871	1

Рис. 2. Результати застосування інструменту «Кореляція / Correlation»

Всі отримані коефіцієнти парної кореляції між незалежними змінними x_1 , x_2 , x_3 (на рис. 2 виокремлені зеленим) більші за 0,8, що є свідченням щільного взаємозв'язку між змінними в моделі та наявністю в них мультиколінеарності.

4) перевірити мультиколінеарність згідно алгоритму Фаррара-Гребера. Даний метод заснований на застосуванні трьох видів статистичних критеріїв:

Порівнявши ці критерії з їх критичними значеннями, можна зробити висновок щодо наявності чи відсутності мультиколінеарності між незалежними змінними.

4.1) перевірка всього масиву незалежних змінних (критерій χ^2);

Треба побудувати кореляційну матрицю незалежних змінних. Використаємо дані, отримані на попередньому кроці (рис. 2). Оскільки ця матриця є симетричною відносно головної діагоналі, заповнимо верхню частину.

	x1	x2	x3
x1	1	0,941919	0,891652
x2	0,941919	1	0,875871
x3	0,891652	0,875871	1

Рис. 3. Кореляційна матриця R

Знайдемо визначник кореляційної матриці **R**. Для цього скористаємось функцією МОПРЕД (MDETERM).

$$\det(R)=0,0218$$

Чим ближче значення $\det(R)$ до нуля, тим сильнішою є мультиколінеарність в моделі. Оскільки в нашому випадку $\det(R)$ дуже близьке до 0, то в масиві незалежних змінних може існувати мультиколінеарність.

Обчислимо критерій Пірсона χ^2 (хі-квадрат) за формулою:

$$\chi^2_{\text{факт}} = -(n-1-1/6*(2m+5))*\ln(\det(R))$$

де, n – обсяг сукупності, або кількість спостережень; m – кількість незалежних змінних в моделі. В нашому випадку $n=20$, $m=3$.

Маємо:

$$\chi^2_{\text{факт}} = -(20-1-1/6*(2*3+5))*\ln(0,0218)=65,66$$

Визначимо табличне χ^2 . Табличне або критичне значення χ^2 визначають за допомогою спеціальних таблиць або за допомогою функції ХІ2ОБР (CHIINV) (рис. 4).

Задаємо рівень значущості $\alpha=0,05$ та обчислюємо ступень свободи $k=1/2*m*(m-1)=1/2*3*(3-1)=3$.

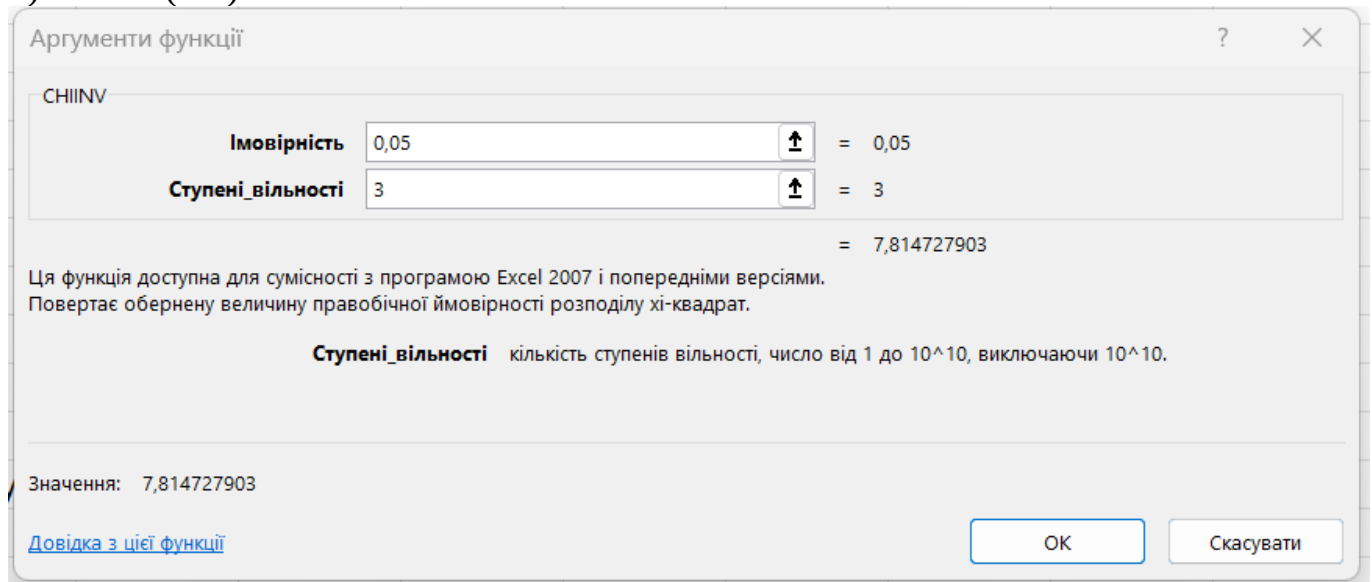


Рис. 4. Діалогове вікно функції ХІ2ОБР (CHIINV)

$$\chi^2_{\text{табл}}=7,814.$$

Оскільки виконується умова $\chi^2_{\text{факт}} > \chi^2_{\text{табл}}$ ($65,66 > 7,814$), то в масиві незалежних змінних (фондовіддача, продуктивність праці та питомі інвестиції) існує мультиколінеарність і необхідно проводити подальше дослідження.

4.2) перевірка кожної незалежної змінної з усіма іншими (F-критерій);

Необхідно визначити матрицю похибок **B** або помилок, яку визначають як обернену до кореляційної матриці **R**. Для цього скористаємось функцією МОБР (MINVERSE). В результаті отримали наступне (рис. 5):

	10,6712	-7,37597	-3,0546
B=	-7,3759	9,3929	-1,65019
	-3,0546	-1,65019	5,168995

Рис. 5. Матриця помилок **B**

Значення F-критеріїв обчислюємо за формулою:

$$F_j = (b_{jj} - 1) \cdot \frac{n-m}{m-1},$$

де b_{ij} – діагональні елементи матриці **B**; n – обсяг сукупності, або кількість спостережень; m – кількість незалежних змінних в моделі.

$$F_1 = (10,6712 - 1) \cdot (20 - 3) / (3 - 1) = 82,2$$

$$F_2 = (9,3929 - 1) \cdot (20 - 3) / (3 - 1) = 71,3$$

$$F_3 = (5,168995 - 1) \cdot (20 - 3) / (3 - 1) = 35,4$$

Обчислимо $F_{\text{критичне}}$ за допомогою функції F.ОБР.ПХ (F.INV.RT).

Задаємо рівень значущості $\alpha = 0,05$ та обчислюємо ступені свободи $k_1 = m - 1 = 3 - 1$, $k_2 = n - m = 20 - 3$.

$$F_{\text{критичне}} = 3,592.$$

Обчислені значення F-критеріїв порівнюємо з табличним значенням $F_{\text{критичне}}$. Якщо обчислений F-критерій для певної незалежної змінної x_i більше за табличне (критичне) значення, то це означає, що ця змінна x_i , для якої обчислений критерій залежить від всіх інших незалежних змінних в масиві, тобто необхідно розглядати питання про її виключення з переліку змінних.

Оскільки виконується умова $F_1 > F_{\text{критичне}}$ ($82,2 > 3,592$), то незалежна змінна x_1 мультиколінеарна з іншими змінними x_2 та x_3 , тобто пояснюючі змінні x_2 та x_3 впливають на цю змінну внаслідок суттєвої кореляції між ними. Аналогічно у випадку $F_2 > F_{\text{критичне}}$ ($71,3 > 3,592$) маємо, що змінна x_2 мультиколінеарна зі змінними x_1 та x_3 . Також при $F_3 > F_{\text{критичне}}$ ($35,4 > 3,592$) змінна x_3 мультиколінеарна зі змінними x_1 та x_2 .

Тому в нашому прикладі кожна з незалежних змінних мультиколінеарна з усіма іншими.

Використовуючи матрицю помилок **B** визначимо значення часткових коефіцієнтів детермінації для кожної незалежної змінної за формулою:

$$R^2_{x_j} = 1 - \frac{1}{b_{jj}}$$

$$R^2_{x_1} = 1 - 1/10,6712 = 0,9063$$

$$R^2_{x_2} = 1 - 1/9,3929 = 0,8935$$

$$R^2_{x_3} = 1 - 1/5,168995 = 0,8065$$

Оскільки коефіцієнти детермінації дуже близькі до одиниці, то кожна з незалежних змінних є мультиколінеарною з іншими. Але найбільш лінійно залежною з іншими змінними є змінна x_1 , оскільки їй відповідає найбільше значення коефіцієнта детермінації $R^2_{x_1} = 0,9063$.

4.3) перевірка кожної пари незалежних змінних (t-критерій).

Визначимо частинні коефіцієнти кореляції між пояснюючими змінними r_{ij} . Вони показують щільність зв'язку між двома незалежними змінними за умови, що інші змінні не впливають на цей зв'язок. Їх обчислюють за формулою:

$$r_{ij}^* = \frac{-b_{ij}}{\sqrt{b_{ii} \cdot b_{jj}}},$$

Де, де b_{ij} – елементи матриці помилок **B**, що містяться на перетині i -го рядку та j -го стовпця; b_{ii} та b_{jj} – діагональні елементи матриці помилок **B**.

Отримали такі значення частинних коефіцієнтів кореляції r_{ij} :

$$r_{12} = (-7,37597) / (\text{SQRT}(10,6712 \cdot 9,3929)) = 0,73674$$

$$r_{13} = (-3,0546) / (\text{SQRT}(10,6712 \cdot 5,168995)) = 0,41129$$

$$r_{23} = (-1,65019) / (\text{SQRT}(9,3929 \cdot 5,168995)) = 0,23683$$

Порівнявши частинні коефіцієнти кореляції з парними з матриці **R** (рис. 3), варто відзначити, що частинні коефіцієнти значно менше парних.

Враховуючи отримані значення частинних коефіцієнтів кореляції, можна зробити висновок, що зв'язок між фондовіддачею та продуктивністю праці є сильним, якщо не враховувати вплив питомих інвестицій $r_{12} = 0,73674$; зв'язок між фондовіддачею та питомими інвестиціями є помірним, якщо не брати до уваги вплив продуктивності праці $r_{13} = 0,41129$; зв'язок між продуктивністю праці та питомими інвестиціями є слабким, якщо не враховувати фондовіддачу $r_{23} = 0,23683$.

Перевіряємо статистичну значущість отриманих частинних коефіцієнтів кореляції r_{ij} за допомогою t-критерію Стюдента.

Визначимо значення t-критеріїв Стюдента для визначення попарної мультиколінеарності двох незалежних змінних за формулою:

$$t_{ij} = \frac{|r_{ij}| \sqrt{n-m}}{\sqrt{1-r_{ij}^2}}.$$

Отже, маємо:

$$t_{12} = (\text{ABS}(0,73674)) \cdot \text{SQRT}(20-3) / (\text{SQRT}(1-0,73674 \cdot 0,73674)) = 4,492$$

$$t_{13} = (\text{ABS}(0,41129)) \cdot \text{SQRT}(20-3) / (\text{SQRT}(1-0,41129 \cdot 0,41129)) = 1,860$$

$$t_{23} = (\text{ABS}(0,23683)) \cdot \text{SQRT}(20-3) / (\text{SQRT}(1-0,23683 \cdot 0,23683)) = 1,005$$

Обчислені значення t-критеріїв Стюдента порівнюємо з табличним $t_{\text{критичне}}$ для рівня значущості $\alpha = 0,05$ і ступеня свободи $k = n - m = 20 - 3 = 17$. Використаємо функцію СТЬЮДЕНТ.ОБР.2Х (СТЬЮДЕНТРАСПОБР, Т.INV.2Т).

$$t_{\text{критичне}} = 2,110.$$

Якщо $t_{ij} > t_{\text{критичне}}$, тоді отриманий коефіцієнт кореляції є статистично значущим, а між незалежними змінними x_i та x_j існує мультиколінеарність.

Оскільки $t_{12} > t_{\text{критичне}}$ ($4,492 > 2,110$), тому дві незалежні змінні фондовіддача (x_1) та продуктивність праці (x_2) утворюють колінеарну пару, тобто між незалежними

змінними (x_1) та (x_2) існує кореляційний зв'язок тобто є відповідно мультиколінеарними між собою;

$t_{13} < t_{\text{критичне}} (1,860 > 2,110)$, тому відповідно фондівдача та питомі інвестиції не є мультиколінеарними між собою;

$t_{23} < t_{\text{критичне}} (1,005 > 2,110)$, тому продуктивність праці та питомі інвестиції не є мультиколінеарними між собою.

Отже між іншими парами змінних x_1 та x_3 , x_2 та x_3 мультиколінеарний зв'язок відсутній.

Висновки: дослідження показали, що в заданому масиві даних мультиколінеарність існує між фондівдачею та продуктивністю праці. Наявність мультиколінеарності означає порушення припущення класичного лінійного регресійного аналізу про незалежність між пояснюючими змінними в моделі. Тому з метою усунення цього явища пропонується вилучення з економетричної моделі однієї зі змінних, що утворюють колінеарну пару.

Отже, для того, щоб можна було застосувати метод 1 МНК для оцінювання параметрів моделі за цією інформацією, необхідно звільнитися від мультиколінеарності. Якщо пара факторів мультиколінеарні, то один з цих факторів залишають, а інший вилучають з моделі. Найчастіше залишають той фактор, який з економічної точки зору є більш вагомим. Якщо із економічних міркувань жодному з них не можна надати перевагу, то залишають той фактор, який має більший коефіцієнт кореляції із залежною змінною, а потім за допомогою класичного методу найменших квадратів будують відповідну модель.

5) побудувати адекватну модель без мультиколінеарності.

Отримані результати використовуємо для подальшого аналізу та побудови моделі. В результаті застосування інструменту «Регресія / Regression» (рис. 1) отримали такі дані:

	<i>Coefficients</i>
Intercept	23,30887089
x1	0,482526065
x2	0,410496492
x3	0,550038753

Та таку модель:

$$\hat{y} = 23,3089 + 0,4825x_1 + 0,4105x_2 + 0,5500x_3 \quad (1)$$

Всі показники t-критеріїв Стюдента для параметрів менші за $t_{\text{критичне}}$, крім a_0 .

<i>t Stat</i>
6,569
1,730
1,460
2,019

Одним з найбільш поширених способів усунення мультиколінеарності є виключення незалежних факторів з моделі.

За звичай, з двох факторів, пов'язаних значною кореляцією, виключається, насамперед, фактор на підставі теоретичних міркувань. Якщо такий підхід не дає результату, то виключається той фактор, якому відповідає менший коефіцієнт кореляції з функцією.

Таким чином, наявність мультиколінеарності між змінними дає статистично незначимі оцінки параметрів моделі. Оскільки в нашому випадку всі параметри є статистично незначущими, то для усунення мультиколінеарності виключенню може підлягати будь-яка з цих змінних.

На попередніх етапах було встановлено, що мультиколінеарність існує між фондовіддачею та продуктивністю праці. Отже, в залежності від мети дослідження з моделі варто вилучити ту змінну, яка є менш важливою з точки зору мети дослідження. Ми спробуємо обидва варіанти.

5.1) вилучаємо змінну x_1 .

З умови задачі прибираємо дані для змінної x_1 (рис. 6) та будуємо нову модель (2) без змінної x_1 використовуючи інструмент «Регресія / Regression» (рис. 7).

Місяць	Прибуток, тис. грн.	Продуктивність праці, тис.грн. /люд.-год.	Питомі інвестиції, %
	Y	x2	x3
1	40	5	15
2	45	7	18
3	40	6	16
4	43	7	17
5	48	6	20
6	39	5	15
7	42	6	16
8	45	9	18
9	38	5	19
10	48	10	20
11	50	11	22
12	48	10	21
13	49	12	21
14	45	8	20
15	49	9	22
16	52	14	23
17	54	15	24
18	51	13	20
19	55	16	24
20	56	18	25

Рис. 6. Дані для нової моделі (2) без змінної x_1

SUMMARY OUTPUT								
Regression Statistics								
Multiple R	0,9556							
R Square	0,9133							
Adjusted R Square	0,9030							
Standard Error	1,6761							
Observations	20							
ANOVA								
	<i>df</i>	<i>SS</i>	<i>MS</i>	<i>F</i>	<i>Significance F</i>			
Regression	2	502,7913369	251,396	89,48588767	0,00000000094			
Residual	17	47,75866312	2,80933					
Total	19	550,55						
	<i>Coefficients</i>	<i>Standard Error</i>	<i>t Stat</i>	<i>P-value</i>	<i>Lower 95%</i>	<i>Upper 95%</i>	<i>Lower 95,0%</i>	<i>Upper 95,0%</i>
Intercept	24,741	3,647	6,784	0,0000032	17,046	32,436	17,046	32,436
x2	0,769	0,201	3,826	0,0013517	0,345	1,193	0,345	1,193
x3	0,744	0,262	2,835	0,0114410	0,190	1,298	0,190	1,298
		СТЬЮДЕНТ.ОБР.2X	2,11					

Рис. 7. Результати застосування інструменту «Регресія / Regression» для нових даних моделі (2) без змінної x_1

В результаті отримаємо модель (2):

$$\hat{y} = 24,74 + 0,769 x_2 + 0,744 x_3 \quad (2)$$

Протестуємо отриману модель (2) на статистичну якість. За F-критерієм Фішера модель є адекватною, оскільки $F_{\text{фактичне}} > F_{\text{критичне}}$, про що свідчить значення $\text{Significance F} = 0,00000000094$, що значно менше за 0,05 (див. рис. 7).

Отримані оцінки параметрів в моделі (2) є статистично значущими, оскільки всі значення $P\text{-value}$ також менші за 0,05 та значення t-статистики (t Stat на рис. 7), що більші за $t_{\text{критичне}}$ (2,11) (див. рис. 7). Отже, можна зробити висновок, що за моделлю (2) можна робити економічний аналіз параметрів та прогнозування.

5.2) вилучаємо змінну x_2 .

З умови задачі прибираємо дані для змінної x_2 (рис. 8) та будуємо нову модель (3) без змінної x_2 використовуючи інструмент «Регресія / Regression» (рис. 9).

Місяць	Прибуток, тис. грн.	Фондовіддача, тис.грн. / тис.грн.	Питомі інвестиції, %
	Y	x1	x3
1	40	12	15
2	45	17	18
3	40	13	16
4	43	14	17
5	48	16	20
6	39	15	15
7	42	14	16
8	45	17	18
9	38	12	19
10	48	18	20
11	50	20	22
12	48	17	21
13	49	18	21
14	45	19	20
15	49	20	22
16	52	22	23
17	54	24	24
18	51	21	20
19	55	25	24
20	56	27	25

Рис. 8. Дані для нової моделі (3) без змінної x_2

SUMMARY OUTPUT								
<i>Regression Statistics</i>								
Multiple R	0,958							
R Square	0,917							
Adjusted R Square	0,907							
Standard Error	1,638							
Observations	20							
ANOVA								
	<i>df</i>	<i>SS</i>	<i>MS</i>	<i>F</i>	<i>Significance F</i>			
Regression	2	504,9572627	252,4786	94,1407993	0,000000000637			
Residual	17	45,59273731	2,681926					
Total	19	550,55						
	<i>Coefficients</i>	<i>Standard Error</i>	<i>t Stat</i>	<i>P-value</i>	<i>Lower 95%</i>	<i>Upper 95%</i>	<i>Lower 95,0%</i>	<i>Upper 95,0%</i>
Intercept	19,970	2,802	7,13	0,000001694	14,0585	25,8808	14,0585	25,8808
x1	0,783	0,195	4,02	0,000892071	0,3716	1,1935	0,3716	1,1935
x3	0,644	0,273	2,36	0,030658540	0,0676	1,2208	0,0676	1,2208
		СТЬЮДЕНТ.ОБР.2X		2,11				

Рис. 9. Результати застосування інструменту «Регресія / Regression» для нових даних моделі (3) без змінної x_2

В результаті отримаємо модель (3):

$$\hat{y} = 19,97 + 0,783 x_1 + 0,644 x_3 \quad (3)$$

Протестуємо отриману модель (3) на статистичну якість. За F-критерієм Фішера модель є адекватною, оскільки $F_{\text{фактичне}} > F_{\text{критичне}}$, про що свідчить $\text{SignificanceF}=0,000000000637$, що значно менше за 0,05 (див. рис. 9).

Отримані оцінки параметрів в моделі (3) є статистично значущими, оскільки всі значення $P\text{-value}$ також менші за 0,05 та значення t-статистики (t Stat на рис. 9), що більші за $t_{\text{критичне}}$ (2,11) (див. рис. 9). Отже, можна зробити висновок, що за моделлю (3) також можна робити економічний аналіз параметрів та прогнозування.

5.3) вилучення змінної на основі оцінки лінійного зв'язку з залежною змінною у.

Якщо, з двох факторів, між якими існує мультиколінеарність, обрати той, що виключається на підставі теоретичних міркувань, не має можливості, то виключається той фактор, якому відповідає менший коефіцієнт кореляції з функцією.

Це пояснюється тим, що мірою кореляційного зв'язку між змінними є коефіцієнт парної кореляції. Тому, доцільним є визначення коефіцієнта кореляції між залежною змінною у (прибутком) та кожною пояснюючою змінною, які склали мультиколінеарну пару.

Отже, подальший аналіз передбачає виключення з моделі мультиколінеарних факторів на основі оцінки лінійного зв'язку з залежною змінною у. Для визначення того, яку саме змінну будемо виключати визначимо коефіцієнти кореляції між кожною незалежною змінною x_i та залежною змінною у. Кожний з цих коефіцієнтів буде вказувати як впливає кожна зі змінних на залежну змінну у.

Ці дані можна взяти з кореляційної матриці R (рис. 2):

	у
у	1
x1	0,943461
x2	0,933945
x3	0,915723

Або можна використати функцію КОРЕЛЛ (CORREL).

$$r_{x1y}=0,943$$

$$r_{x2y}=0,934$$

$$r_{x3y}=0,916$$

Серед двох змінних, між якими існує мультиколінеарність (фондовіддача та продуктивність праці) менший коефіцієнт кореляції має друга змінна (продуктивність праці), хоча різниця є несуттєвою. Отже, у нас виходить, що найменше на залежну змінну у впливає x_2 . В цьому випадку отримаємо модель (3), яка була отримана нами на етапі 5.2.

Висновки загальні:

1) В процесі дослідження на основі наявних даних (табл. 1) було побудовано економетричну модель впливу фондовіддачі, продуктивності праці та питомих інвестицій на прибуток підприємства:

$$\hat{y} = 23,3089 + 0,4825x_1 + 0,4105 x_2 + 0,5500 x_3 \quad (1)$$

Тестування моделі на адекватність показало, що модель є адекватною. Оскільки розраховане значення F-критерію Фішера (67,65) є більшим критичне (3,239), отже виконується умова $F_p > F_{\alpha, k_1, k_2}$ та значення «Significance F» = 0,000000002625, що значно менше ніж 0,05.

2) перевірка зовнішніх ознак на наявність мультиколінеарності показала, що в моделі присутня мультиколінеарність. А саме:

2.1) одночасно виконуються дві умови – високе значення коефіцієнта детермінації (близьке до одиниці) $R^2 = 0,9269$ (що є свідченням якості моделі в цілому) та незначущість t-статистики (виконується умова $t_{\text{фактичне}} < t_{\text{критичне}} - 1,73 < 2,12; 1,46 < 2,12; 2,02 < 2,12$) що вказує на незначущість отриманих оцінок параметрів a_i в моделі (1). З чого робимо висновок про наявність мультиколінеарності в моделі.

2.2) Всі отримані коефіцієнти парної кореляції між незалежними змінними x_1, x_2, x_3 (0,95, 0,89 та 0,88) більші за 0,85, що є свідченням щільного взаємозв'язку між змінними в моделі та наявністю в них мультиколінеарності.

3) тестування моделі (1) алгоритмом Фаррара-Глобера показало:

3.1) перевірка всього масиву незалежних змінних (критерій χ^2) довела, що оскільки виконується умова $\chi^2_{\text{факт}} > \chi^2_{\text{табл}}$ (65,66 > 7,814), то в масиві незалежних змінних (фондовіддача, продуктивність праці та питомі інвестиції) існує мультиколінеарність і необхідно проводити подальше дослідження.

3.2) перевірка за F-критеріями, показала, що для всіх змінних моделі x_1, x_2 та x_3 виконується нерівність $F_{i \text{ розрахункове}} > F_{\text{критичне}}$, а саме: $F_1 > F_{\text{критичне}}$ (82,2 > 3,592); $F_2 > F_{\text{критичне}}$ (71,3 > 3,592); $F_3 > F_{\text{критичне}}$ (35,4 > 3,592). Це означає, що ці змінні x_i залежать від всіх інших незалежних змінних в масиві. Тому в нашому прикладі кожна з незалежних змінних мультиколінеарна з усіма іншими.

Тестування за частковими коефіцієнтами детермінації ($R^2_{x_1} = 0,9063$; $R^2_{x_2} = 0,8935$; $R^2_{x_3} = 0,8065$) показало, що оскільки коефіцієнти детермінації дуже близькі до одиниці, то кожна з незалежних змінних є мультиколінеарною з іншими. Але найбільш лінійно залежною з іншими змінними є змінна x_1 , оскільки їй відповідає найбільше значення коефіцієнта детермінації $R^2_{x_1} = 0,9063$.

3.3) перевірка кожної пари незалежних змінних за t-критерієм. Враховуючи отримані значення частинних коефіцієнтів кореляції, можна зробити висновок, що зв'язок між фондовіддачею та продуктивністю праці є сильним, якщо не враховувати вплив питомих інвестицій $r_{12} = 0,73674$; зв'язок між фондовіддачею та питомими інвестиціями є помірним, якщо не брати до уваги вплив продуктивності праці $r_{13} = 0,41129$; зв'язок між продуктивністю праці та питомими інвестиціями є слабким, якщо не враховувати фондовіддачу $r_{23} = 0,23683$.

Перевірка статистичної значущості частинних коефіцієнтів кореляції за t-критерієм показала наступне. Оскільки $t_{12} > t_{\text{критичне}}$ (4,492 > 2,110), тому фондовіддача та продуктивність праці є відповідно мультиколінеарними між собою; $t_{13} < t_{\text{критичне}}$ (1,860 > 2,110), тому відповідно фондовіддача та питомі інвестиції не є мультиколінеарними між собою; $t_{23} < t_{\text{критичне}}$ (1,005 > 2,110), тому продуктивність праці та питомі інвестиції не є мультиколінеарними між собою.

4) Дослідження показали, що в заданому масиві даних мультиколінеарність існує між фондовіддачею та продуктивністю праці. Отже, для того, щоб можна було застосувати метод 1 МНК для оцінювання параметрів моделі за цією інформацією,

необхідно звільнитися від мультиколінеарності шляхом вилучення з моделі (1) фондовіддачі або продуктивності праці.

5.1) В результаті вилучення з моделі (1) змінної x_1 отримаємо модель (2):

$$\hat{y} = 24,74 + 0,769 x_2 + 0,744 x_3 \quad (2)$$

А) За F-критерієм Фішера модель (2) є адекватною, оскільки $F_{\text{фактичне}} > F_{\text{критичне}}$, про що свідчить значення $\text{Significance } F = 0,00000000094$, що значно менше за 0,05 (див. рис. 7).

Б) Отриманий коефіцієнт множинної кореляції, $R = 0,9556$, вказує на те, що між прибутком підприємства та продуктивністю праці і питомими інвестиціями, існує лінійний, прямий та сильний кореляційний зв'язок оскільки, його значення наближується до одиниці.

В) Коефіцієнт детермінації $R^2 = 0,9133$, показує, що зміна значення прибутку підприємства на 91,33% пояснюється зміною значень продуктивності праці та питомими інвестиціями, а на 8,67% - іншими випадковими факторами, що свідчить про високий рівень адекватності моделі в цілому.

Г) Отримані оцінки параметрів в моделі (2) є статистично значущими, оскільки всі значення $P\text{-value}$ також менші за 0,05 та значення t-статистики (t Stat на рис. 7), що більші за $t_{\text{критичне}}$ (2,11) (див. рис. 7). Отже, можна зробити висновок, що за моделлю (2) можна робити економічний аналіз параметрів та прогнозування.

Д) Оскільки нами було доведено статистичну значущість параметрів моделі, надамо їх економічну інтерпретацію.

На основі обчислених коефіцієнтів регресії a_2 , a_3 оцінюємо граничний вплив кожного фактору на прибуток. На підставі обчислених значень можна зробити висновки про те, що у випадку збільшення продуктивності праці на 1 тис.грн./людино-год. прибуток в середньому збільшиться на 0,769 тис.грн. за умови, що інший фактор – питомі інвестиції залишиться незмінним. При збільшенні питомих інвестицій на 1% прибуток в середньому збільшиться на 0,744 тис.грн. за умови, що інший фактор – продуктивність праці залишиться без змін.

5.2) В результаті вилучення з моделі (1) змінної x_2 отримаємо модель (3):

$$\hat{y} = 19,97 + 0,783 x_1 + 0,644 x_3 \quad (3)$$

А) За F-критерієм Фішера модель (2) є адекватною, оскільки $F_{\text{фактичне}} > F_{\text{критичне}}$, про що свідчить значення $\text{Significance } F = 0,000000000637$, що значно менше за 0,05 (див. рис. 9).

Б) Отриманий коефіцієнт множинної кореляції, $R = 0,958$, вказує на те, що між прибутком підприємства та фондовіддачею і питомими інвестиціями, існує лінійний, прямий та сильний кореляційний зв'язок оскільки, його значення наближується до одиниці.

В) Коефіцієнт детермінації $R^2 = 0,917$, показує, що зміна значення прибутку підприємства на 91,7% пояснюється зміною значень фондовіддачі та питомими інвестиціями, а на 8,3% - іншими випадковими факторами, що свідчить про високий рівень адекватності моделі в цілому.

Г) Отримані оцінки параметрів в моделі (3) є статистично значущими, оскільки всі значення $P\text{-value}$ також менші за 0,05 та значення t-статистики (t Stat на рис. 9),

що більші за $t_{\text{критичне}} (2,11)$ (див. рис. 9). Отже, можна зробити висновок, що за моделлю (3) також можна робити економічний аналіз параметрів та прогнозування.

Д) Оскільки нами було доведено статистичну значущість параметрів моделі, надамо їх економічну інтерпретацію.

На основі обчислених коефіцієнтів регресії a_1, a_3 оцінюємо граничний вплив кожного фактору на прибуток. На підставі обчислених значень можна зробити висновки про те, що у випадку збільшення фондоддачі на 1 тис.грн./ тис.грн. прибуток в середньому збільшиться на 0,783 тис.грн. за умови, що інший фактор – питоми інвестиції залишиться незмінним. При збільшенні питомих інвестицій на 1% прибуток в середньому збільшиться на 0,644 тис.грн. за умови, що інший фактор – фондоддача залишиться без змін.