Східноєвропейський національний університет
імені Лесі Українки
Інститут економіки та менеджменту
Кафедра обліку і аудиту

Світлана Бегун

ЕКОНОМЕТРИКА

Методичні вказівки
для самостійної роботи

Луцьк - 2014

УДК 330.43 (075.8)

ББК 65.050я7
Б-87

Рекомендовано до друку науково-методичною радою Східноєвропейського національного університету імені Лесі Українки (протокол № від 2014).

Рецензенти: Грудзевич І.Т., к.е.н., доцент, зав. кафедри фінансів і кредиту СНУ ім..Лесі Українки;

Зорій Н.М., к.е.н., доцент кафедри обліку у бюджетній та соціальній сфері Тернопольського національного економічного університету

Бегун С.І.

Б-87 Економетрика : Методичні вказівки для самостійної роботи/ Світлана Іванівна Бегун. – Луцьк : Вид-во ПП Іванюк В.П., 2014. – 60 с.

В методичних вказівках розглянуто основні питання курсу економетрики, винесені на самостійну роботу студентів: найпростіша економетрична модель, множинна модель, мультиколінеарність та методи її визначення, гетероскедастичність та методи її усунення тощо. Наведено приклади розрахункових завдань.

Рекомендовано студентам, що навчаються на 3 курсі за напрямом 6.030509 «Облік і аудит», 6.030508 «Фінанси і кредит», 6.030504 «Економіка підприємства».

УДК
ББК

Зміст

Вступ

Підготовка кваліфікованих економістів вимагає отримання знань з основних фундаментальних курсів: мікроекономіки, макроекономіки та економетрики. Вивчення останьої набуває дуже важливого значення в зв'язку з переходом країни до світових стандартів освіти.

Предметом курсу «Економетрія» є методи побудови та застосування економетричних моделей, що характеризують взаємозв'язок між економічними показниками. Уже виходячи із визначення очевидно, що питома вага теоретичних викладок методів та їх практичної реалізації для конкретних економічних задач має бути приблизно однаковою. Не можна вивчити методи оцінки параметрів економетричних моделей, не застосувавши той чи інший алгоритм методу для конкретної реалізації на прикладах. Студент може знати ідею методу та вивчити його алгоритм, але щоб вміти методом користуватись, необхідно мати досвід розв'язання конкретних економічних задач на його базі.

Мета даного курсу:

- навчити студентів економічних спеціальностей методам побудови економетричних моделей, які кількісно описують взаємозв'язки між економічними показниками;
- дати їм навички використання цих моделей в конкретних економічних дослідженнях;
- показати, як знання, набуті при вивченні економетрики, широко застосовуються у різних економічних курсах: мікроекономіки, маркетингу, менеджменту, макроекономіки, аналізу господарської діяльності тощо;
- надати додаткові можливості для використання обчислювальної техніки, сприяти розвитку аналітичних навичок.

Дані методичні вказівки включають в себе основні теоретичні питання та практичні завдання курсу "Економетрика" для самостійної роботи.

Тема 1. Вступ до економетрії

В перекладі з латинської "економетрія" означає вимірювання в економіці, але це дуже розпливчасте визначення економетрії. Слід пам'ятати, що в економіці все залежить від всього, тобто всі явища та процеси взаємозв'язані між собою. Економічна теорія вивчає причини та наслідки цих зв'язків, а економетрія займається зв'язками взагалі. Таким чином, економетрія – це наука, яка базується на вивченні кількісних закономірностей та взаємозв'язків економічних об'єктів і процесів за допомогою математико-статистичних методів і моделей.

Знання економетрії необхідно для перевірки емпіричних залежностей, для отримання нових таких залежностей, для побудови надійних прогнозів, що сприяє успіху в бізнесі, банківській справі тощо. Саме тому економетрія займає одне з ключових місць серед фундаментальних курсів підготовки бакалаврів з економіки.

Предметом економетрії є методи оцінювання параметрів економіко-математичних моделей, які описують кількісні взаємозв'язки між економічними показниками, а також основні напрямки застосування цих моделей в економічних дослідженнях.

Економіко-математичні моделі, які використовуються в економетрії називаються економетричними моделями, а математичні методи їх побудови – відповідно економетричними методами.

Виходячи з предмету економетрії основними завданнями цієї дисципліни є:

1. вивчення і розробка математичних методів побудови економетричних моделей;
2. використання економетричних моделей в економічних дослідженнях.

Об'єктами економетрії є різноманітні економічні явища і процеси, як на мікро- так і на макрорівнях, такі як: попит і пропозиція, виробництво, споживання ,зайнятість і т.п.

За допомогою економетричних методів вирішуються наступні важливі питання :

1. підтвердження або відхилення економічних гіпотез і законів ;
2. емпіричний вивід економічних законів ;
3. прогнозування різноманітних економічних показників ;
4. економіко-математичний аналіз явища або процесу, що досліджується .

Економетрія тісно пов'язана з економічною теорією, економічною математикою, економічною статистикою. Так, економічна теорія (на мікро – та макрорівні) пропонує якісні гіпотези, але не наводить жодного кількісного виміру показників, що підтвердило б цю гіпотезу. Завдання економетрії полягає в обчисленні відповідних кількісних оцінок. Математична економіка тільки відображає економічну теорію в економічній формі, тоді як економетрія займається емпіричним підтвердженням теоретичних тверджень. Зібрані статистичні дані є основою для економетричного дослідження. В свою чергу економетрія широко використовує поняття та методи лінійної алгебри, математичної статистики, теорії ймовірностей, математичного аналізу та ін.

Крім зазначених дисциплін економетрія також широко застосовує методи вищої математики і сучасні комп'ютерні технології. Оскільки економетричні дослідження пов'язані, як правило, з великим обсягом обчислень, вони потребують відповідної комп'ютерної підтримки. На даний час розроблено і використовується достатня кількість як спеціалізованих програмних продуктів, таких як STATISTICA, SPSS, STATA, StatGraphics, Econometric Views і т.і., так і універсальні програмні продукти, які мають достатньо потужні можливості для проведення економетричного аналізу - MS Excel, MathCAD, Mathematica та інші.

Термін “Економетрія” вперше запропонував львівський вчений П. Чомпа, який опублікував у Львові у 1910 році книгу “Нариси економетрії і природної теорії бухгалтерії, яка ґрунтується на політичній економії”.

Остаточно економетрія, як галузь науки, під такою назвою стала відома у 1930 році, коли було засновано “Економетричне товариство”, яке визнало себе як “Міжнародне товариство для розвитку економічної теорії і її зв’язку зі статистикою та математикою”. Засновниками економетрії вважаються Р. Фріш (Норвегія), Е. Шумпетер (Швейцарія), Я. Тінберген (Нідерланди).

Після застосування “Економетричного товариства”, починаючи з 30-х років 20-го століття відомими економістами Я. Тінбергеном (Нідерланди), Л. Клейном (США) та Р. Стоуном (США), почали розроблятися макроекономічні моделі, які описували статистичні зв’язки виробництва, кінцевого індивідуального і державного попиту, цін, податків, зовнішньої торгівлі, пропозиції робочої сили і т.і. Такі моделі складалися вже з багатьох рівнянь, що дали поштовх для розробки нових методів оцінювання параметрів економетричних моделей. Цей напрямок було продовжено і після 2-ї світової війни.

Особливі досягнення пов’язані з розвитком економетрії в останні 30 років. Про це свідчить той факт, що серед лауреатів Нобелівської премії у галузі економіки достатньо велика кількість економістів, які відзначились значним внеском в економетрію. Серед цих лауреатів : проф. Рагнар Фріш – 1969 (Норвегія) ; проф. Ян Тінберген – 1969 (Нідерланди) ; проф. Теллінг До Купманс – 1975 (США) ; проф. Лоуренс Р. Клейн – 1980 (США) ; проф. Трігве Хаавелмо –1989 (Норвегія) .

У 2000 році Дж. Хекман і Д. Макфеден отримали Нобелівську премію за розробку мікро економетрії та методів статистичного аналізу.

До типових проблем сьогоденяшньої економетрії відносять розробку і дослідження властивостей:

- виробничих функцій;
- функцій попиту різних груп споживачів та цільових функцій переваги споживачів;
- статичних і динамічних міжгалузевих моделей виробництва, розподілу та споживання продукції;
- моделей загальної рівноваги (наприклад між попитом на робочу силу та її пропозицією і т.і.).

Тема 2. Математичне моделювання як метод наукового пізнання економічних явищ і процесів

Сучасні методи дослідження і управління економічними системами та процесами базуються на широкому використанні математичних методів, математичних моделей і комп'ютерних технологій. В сучасній економіці сформувався цілий напрямок теоретико-практичних досліджень – економіко-математичне моделювання, який є виразом процесу математизації науково-економічних знань.

Особливістю економетрії як науки є те, що вона вивчає зв'язки між явищами та процесами за допомогою побудови моделей. При розв'язанні багатьох практичних задач використовують метод математичного моделювання. Він дає змогу використовувати аналоги відомих процесів і приймати правильні рішення. Метод побудови моделей широко використовується в різних галузях науки.

Модель – досліджуваний об'єкт, представлений в найбільш загальному вигляді.

Основна вимога, яку ставлять до моделі, - відображення суті об'єкту, що досліджується. Водночас вона має бути позбавлена несуттєвих деталей, що дозволяє виконувати експеріменти та знайти більш ефективне рішення розв'язання певної проблеми.

Окремо розглядаються економіко-математичні моделі, на яких досліджуються економічні закономірності, подані в абстрактному вигляді. Таким чином.

Економетрична модель – це система математичних співвідношень, яка відображає певний економічний об'єкт, явище або процес. Типовими економіко-математичними моделями є виробничі функції, функції попиту споживачів, статистичні та

В залежності від конкретного виду економіко-математичні моделі можуть мати вигляд рівнянь, нерівностей, тотожностей, систем рівнянь і т.д.

Сам процес економіко-математичного моделювання включає наступні етапи:

1. постановка задачі ;
2. побудова математичної моделі;
3. реалізація математичної моделі (знаходження розв'язку моделі) ;
4. перевірка адекватності моделі ;
5. використання економіко-математичної моделі для економічного аналізу, прогнозу і оптимізації.

Що ж дає економіко-математичне моделювання, як метод наукового пізнання економічних явищ і процесів ? Він дає можливість :

1. виділити і формально описати найбільш важливі, суттєві зв'язки економічних змінних(показників) і об'єктів ;
2. шляхом формальних математичних перетворень виконати більш детальний аналіз модельованого явища;
3. розробляти прогнози функціонування економічних систем при зміні її зовнішніх параметрів.

Усі відомі економіко-математичні моделі класифікують за наступними ознаками.

1. В залежності від способу опису зв'язків між параметрами і змінними моделі розрізняють наступні економіко-математичні моделі :

- структурні , які відтворюють внутрішню організацію об'єкту ;

- функціональні, які описують поведінку об'єкта без знання його внутрішньої структури

2.За рівнем моделювання:

- мікромоделі ;
- макромоделі .

3.За способом побудови економіко-математичні моделі поділяються на :

- теоретичні ;
- прикладні .

Теоретичні моделі дозволяють вивчати загальні властивості економічних

систем за допомогою дедукції на основі формальних вихідних посилок і носять достатньо абстрактний і узагальнений характер . Параметри таких моделей мають тільки загальне позначення і не мають конкретних числових значень.

Прикладні моделі базуються на фактичному, статистичному матеріалі і дають можливість кількісно оцінити характеристики конкретного економічного процесу і сформулювати рекомендації для прийняття практичних рішень.

4.В залежності від мети економіко-математичні моделі поділяються на:

- дескриптивні (описові) ;
- оптимізаційні .

5.В залежності від врахування фактору часу економіко-математичні моделі поділяються на :

- статичні, в яких не враховується фактор часу;
- динамічні, в яких враховується фактор часу .

6.В залежності від врахування випадкових величин економіко-математичні моделі поділяються на :

- детерміновані ;
- стохастичні .

Детерміновані моделі припускають тільки жорсткі функціональні зв'язки між змінними (параметрами) моделі.

Стохастичні моделі припускають наявність випадкових впливів на показники, що досліджуються і використовують інструментарій теорії ймовірності і математичної статистики .У таких моделях зв'язки між змінними – кореляційно-регресійні.

Будь-які економічні показники, зазвичай, перебувають під впливом багатьох випадкових факторів, а тому з математичної точки зору вони повинні розглядатися як випадкові величини. Внаслідок цього залежності між економічними показниками не є однозначними, не є функціональними. Це означає, що кожному фіксованому значенню однієї економічної змінної (або фіксованому набору змінних) відповідає не одне єдине, а множина значень іншої змінної, тобто деякий ймовірностний розподіл. Тому в економіці спостерігаються і розглядаються так звані статистичні (або кореляційні) залежності.

Статистичною називається залежність, коли зі зміною однієї випадкової величини змінюється закон розподілу ймовірностей іншої.

Кореляційною називається статистична залежність, коли зі зміною однієї випадкової величини змінюється математичне сподівання (середнє значення) іншої.

Економетричні моделі використовуються у такій послідовності: спочатку ставиться економічна задача, яка описує реальну чи проблемну ситуацію. На основі аналізу проблеми створюють математичну модель. За допомогою математичних методів аналізують модель, в наслідок чого дістають розвиток проблеми, який після всебічного аналізу рекомендується до впровадження у практику.

Найпростішою економічною моделлю є лінійна, яка в загальному має такий вигляд:

$$A_{y_z} = B_{x_t} + E_t ,$$

де A – матриця невідомих параметрів (розміром m*),

y_z - вектор ендогенних змінних (m*1),

B – матрица невідомих параметрів (m*k),

x_t - вектор екзогенних змінних (k*1),

E_t - випадковий вектор відхилень,

t – номер спостереження (або момент часу).

Прикладом лінійної моделі може бути залежність попиту на ринку на певний товар від його ціни. В даному випадку, ціна – екзогенна змінна, тобто незалежна, вхідна змінна, яка характеризує причину (або факторна), а попит на товар – це ендогенна змінна,тобто, залежна, вихідна змінна, що характеризує наслідок (або результативна).

Лінійні моделі відносяться до класу регресійних моделей з одним рівнянням. До нього також належать і нелінійні моделі, які представленні нелінійними функціями (гіпербола, парабола, степенева тощо).

Крім класу регресійних моделей виділяють ще класи моделей динамічних рядів та системи одночасних рівнянь. Серед моделей динамічних рядів найчастіше використовуються моделі тренду та моделі сезонності. Загальною рисою цих моделей є те, що вони пояснюють поведінку динамічного ряду, виходячи з його попередніх значень. Останній клас моделей представлений системами рівнянь. Прикладом може бути модель попиту та пропозиції. Особливістю цих моделей є більш складний математичний апарат.

Економетричні моделі представляють собою окремий клас економіко-математичних моделей і характеризуються наступними особливостями:

1) економетричні моделі є моделі функціональні ;
2) економетричні моделі є моделі прикладні (емпіричні) ;
3) економетричні моделі є моделі дескриптивні ;
4) економетричні моделі є моделі стохастичні.

У загальному вигляді економетрична модель у вигляді однієї функції (рівняння) має наступний вигляд :

$$y = f(x_1, x_2, ..., x_m, \varepsilon),$$

де y - залежна змінна; $x_i(i=\overline{1,m})$- незалежні змінні; ε - випадкова (стохастична) складова моделі. Прикладом такої моделі може бути відома виробнича функція Кобба-Дугласа:

$$Y = a_0 k^{\alpha} L^{\beta} e^{\varepsilon},$$

де Y - випуск продукції; К – основний капітал; L – затрати праці (людський капітал); a_0, α і β - параметри моделі.

Якщо економетрична модель представляє собою систему функцій (рівнянь) вона у загальному вигляді має наступний вигляд:

$$y_s = f_s(x_{s1}, x_{s2}, \ldots, x_{sm}, \varepsilon_s), (s=\overline{1,k}),$$

де к – кількість рівнянь. Прикладом такої моделі може бути відома модель формування доходу Дж. М. Кейнса:

$$\begin{cases} C_t = \beta_0 + \beta_1 y_t + \varepsilon_t, \\ y_t = C_t + I_t, \end{cases}$$

де C_t - сукупне споживання, y_t - національний дохід, I_t - інвестиції, β_0, β_1 - параметри моделі.

Незалежні змінні економетричних моделей $x_1, x_2, ..., x_m$,як і регресійних, називають пояснюючими змінними (або факторами, інколи регресорами). Залежні змінні $y(y_s)$ називають пояснюваними змінними (або регресандами). Крім цього усі змінні економетричних моделей, як і будь-якої економіко-математичної моделі, поділяють на екзогенні і ендогенні.

Екзогенними (зовнішніми) називаються змінні, значення яких є наперед визначеними перед використанням моделі, а ендогенними (внутрішніми) – такі, значення яких визначаються тільки із самої моделі.

Так, для моделі змінні K і L є екзогенними змінними, а Y - ендогенною. Для моделі тільки змінна I_t є екзогенною, а змінні C_t і y_t - ендогенними, при чому вони виступають одночасно як залежні так і незалежні змінні.

Випадкову складову економетричної моделі за аналогією з регресійною моделлю прийнято називати збуренням (похибкою, відхиленням) моделі, а для вибіркової моделі при означені оцінки цієї величини в основному використовується термін залишки.

Введення до економетричної моделі стохастичної складової має наступні підстави:

1) до будь-якої економетричної моделі включаються не всі фактори, які можуть впливати на залежну змінну, а тільки основні ;

2) на залежну змінну при моделюванні таких складних об'єктів як економічні системи можуть впливати і численні випадкові фактори, які взагалі неможливо передбачити ;

3) частина факторів не піддається квантифікації (тобто кількісному вимірюванню), а для тих, що вимірюються можлива похибка вимірювання даних.

Стохастична складова моделі якраз і акумулює в собі всі відхилення фактичних спостережень залежної змінної від обчисленої згідно рівняння регресії за рахунок наведених вище обставин.

Будь-яке економічне дослідження проходить такі етапи аналізу:

1. Формування теоретичної проблеми або гіпотези (теоретичний опис вивчаємих явищ, визначення цілей дослідження, відбір економічних змінних тощо).
2. Вібір форми економіко-математичної моделі.
3. Розрахунок та оцінка параметрів моделі.
4. Інтерпретація отриманих результатів.
5. Застосування моделі (для прогнозування, для контролю тощо).

Економетрія вивчає кількісні взаємозв'язки і залежності між економічними показниками, явищами і процесами. При вивчені і описі таких взаємозв'язків і залежностей економічні показники повинні розглядатися як випадкові величини, що пов'язано з впливом на останні різноманітних випадкових факторів. Внаслідок цього залежності між економічними показниками не є однозначними, не є функціональними. Це означає, що кожному фіксованому значенню однієї економічної змінної (або фіксованому набору змінних) відповідає не одне єдине, а множина значень іншої змінної, тобто деякий імовірнісний розподіл. Тому в економіці спостерігаються і розглядаються так звані статистичні (або кореляційні) залежності.

Кореляційна залежність між економічними показниками (змінними) (або залежність у середньому) описується за допомогою функції (рівняння) регресії. Функція регресії описує залежність між незалежною змінною (або незалежними змінними) і умовним математичним сподіванням (середнім) залежної змінної. У якості змінних такої функції виступають економічні показники. Регресією прийнято називати функціональну залежність між незалежною змінною (або незалежними змінними) і умовним математичним сподіванням (середнім значенням) залежної змінної.

В залежності від числа змінних розрізняють парну і багатофакторну (множинну) регресію, а в залежності від виду функції регресії – лінійну та нелінійну регресію.

Функція регресії має графічну форму представлення. Так для парної регресії графічне зображення функції регресії на площині x0y представляє собою так звану криву (лінію) регресії. Для множинної регресії крива регресії перетворюється на поверхню (або гіперповерхню) регресії.

Фактична, статистична залежність між економічними показниками подається як модель регресії (або регресійна модель). Модель регресії – це математична модель, яка описує статистичну залежність деякого економічного показника (змінної) від інших. Іншими словами модель регресії – це математична модель, яка описує кореляційно-регресійні зв’язки між економічними показниками, один з яких розглядається як залежна змінна, а усі інші – як незалежні.

В залежності від числа змінних розрізняють модель парної і багатофакторної (множинної) регресії, а в залежності від виду функції регресії – лінійну та нелінійну моделі регресії.

В залежності від статистичної бази розрізняють теоретичну і вибіркову (емпіричну) моделі регресії. Теоретична модель відповідає генеральній сукупності спостережень за змінними моделі, а вибіркова – побудована на основі окремої статистичної вибірки з генеральної сукупності спостережень.

Теоретична модель регресії є ідеалізованою конструкцією, оскільки у практиці економетричного моделювання, як правило, не доводиться мати справу з генеральною сукупністю спостережень, а тільки з деякою окремою статистичною вибіркою з неї. Тому реально, ніколи не можливо побудувати «дійсну» теоретичну регресію, а тільки вибіркову (емпіричну), яка є тільки наближенням (оцінкою) «дійсної» теоретичної регресії.

Питаннями побудови якісних і статистично надійних регресійних моделей займається кореляційно-регресійний аналіз, який представляє собою сукупність методів, за допомогою яких на основі вибіркових статистичних даних досліджуються та узагальнюються взаємозв'язки кореляційно пов'язаних змінних.

Кореляційно-регресійний аналіз відіграє дуже важливе значення в економетричному аналізі, оскільки саме його методи і підходи покладено в основу більшості економетричних методів. Основними задачами такого аналізу є :

- визначення наявності і сили кореляційного взаємозв'язку між економічними показниками ;
- вибір виду (аналітичної форми) функції регресії;
- оцінювання (визначення) параметрів вибраного рівняння регресії на основі статистичної вибірки;
- аналіз якості регресійної моделі і перевірка адекватності моделі статистичним даним.

Тема 3. Загальна лінійна економетрична модель

При вивченні економетричних моделей слід пам'ятати, що всі вони побудовані на вибіркових даних, тобто не відомо точний імовірнісний розподіл або щільність розподілу змінних генеральної сукупності. Отже, при побудові будь-якої моделі отримуємо тільки оцінки параметрів чи коефіцієнтів.

До першочергових завдань економетричного аналізу відноситься специфікація моделі (вираз її у математичній формі), яка б відповідно до нашої уяви адекватно відтворювала зв'язки між явищами. Найпростішою є лінійна модель з двома змінними – екзогенною - х та екзогенною – у.

Серед багаточисельних зв'язків між економічними показниками завжди можна виділити такий показник, вплив якого на результативну ознаку є основним, найбільш важливим. Щоб виміряти цей зв'язок кількісно, необхідно побудувати економетричну модель з двома змінними (просту модель). Загальний вигляд такої моделі:

$$Y = f(X, u),$$

де Y — залежна змінна (результативна ознака); X — незалежна змінна (фактор); u — стохастична складова.

Аналітична форма цієї моделі може бути різною залежно від економічної сутності зв'язків. Найбільш поширені форми залежностей:

$$Y = a_0 + a_1 X \; ;$$

$$Y = a_0 e^{a_1 X}$$

$$Y = a_0 + \frac{a_1}{X} \; ;$$

$$Y = a_0 X^{a_1} \; ,$$

де a_0, a_1 — невідомі параметри моделі.

У загальному вигляді вона записується так:

$$Y = a_0 + a_1 X$$

де a_0 і a_1 - невідомі оцінки параметрів рівняння,

a_0 - визначає перетин прямої з віссю ординат,

a_1 - тангенс кута нахилу прямої регресії до осі абсцис.

Але такий запис моделі не забезпечує її адекватності. Вирішення проблеми адекватності можливо за допомогою введення стохастичного (випадкового) члену Е в рівняння прямої:

$y = a_0 + a_1 x + E$

Ввод випадкової змінної Е спричинений тим, що в дійсності на ендогенну змінну y впливає множина ендогенних змінних x –

$y = f(x_1, x_2, x_3, ... x_n)$

Саме тому випадковий член Е відображає сумарний ефект від впливу неврахованих в моделі факторів.

- теоретична (“канонічна”) модель парної лінійної регресії

$$y = \beta_0 + \beta_1 x + \varepsilon,$$

- вибіркова (емпірична) модель парної лінійної регресії

$$y = b_0 + b_1 x + e,$$

- вибіркова функція парної лінійної регресії

$$\hat{y} = b_0 + b_1 x.$$

Рівняння представляє собою параметричне рівняння прямої, тому на площині x0y вибірковій функції парної лінійної регресії відповідає вибіркова (емпірична) пряма регресії.

Для оцінювання параметрів вибіркової моделі використовуються різні методи, такі як: метод моментів, метод найменших квадратів, метод максимальної правдоподібності та інші. До найбільш простих і розповсюджених методів відноситься широко відомий метод найменших квадратів, який становить основу регресійного аналізу. В принципі цей метод може застосовуватися для оцінювання параметрів будь-яких лінійних економетричних моделей. Але результат застосування цього методу залежить від того, чи задовольняє економетрична модель усім припущенням класичного лінійного регресійного аналізу, чи ні. Якщо модель задовольняє усім припущенням класичного лінійного регресійного аналізу, то вона є по суті багатофакторною (або однофакторною) класичною регресійною моделлю. У цьому випадку для оцінювання її параметрів може коректно застосовуватися однокрокова процедура методу найменших квадратів (1МНК) і отримані при цьому оцінки параметрів моделі будуть найкращими серед усіх можливих. Якщо ж для моделі порушується хоча б одне з положень класичної лінійної регресії така модель називається узагальненою моделлю і для оцінювання її параметрів застосовуються інші більш ефективні методи.

В основі методу МНК лежить принцип мінімізації суми квадратів залишків моделі. Реалізація цього принципу дає можливість отримати систему нормальних рівнянь:

$$\begin{cases} na_0 + \sum_i x_i a_1 = \sum_i y_i \\ \sum_i x_i a_0 + \sum_i x_i^2 a_1 = \sum_i x_i y_i \end{cases}$$

В даній системі m — кількість спостережень, $\sum_i x_i$, $\sum_i y_i$, $\sum_i x_i^2$, $\sum_i x_i y_i$ — величини, які можна розрахувати на основі вихідних спостережень над змінними і . Звідки

$$a_0 = \bar{y} - a_1\bar{x}, \quad a_1 = \frac{\overline{xy} - \bar{x} * \bar{y}}{\overline{x^2} - \bar{x}^2} = \frac{Cov(x,y)}{Var(x)},$$

де Cov(x,y) – це змішана дисперсія, а

Var(x) – дисперсія факторної ознаки.

Розв'язавши систему нормальних рівнянь, одержимо оцінки невідомих параметрів моделі $\hat{a}_0$ і $\hat{a}_1$:

$$\hat{Y} = \hat{a}_0 + \hat{a}_1 X \quad .$$

Звідки

$$a_0 = \bar{y} - a_1\bar{x}, \quad a_1 = \frac{\overline{xy} - \bar{x} * \bar{y}}{\overline{x^2} - \bar{x}^2} = \frac{Cov(x,y)}{Var(x)},$$

де Cov(x,y) – це змішана дисперсія, а

Var(x) – дисперсія факторної ознаки.

Достовірність побудованої економетричної моделі можна перевірити, користуючись елементами дисперсійного аналізу. Перш за все слід розрахувати залишки моделі $u_i = y_i - \hat{y}_i$

та знайти їх дисперсію:

$$\sigma_u^2 = \frac{\sum_i u_i^2}{n-m},$$

де m— кількість змінних моделі (m=2).

$$S_{\hat{a}_j} = \sigma_u \sqrt{c_{jj}}$$

необхідно визначити стандартну помилку кожного параметра моделі. В цій формулі характеризує відповідний діагональний елемент матриці помилок (матриці, оберненої до матриці системи нормальних рівнянь).

До основних припущень класичного лінійного регресійного аналізу належать наступні умови Гауса-Маркова .

1. Математичне сподівання стохастичної складової моделі дорівнює нулю для всіх спостережень:

$$M(\varepsilon_i) = 0,\ i = 1,2,\ldots,n$$

<u>Перше припущення</u> стверджує що, невраховані у моделі і тому віднесені до стохастичної складової ε фактори не впливають систематично на математичне сподівання залежної змінної y.У протилежному ж випадку існує систематичний вплив на залежну змінну і до моделі не введено всі основні пояснюючі змінні.

2. Гомоскедастичність.

Це означає, що дисперсія стохастичної складової у всіх спостереженнях повинна бути однаковою

$$\mathrm{var}(\varepsilon_i) = \sigma_\varepsilon^2 = \mathrm{const},\ i = 1,2,\ldots,n$$

3. Відсутність автокореляції залишків.

Це означає, що значення $\varepsilon_i, \left(i = \overline{1,n}\right)$ вектора стохастичної складової моделі повинні бути незалежні між собою

$$\mathrm{cov}(\varepsilon_i, \varepsilon_j) = 0 \text{ для } i \neq j.$$

Друге і третє припущення передбачають відсутність автокореляції залишків (залежність між залишками у вибірці) і гомоскедастичність моделі (сталість дисперсії залишків). Відсутність автокореляцій залишків дає можливість вивчати тільки систематичний вплив незалежних змінних на залежну. Сталість дисперсії залишків (гомоскедастичність) дає підстави вважати всі значення залежної змінної y_i, які відносяться до різних спостережень однаково важливими при оцінюванні параметрів моделі.

4.Незалежні (пояснюючи) змінні не пов'язані із стохастичною складовою моделі :

$$\mathrm{cov}(\varepsilon_i, x_{ji}) = 0 \text{ для } i = 1,2,\dots,n;\ j = 1,2,\dots,m$$

5.Матриця спостережень X *не є стохастичною.*

П'яте припущення означає, що пояснюючі змінні моделі утворюють лінійно незалежну систему векторів, внаслідок чого матриця спостережень за незалежними змінними моделі має повний ранг і $\det(X'X) \neq 0$.

6.Відсутність мультиколіарності .

Незалежні змінні моделі не повинні бути мультиколінеарними, тобто між ними не повинно існувати лінійного функціонального або тісного кореляційного зв'язку. Слід зазначити, що дане припущення має місце тільки для багатофакторної регресії.

7.Випадкова величина $\boldsymbol{\varepsilon_i}, \boldsymbol{i = 1,2,\dots,n}$ *має нормальний закон розподілу з математичним сподіванням О і сталою дисперсією:*

$$\varepsilon_i \sim N(0, \sigma_\varepsilon^2) .$$

Після специфікації моделі виникає питання: які оцінки більш точні, коректні? Найбільш поширений метод визначення оцінок параметрів рівняння є метод найменших квадратів(МНК). Суть його полягає в мінімізації квадратів відхилень емпіричних значень результативної ознаки у від теоретичних (знайдених за рівнянням регресії) значень $\tilde{y}$:

$$\sum E^2 = \sum (y - \tilde{y})^2 \to \min$$

За МНК рівняння оцінюванної лінії регресії має вигляд $\tilde{y} = a_0 + a_1 x$,

де a_0 - вільний член рівняння, найчастіш не має економічного змісту;

a_1 - коефіцієнт регресії, який показує наскільки одиниць власного виміру зміниться результативна ознака при збільшенні факторної ознаки на 1 власного виміру.

Для позначення вибіркових характеристик в економетрії застосовуються свої позначення , які відрізняютьсЯ інколи від прийнятих в статистиці.

Для оцінки тісноти зв’язку між ознаками використовуються такі коефіцієнти:

Лінійний коефіцієнт кореляції Пірсона:

$$r = \frac{Cov(x,y)}{\sqrt{Var(x)} * \sqrt{Var(y)}}, \ -1 \le r \le 1,$$

характеризує напрям і силу зв’язку (за шкалою Чеддока).

$$R^2 = \frac{Var(\tilde{y})}{Var(y)}, \ 0 \le R^2 \le 1,$$

де $Var(\tilde{y}) = \dfrac{\sum(\tilde{y}_i - \bar{y})^2}{n}$ - дисперсія теоретичних значень (пояснена дисперсія),

$Var(y) = \dfrac{\sum(y_i - \bar{y})^2}{n}$ - загальна дисперсія.

R^2 характеризує частку варіації результативної ознаки, яка залежить від варіації факторної ознаки при лінійному рівнянні регресії.

Для перевірки значимості параметрів використовують t – критерії Стьюдента, які дозволяють також побудувати довірчі інтервали функції регресії.

$t_{a_0} = a_0 \sqrt{\dfrac{n-2}{Var(E)}}$; $t_{a_1} = a_1 \sqrt{\dfrac{(n-2)Var(x)}{Var(E)}}$, де

$Var(E) = \dfrac{\sum(y - \tilde{y})^2}{n}$ - дисперсія відхилень.

Таке позначення $t_{кр}$ як $t_{\alpha/2}$ пояснюється тим, що t-розподіл Ст'юдента є симетричним розподілом і критична область для нього складається з двох фрагментів, межі яких задаються квантилем $\alpha/2$. У самій же таблиці t-розподілу Ст'юдента критичне значення $t_{кр}$ слід вибирати для значення α при двосторонньому тесті і для значення $\alpha/2$ - для одностороннього. Оскільки розрахункові значення значення можуть бути як додатними так і від'ємними, слід орієнтуватися на двосторонній тест на значимість.

Порівнюючи фактичні значення t_{a_0} і t_{a_1} з критичними, що знаходяться в спеціальних таблицях, визначаємо типовість параметрів. Якщо $t_{факт} > t_{крит}$ - параметри типові. Довірчі інтервали для параметрів має такий вигляд (при заданому рівні довіри):

$$\alpha_i = a_s \pm t_{крит} * \sigma_{a_i}$$

Для перевірки істотності зв'язку між ознаками використовують F – критерій Фішера $F = \frac{R^2}{1-R^2} * \frac{n-m}{m-1}$, якщо $F > F_{крит}$, зв'язок істотний.

Для усунення можливого впливу різних одиниць виміру ознак використовують стандартизовану форму моделі, коли змінні замінюються стандартизованні змінні, а параметри рівняння – на β – коефіцієнти, які характеризують зміну ознак в середньоквадратичних відхиленнях.

Приклад .1. На основі даних про роздрібний товарообіг і доходи населення побудувати економетричну модель роздрібного товарообігу. Дати загальну характеристику достовірності моделі та зробити висновки.

Вихідні дані та елементарні перетворення цих даних для побудови моделі наведені в табл. 1.1.

Розв'язання:

1. Ідентифікуємо змінні:

y — роздрібний товарообіг (залежна змінна);

x — доходи населення (незалежна змінна).

Таблиця 1.1

N п/п	y	x	x^2	xy	$\hat{Y}$	$X-\overline{X}$	$Y-\overline{Y}$	$(X-\overline{X})^2$	$(X-\overline{X})*(Y-\overline{Y})$	$\varepsilon = Y-\hat{Y}$	ε^2	$(Y-\overline{Y})^2$
1	17	18	324	306	16.67	-6.5	-5	42.25	32.5	0.33	0.1089	25
2	18	20	400	360	18.31	-4.5	-4	20.25	18.0	-0.31	0.0961	16
3	19	21	441	399	19.31	-3.5	-3	12.25	10.5	-0.13	0.0169	9
5	21	24	576	504	21.59	-0.5	-1	0.25	0.5	-0.59	0.3481	1
6	23	25	625	575	22.41	0.5	1	0.25	0.5	0.59	0.3481	1
7	24	27	729	648	24.05	2.5	2	6.25	5.0	-0.05	0.0125	4
8	25	28	784	700	24.87	3.5	3	12.25	10.5	0.13	0.0169	9
9	26	29	841	754	25.69	4.5	4	20.25	18.0	0.31	0.0961	16
10	27	31	961	837	27.33	6.5	5	42.25	32.5	-0.33	0.1089	25
Σ	220	245	6165	5523	x	x	x	162.5	133	x	1.145	110

2. Нехай специфікація моделі y=f(x) $+\varepsilon$ визначається лінійною функцією; вона має такий вигляд:

$$y = a_0 + a_1 x + \varepsilon\ ,$$

де a_0, a_1 — параметри моделі;

u — стохастична складова, залишки.

3. Оцінимо параметри моделі $\hat{Y} = \hat{a}_0 + \hat{a}_1 X$ за методом МНК. Для цього запишемо систему нормальних рівнянь:

$$\begin{cases} n\hat{a}_0 + \sum\limits_i x_i \hat{a}_1 = \sum\limits_i y_i & (i = \overline{1,n}); \\ \sum\limits_i x_i \hat{a}_0 + \sum\limits_i x_i^2 \hat{a} = \sum\limits_i x_i y_i & (i = \overline{1,n}); \end{cases}$$

n = 10 — кількість спостережень.

Підставимо в цю систему величини *n*, $\sum\limits_i x_i$, $\sum\limits_i y_i$, $\sum\limits_i x_i^2$ $\sum\limits_i x_i y_i$, які розраховані на основі вихідних даних табл. 1.1; тоді система набуде такого вигляду:

$$\begin{cases} 10a_0 + 245a_1 = 220 \\ 245a_0 + 6165a_1 = 5523. \end{cases}$$

Розв’яжемо цю систему відносно невідомих параметрів a_0 і a_1.

Економетрична модель запишеться так:

$$\hat{Y} = 1{,}91 + 0{,}82x\ .$$

4. Розрахуємо дисперсії залежної змінної та залишків:

$$\sigma_Y^2 = \frac{\sum_i (y_i - \bar{y})^2}{n-1} = \frac{110}{9} = 12{,}2\ ;$$

$$\sigma_\varepsilon^2 = \frac{\sum_i u_i}{n-2} = \frac{1{,}145}{8} = 0{,}14\ .$$

5. Визначимо коефіцієнти детермінації та кореляції:

$$R^2 = \frac{\sigma_Y^2}{\sigma_Y^2} = \frac{12{,}2 - 0{,}14}{12{,}2} \approx 0{,}99\ ;$$

$$R = \sqrt{R^2} = \sqrt{0{,}99} = 0{,}994\ .$$

Оскільки коефіцієнт детермінації $R^2 = 0{,}994$, це свідчить, що варіація обсягу роздрібного товарообігу на 99,4% визначається варіацією доходів населення. Коефіцієнт кореляції $R = 0{,}994$ характеризує тісний зв'язок між цими соціально-економічними показниками. Величини R^2 і R для парної економетричної моделі свідчать про її достовірність, якщо вони наближаються до одиниці.

6. Визначимо стандартні помилки оцінок параметрів моделі, враховуючи дисперсію залишків:

$$S_{\hat{a}_0} = \sqrt{\sigma_u^2 * c_{00}} = \sqrt{0{,}14*3{,}79} = \sqrt{0{,}53} = 0{,}73\ ;$$

$$S_{\hat{a}_1} = \sqrt{\sigma_u^2 * c_{11}} = \sqrt{0{,}14*0{,}006} = \sqrt{0{,}0008} = 0{,}028$$

Порівняємо стандартні помилки оцінок параметрів моделі з величиною цих оцінок. В результаті визначимо, що стандартна помилка оцінки параметра $\hat{a}_1$ становить 3,4% абсолютного значення цієї оцінки(0,82), що свідчить про незміщеність даної оцінки параметра моделі. Стандартна помилка оцінки параметра $\hat{a}_0$ становить 38% абсолютного значення цієї оцінки (1,91), а це означає, що даний параметр може мати зміщення, яке зумовлюється невеликою сукупністю спостережень (n = 10).

7. Висновки. Економетрична модель $\hat{Y} = 1{,}91 + 0{,}82X$ кількісно описує зв'язок роздрібного товарообігу і доходів населення.

Параметр $a_1 = 0{,}82$ характеризує граничну величину витрат на купівлю товарів у роздрібній торгівлі, коли дохід збільшується на одиницю, тобто при збільшенні доходів на одиницю обсяг роздрібного товарообігу зростає на 0,82 одиниці.

Визначимо коефіцієнт еластичності роздрібного товарообігу залежно від доходів населення:

$$E_{Y/X} = \frac{\partial \hat{Y}}{\partial X_1} : \frac{\overline{Y}}{\overline{X}} = 0,82 : \frac{2,0}{24,5} = 0,82 : 0,898 = 0,91 \ .$$

На основі коефіцієнта еластичності можна стверджувати, що при збільшенні доходів населення на один процент роздрібний товарообіг зросте на 0,91%.

Завдання для самостійної роботи

Завдання 3.1. На основі даних по дев'яти металобазах побудувати економетричну модель, яка характеризує залежність між витратами обігу та вантажооборотом. Проаналізувати достовірність моделі та її параметрів. Зробити економічні висновки. Вихідні дані наведені в табл. 1.2 — 1.4.

Таблиця 1.2

N п / п	Витрати обігу	Вантажо-оборот
1	2,7	15,6
2	3,0	15,3
3	2,8	14,9
4	2,9	15,1
5	2,6	16,1
6	2,5	16,7
7	2,8	15,4
8	2,6	17,1
9	2,5	16,8

Таблиця 1.3

N п / п	Витрати обігу	Вантажо-оборот
1	2,6	16,9
2	2,9	16,1
3	2,7	15,0
4	2,5	18,0
5	2,7	17,2
6	2,6	17,1
7	2,7	16,4
8	2,6	16,7
9	2,8	16,9

Таблиця 1.4

N п / п	Витрати обігу	Вантажо-оборот
1	2,9	14,1
2	2,.6	17,2
3	2,8	17,1
4	2,7	17,8
5	2,7	16,2
6	2,9	17,2
7	2,4	16,8
8	2,9	14,8
9	2,3	19,6

Тема 4. Множинна лінійна модель

Складність соціально-економічних явищ приводить до необхідності включення в модель багатьох факторів. Найчастіше використовуються лінійні рівняння множинної регресії для n факторів

$\tilde{y} = a_0 + a_1 x_1 + a_2 x_2 + ... + a_n x_n + E,$

де $a_1, a_2, ..., a_n$ - часткові коефіцієнти регресії, які показують як в середньому зміниться результативна змінна при збільшенні і-ої факторної ознаки на 1 власного виміру, за умови, що решта факторів залишаються незмінними.

Оцінки параметрів загальної економетричної моделі повинні мати такі *властивості:*

1) незміщеності;

2) обгрунтованості;

3) ефективності;

4) інваріантності.

Оцінка параметра моделі буде *незміщеною,* коли дотримується рівність:

$$M(\hat{A}) = A .$$

Якщо ця рівність не дотримується, то різниця $M(\hat{A}) - A = Q$ називається зміщенням оцінки.

Оцінка параметра моделі буде *обгрунтованою,* якщо при заданій малій величині $\varepsilon > 0$ справедливе відношення:

$$\lim_{n \to \infty} P \left\{ \left| \hat{A} - A \right| < \varepsilon \right\} = 1 .$$

Оцінки $\hat{A}$ параметрів A називаються *ефективними,* коли вони мають найменшу дисперсію.

Якщо функція $g(\hat{A})$ відповідає функції $g(A)$, то оцінки $\hat{A}$ параметрів A є інваріантними.

Існує декілька методів побудови "найкращої" лінійної регресії, найбільш поширеними серед яких є:

1) метод усіх можливих регресій ;

2) метод покрокової регресії ;

3) метод виключень .

1) Метод усіх можливих регресій

Метод усіх можливих регресій – історично перший метод побудови лінійних регресійних моделей і найбільш громіздкий серед усіх методів. Ідея методу полягає у побудові множини регресійних рівнянь, які містять усі можливі комбінації попередньо відібраних факторів, і у порівнянні цих рівнянь за трьома критеріями : коефіцієнтом детермінації R^2, стандартною похибкою $\hat{\sigma}_\varepsilon$ і критерієм Меллоуза C_p. У загальному випадку для **m** відібраних факторів (пояснюючих змінних) можна побудувати 2^m рівнянь регресії і виконати їх порівняння.

Побудова і аналіз усіх можливих регресійних рівнянь є доволі громіздка і ненадійна процедура, тому цей метод рекомендується використовувати при невеликій кількості відібраних факторів.

2) Метод покрокової регресії

Цей метод є найпоширенішим на практиці і більш економним у порівнянні з попереднім. Ідея методу полягає у послідовному включенні до моделі факторів (пояснюючих змінних) до тих пір, поки модель не стане задовільною. Порядок включення факторів до моделі вибирається на основі значень коефіцієнтів парної кореляції між пояснюючими і залежною змінною моделі. Алгоритм методу покрокової кореляції можна подати у наступному вигляді :

1. Розраховується кореляційна матриця **r** для усіх змінних моделі, які планується включити до моделі.

2. Спочатку з кореляційної матриці вибирається і включається до моделі той фактор $x_j, \left(j=\overline{1,m}\right)$, якому у кореляційній матриці відповідає найбільший за модулем коефіцієнт парної кореляції з залежною змінної моделі **y** (нехай це буде змінна x_1). Будується регресійне рівняння з однією незалежною змінною $\hat{y}=b_0+b_1x_1$ і для нього обчислюється коефіцієнт детермінації. Після цього перевіряється чи значима ця змінна за коефіцієнтом

детермінації і за частковим F- критерієм. Якщо ні, то приймаємо $y = \bar{y}$ і процес побудови моделі припиняється. Якщо так, то переходимо до наступного кроку 3.

3. На основі аналізу кореляційної матриці серед тих пояснюючих змінних, що залишились, шукаємо нову змінну, яка має найбільший за модулем коефіцієнт кореляції з **y** і включаємо її до моделі (нехай це буде змінна x_2) .

4. Будується нове рівняння регресії :

$$\hat{y} = b_0 + b_1 x_1 + b_2 x_2$$

і для нього розраховується звичайний R^2 і оцінений $\bar{R}^2$ коефіцієнт детермінації. Аналізується зміна цих показників у порівнянні з попередньою моделлю. Потім розраховуються часткові F- критерії для кожного фактора. Серед них обирається найменше значення і порівнюється із заздалегідь обраним критичним значенням F - критерію. В залежності від результатів перевірки додана на цьому кроці змінна або залишається у моделі, або відкидається.

5. Після цього модель перераховується в залежності від факторів, які залишились і здійснюється перехід до кроку 3.

Процес побудови моделі за наведеним алгоритмом припиняється, якщо жодний фактор, що знаходиться у поточному рівнянні , не вдається виключити, а новий претендент на включення не відповідає частковому F - критерію.

3). Метод виключень

Метод виключень діє у зворотному порядку порівняно з методом покрокової регресії і є також досить поширеним. Загальний алгоритм методу складається з 5 кроків.

1. Будується рівняння регресії, яке включає всі відібрані фактори . Якщо попередньо було відібрано m факторів , то вихідне базове рівняння має вигляд :

$$\hat{y} = b_0 + b_1 x_1 + b_2 x_2 + \ldots + b_m x_m \ .$$

2. Для кожного фактора (пояснюючої змінної) $x_j, \left(j=\overline{1,m}\right)$ обчислюється значення часткового F- критерію.

3. Серед розрахованих значень часткового F- критерію вибирається найменше Fmin і порівнюється із заздалегідь обраним критичним значенням розподілу Фішера Fкр .

4. Якщо Fmin < Fкр , то відповідний фактор виключається з рівняння. Проводиться новий розрахунок регресійного рівняння вже без виключеного фактора і здійснюється перехід знову до кроку 2.

5. Якщо Fmin > Fкр , то регресійне рівняння залишається без змін.

При розгляді методів побудови загальної лінійної економетричної моделі були використані такі нові поняття і статистичні показники, як частковий F – критерій ,оцінений коефіцієнт детермінації ,кореляційна матриця. Розглянемо ці показники більш детальніше.

Одним з головних питань будь-якого методу побудови багатофакторної регресійної моделі є питання визначення суттєвості впливу на залежну змінну *у* окремих факторів. Таку оцінку можна зробити з використанням F- статистики на основі часткового F- критерію Фішера. Зміст часткового F- критерію розглянемо на наступному прикладі.

Кореляційна матриця дозволяє оцінити щільність лінійного кореляційного зв'язку між залежною змінною моделі і окремими факторами, а також між окремими незалежними змінними. У загальному випадку вона представляє собою квадратну симетричну матрицю, елементами якої є коефіцієнти парної кореляції між залежною змінною моделі і кожною пояснюючою змінною, а також коефіцієнти парної кореляції між самими пояснюючими змінними моделі.

Для випадку **m** пояснюючих змінних кореляційна матриця має наступний вигляд і структуру :

$$r = \begin{pmatrix} r_{yy} & r_{yx_1} & r_{yx_2} & \cdots & r_{yx_m} \\ r_{x_1y} & r_{x_1x_1} & r_{x_1x_2} & \cdots & r_{x_1x_m} \\ r_{x_2y} & r_{x_2x_1} & r_{x_2x_2} & \cdots & r_{x_2x_m} \\ \cdots & \cdots & \cdots & \cdots & \cdots \\ r_{x_my} & r_{x_mx_1} & r_{x_mx_2} & \cdots & r_{x_mx_m} \end{pmatrix}, \dim r = (m+1)\times(m+1). \qquad (67)$$

Діагональні елементи матриці r дорівнюють 1. Коефіцієнти парної кореляції r_{yx_j} та $r_{x_jx_k}$ обчислюються за відомою формулою

Побудова багатофакторних моделей відбувається на основі покрокової регресії. В комп'ютерних пакетах використовуються такі основні процедури покрокового відбору: послідовного приєднання регресорів (факторних змінних), приєднання – вилучення та послідовного вилучення регресорів.

Найпростішими економетричними моделями є виробничі функції; які мають вигляд:

$Y = AK^{\alpha}L^{\beta}V$, де

Y – обсяг випуску проукції;

K – капітальні витрати;

L – витрати праці;

V – випадковий член;

A – числовий параметр;

α і β - коефіцієнти еластичності випуску за капіталом і працею, незалежні між собою.

Окремим випадком є виробнича функція Кобба-Дугласа, коли $\alpha + \beta = 1$:

$Y = FK^{\alpha}L^{\beta}$.

Виробничі функції приводять до лінійного вигляду за допомогою логарифмування:

$\log Y = \log A + \alpha \log K + \beta \log L + \log V$,

де α і β характеризуються як елементи еластичності

Приклад 4.1. Побудувати економетричну модель, яка характеризує залежність між витратами на харчування, загальними затратами та складом сім'ї на основі даних, наведених у табл. 2.1. Проаналізувати зв'язок, визначений на основі побудованої моделі.

Таблиця 2.1

№ п / п	Витрати на харчування	Загальні затрати	Склад сім'ї
1	*2*	*3*	*4*
1	20	45	1,5
2	32	75	1,6
3	48	125	1,9
4	65	223	1,8
5	45	92	3,4
6	64	146	3,6
7	79	227	3,5
8	104	358	5,5
9	68	135	5,4
10	93	218	5,4
11	117	331	5,3
12	145	490	8,5
13	91	175	8,3
14	131	205	8,1
15	167	468	7,3
16	195	749	8,4

Розв'язання:

1. Ідентифікуємо змінні моделі:

Y — витрати на харчування (залежна змінна);

X_1 — загальні витрати (незалежна змінна);

X_2 — розмір сім'ї (незалежна змінна);

u — залишки (стохастична складова).

Загальний вигляд моделі:

$$Y = f(X_1, X_2, u) \ .$$

2. Специфікуємо модель, тобто в даному випадку визначимо її аналітичну форму:

$$Y = a_0 + a_1 X_1 + a_2 X_2 + u;$$

$$\hat{Y} = \hat{a}_0 + \hat{a}_1 X_1 + \hat{a}_2 X_2 .$$

3. Оцінимо параметри моделі на основі методу 1МНК, попередньо висунувши гіпотезу, що всі чотири передумови для його застосування дотримані.

Оператор оцінювання на основі 1МНК:

$$\hat{A} = (X'\,X)^{-1}\,X'\,Y.$$

У даному операторі матриця X характеризує всі незалежні змінні моделі. Оскільки економетрична модель має вільний член $\hat{a}_0$, для якого всі $x_i = 1$, то матрицю X треба доповнити першим стовпцем, в якому всі шістнадцять членів є одиницями. X' — матриця, транспонована до матриці X, а вектор Y — вектор залежної змінної.

Таким чином, $\hat{a}_0 = 9{,}596$; $\hat{a}_1 = 0{,}1837$; $\hat{a}_2 = 6{,}854$. Звідси еконо-метрична модель має вигляд:

$$\hat{Y} = 9{,}596 + 0{,}1837X_1 + 6{,}854X_2.$$

4. Визначимо розрахункові значення залежної змінної $\hat{Y}$ на основі моделі, підставивши в неї значення незалежних змінних X_1 та X_2 . Потім віднімемо розрахункові значення $\hat{Y}$ від фактичних Y, в результаті отримаємо залишки: $u = Y - \hat{Y}$. Всі ці розрахунки наведені в табл. 2.2.

Таблиця 2.2

№ п / п	$\hat{Y}$	u	u^2	$Y - \overline{Y}$	$\left(Y - \overline{Y}\right)^2$
1	28,1424	–8,1424	66,2979	–71,5000	5112,2500
2	34,3382	–2,3382	5,4673	–59,5000	3540,2500
3	45,5785	2,4215	5,8637	–43,5000	1892,2500
4	63,9961	1,0039	1,0077	–26,5000	702,2500
5	49,7976	–4,7976	23,0169	–46,5000	2162,2500
6	61,0872	2,9128	8,4843	–27,5000	756,2500
7	75,2802	3,7198	13,8372	–12,5000	156,2500
8	113,0501	–9,0501	81,9051	12,5000	156,2500
9	71,4035	–3,4035	11,5837	–23,5000	552,2500
10	86,6492	6,3508	40,3332	1,5000	2,2500
11	106,7200	10,2800	105,6793	23,5000	650,2500
12	157,8576	–12,8576	165,3171	53,5000	2862,2500
13	98,6267	–7,6267	58,1665	–0,5000	0,2500
14	121,1347	9,8653	97,3237	39,5000	1560,2500
15	145,5920	21,4080	458,3011	75,5000	5700,2500
16	204,7461	–9,7461	94,9859	103,5000	10712,2500
Всього		0,0000	1237,5704		36518,0000

5. Розрахуємо дисперсії залишків та залежної змінної Y:

$$\sigma^2_{\hat{a}_0} = 29{,}303\ ;$$
$$\sigma^2_{\hat{a}_1} = 0{,}0003787\ ;$$
$$\sigma^2_{\hat{a}_2} = 1{,}95043834\ .$$

Інші елементи даної матриці визначають рівень коваріації між оцінками параметрів моделі.

7. Знайдемо стандартні помилки оцінок параметрів:

$$S_{\hat{a}_0} = \sqrt{\sigma^2_{\hat{a}_0}} = \sqrt{29{,}303} = 5{,}413\ ;$$
$$S_{\hat{a}_1} = \sqrt{\sigma^2_{\hat{a}_1}} = \sqrt{0{,}0003787} = 0{,}01946\ ;$$
$$S_{\hat{a}_2} = \sqrt{\sigma^2_{\hat{a}_2}} = \sqrt{1{,}95043834} = 1{,}39658\ .$$

Порівняємо стандартні помилки оцінок параметрів моделі з величиною оцінки. Так, співвідношення стандартної помилки й абсолютного значення параметра $\hat{a}_0$ становить 56% , параметра $\hat{a}_1$ — 10,6%, параметра $\hat{a}_2$ — Побудова багатофакторних моделей відбувається на основі покрокової регресії. В комп'ютерних пакетах використовуються такі основні процедури покрокового відбору: послідовного приєднання регресорів (факторних змінних), приєднання – вилучення та послідовного вилучення регресорів.

Перше й третє співвідношення свідчать про те, що оцінки параметрів моделі $\hat{a}_0$ і $\hat{a}_2$ можуть мати зміщення, а друге співвідношення підтверджує незміщеність оцінки параметра $\hat{a}_1$.

8. Дамо змістовне тлумачення параметрів моделі.

Оцінка параметра $\hat{a}_1$ характеризує граничну зміну величини витрат на харчування залежно від зміни загальних затрат на одиницю. Тобто, якщо загальні затрати сім'ї зростуть на одиницю, то витрати на харчування в них збільшаться на **0,18** одиниці при незмінному складі сім'ї.

Оцінка параметра $\hat{a}_2$ характеризує граничне зростання витрат на харчування при збільшенні сім'ї на одного члена. Так, якщо склад сім'ї збагатиться ще одним членом, то витрати на харчування зростуть на **6,854** одиниці при незмінній величині доходу.

Завдання для самостійної роботи

Завдання 2.1. Побудувати економетричну модель, що характеризує залежність між витратами обігу, обсягом вантажообороту та фондомісткістю бази. Визначити стандартні помилки параметрів. Дати змістовне тлумачення взаємозв'язку. Вихідні дані наведені в табл. 2.3-2.4.

Таблиця 2.3

N п/п	Витрати	Вантажо-оборот	Фондо-місткість
1	2,72	15,6	106,3
2	3,04	13,5	128,5
3	2,84	15,3	118,0
4	2,89	14,9	121,2
5	2,58	15,1	120,0
6	2,64	16,1	118,4
7	2,52	16,7	108,4
8	2,75	15,4	110,0
9	2,63	17,1	105,9

Таблиця 2.4

N п/п	Витрати	Вантажо-оборот	Фондо-місткість
1	2,58	15,1	120,0
2	2,64	16,1	118,4
3	2,52	16,7	108,4
4	2,75	15,4	110,0
5	2,63	17,1	105,9
6	2,48	16,8	117,7
7	2,62	16,9	97,5
8	2,88	16,1	113,7
9	2,68	15,0	122,3

Тема 5. Мультиколінеарність

У попередніх темах розглядалися економетричні моделі, які будувалися на основі припущень класичної лінійної регресії. Параметри цих моделей, які оцінювалися на основі МНК, внаслідок виконання цих припущень мали властивості BLUE – оцінок, а самі моделі по суті були класичними регресійними моделями.

Мультиколінеарність – явище, при якому між факторними ознаками існує зв'язок близький до функціонального $(rx_i x_j \to 1)$.

При двофакторній моделі для оцінки наявності мультиколінеарності достатньо розрахувати парний лінійний коефіцієнт кореляції. Прибагатофакторній моделі високі коефіцієнти кореляції $(rx_i x_j > 0{,}8)$ - достатня, але не необхідна умова наявності мультиколінеарності.

Ознаки мультиколінеарності

1. Якщо серед парних коефіцієнтів кореляції незалежних змінних є такі, рівень яких наближається або дорівнює множинному коефіцієнту кореляції, то

це свідчить про можливість існування мультиколінеарності. Інформацію про парну залежність може дати симетрична матриця коефіцієнтів парної кореляції, або кореляції нульового порядку:

$$r = \begin{pmatrix} r_{X_1X_1} & r_{X_1X_2} & r_{X_1X_3} & \cdots & r_{X_1X_K} \\ r_{X_2X_1} & r_{X_2X_2} & r_{X_2X_3} & \cdots & r_{X_2X_K} \\ r_{X_3X_1} & r_{X_3X_2} & r_{X_3X_3} & \cdots & r_{X_3X_K} \\ \cdots & \cdots & \cdots & \cdots & \cdots \\ r_{X_KX_1} & r_{X_KX_2} & r_{X_KX_3} & \cdots & r_{X_KX_K} \end{pmatrix}.$$

Але якщо в моделі фігурує більше двох незалежних змінних, вивчення питання про мультиколінеарність не може обмежуватись інформацією, що дає ця матриця. Явище мультиколінеарності ні в якому разі не зводиться тільки до існування парної кореляції між незалежними змінними.

Більш загальна перевірка передбачає визначення визначника (детермінанта) матриці r, який називається детермінантом кореляції і позначається $|r|$. Числові значення детермінанта кореляції знаходяться на множині: $|r| \in [0,1]$.

2. Якщо $|r| = 0$, то існує повна мультиколінеарність, якщо $|r| = 1$— мультиколінеарність відсутня, чим ближче $|r|$ до нуля, тим певніше можна стверджувати, що між незалежними змінними існує мультиколінеарність. Незважаючи на те, що на числове значення $|r|$ впливає дисперсія незалежних змінних, цей показник можна вважати точковою мірою тісноти мультиколінеарності.

3. Якщо в економетричній моделі одержано мале значення параметра $\hat{a}_k$ при високому рівні коефіцієнта детермінації і при цьому F-критерій суттєво відрізняється від нуля, то це також свідчить про наявність мультиколінеарності.

4. Якщо коефіцієнт детермінації R^2, що розрахований для регресійних залежностей між однією незалежною змінною та іншими, має значення, яке близьке до одиниці, то можна говорити про наявність мультиколінеарності.

5. Якщо при побудові економетричної моделі на основі покрокової регресії включення нової незалежної змінної суттєво змінює оцінку параметрів моделі при незначному підвищенні (або зниженні) коефіцієнтів кореляції чи

детермінації, то ця змінна, очевидно, знаходиться в лінійній залежності від інших, які введені в модель раніше.

Всі ці методи виявлення мультиколінеарності мають один загальний недолік: жоден із них не проводить чіткої межі між тим, що треба вважати «суттєвою» мультиколінеарністю, яку треба враховувати, і тим, коли мультиколінеарністю можна знехтувати.

Таким чином природа мультиколінеарності полягає у неможливості статистично оцінити і обґрунтувати вплив кожної пояснюючої змінної на залежну зміну моделі, що ,в свою чергу, робить ненадійною економічну інтерпретацію оціненого рівняння регресії.

Це в свою чергу призводить до наступних негативних практичних наслідків :

1. Збільшення інтервалів довіри параметрів моделі .
2. Статистична незначимість оцінок деяких параметрів моделі. Це пов'язано із зменшенням t – статистики для деяких параметрів. Внаслідок цього із моделі можуть бути вилучені змінні, які за економічним змістом як раз суттєво впливають на залежну змінну.
3. Оцінки параметрів стають чутливими до розміру статистичної вибірки. Збільшення сукупності спостережень внаслідок цього іноді може привести до істотних змін в оцінках параметрів.

Таким чином, мультиколінеарність загалом негативно впливає на кількісні характеристики економетричної моделі, або робить її побудову взагалі неможливою. Внаслідок цього важливим стає питання тестування наявності мультиколінеарності у моделі і вилучення її.

Алгоритм Феррара—Глобера

Найбільш повне дослідження мультиколінеарності можна здійснити на основі алгоритму Феррара—Глобера. Цей алгоритм включає три види статистичних критеріїв, на основі яких перевіряється мультиколінеарність всього масиву незалежних змінних (χ^2, хі-квадрат); кожної незалежної змінної зі всіма незалежними змінними (F-критерій) і мультиколінеарність кожної пари незалежних змінних (t-критерій).

Всі ці критерії при порівнянні з їх критичними значеннями дають можливість зробити конкретні висновки відносно наявності чи відсутності мультиколінеарності незалежних змінних.

Опишемо алгоритм Феррара—Глобера.

Крок 1. Стандартизація (нормалізація) змінних.

Позначимо вектори незалежних змінних економетричної моделі через $X_1, X_2, X_3, \ldots, X_m$. Елементи стандартизованих векторів розрахуємо за формулою:

$$X^*_{ik} = \frac{X_{ik} - \overline{X}_k}{\sqrt{n\sigma^2_{X_k}}},$$

де n — число спостережень, $(i = \overline{1,n})$;

m — число незалежних змінних, $(k = \overline{1,m})$;

$\overline{X}_k$ — середня арифметична k — її незалежної змінної;

$\sigma^2_{X_k}$ — дисперсія k-ї незалежної змінної.

Крок 2. Знаходження кореляційної матриці (матриці моментів стандартизованої системи нормальних рівнянь):

$$R = X^{*'} X^*,$$

де X^* — матриця стандартизованих незалежних змінних;

$X^{*'}$ — матриця, транспонована до матриці X^*.

Крок 3. Розрахунок F- критеріїв:

$$F_k = (c_{kk} - 1)\,\frac{n-m}{m-1},$$

де c_{kk} — діагональні елементи матриці C. Фактичні значення критеріїв F_k порівнюються з табличними при $n-m$ і $m-1$ ступенях свободи і рівні значущості α. Якщо $F_{k\ \text{факт}} > F_{\text{табл}}$, відповідна k-та незалежна змінна мультиколінеарна з іншими.

Коефіцієнт детермінації для кожної змінної розраховується таким чином:

$$R_{X_k}^2 = 1 - \frac{1}{c_{kk}}.$$

Крок 4. Знаходження часткових коефіцієнтів кореляції:

$$r_{kj} = \frac{-c_{kj}}{\sqrt{c_{kk} * c_{jj}}} ,$$

де c_{kj} — елемент матриці C, що заходиться в k-му рядку і j-му стовпці, $k = \overline{1,m}$, $j = \overline{1,m}$, c_{kk} і c_{jj} — діагональні елементи матриці C.

Крок 5. Розрахунок t критеріїв:

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

Фактичні значення критеріїв t_{kj} порівнюються з табличними при $n-m$ ступенях свободи і рівні значущості α. Якщо $t_{kj\ \text{факт}} > t_{\text{табл}}$, між незалежними змінними X_k і X_j існує мультиколінеарність.

У випадку виявлення наявності мультиколінеарності існує декілька простих шляхів її усунення. Основними серед них є наступні.

1. Вилучення змінної (або змінних) з моделі. При цьому з моделі вилучається одна із змінних колінеарної пари. Слід зазначити, що таке вилучення змінних можливе тільки у випадку коли це не суперечить логіці економічних зв’язків. У протилежному випадку це може призвести до помилки специфікації.

2. Зміна аналітичної форми економетричної моделі. Іноді заміна однієї функції регресії іншою (наприклад лінійної нелінійною) , якщо це не суперечить апріорній інформації, дає змогу уникнути явища мультиколінеарності.

3. Збільшення спостережень. З точки зору теорії, мультиколінеарність та невелика кількість спостережень у вибірці – це одна і та ж проблема. Тому збільшення спостережень у статистичній вибірці або використання іншої статистичної вибірки може усунути, або принаймні зменшити вплив мультиколінеарності.

4. Перетворення статистичних даних. Позбутися мультиколінеарності можна і шляхом наступних перетворень вихідних даних стосовно пояснюючих змінних :

а) замість самих даних узяти їхні відхилення від середніх;

б) замість абсолютних значень даних взяти відносні значення ;

в) стандартизувати змінні.

Приклад 5.1. На середньомісячну заробітну плату впливає ряд факторів. Виділимо серед них продуктивність праці, фондомісткість та коефіцієнт плинності робочої сили. Щоб побудувати економетричну модель заробітної плати від наведених чинників на основі методу найменших квадратів, треба переконатись, що продуктивність праці, фондомісткість та коефіцієнт плинності робочої сили як незалежні змінні — не мультиколінеарні.

Вихідні дані наведені в табл. 4.1.

Таблиця 5.1

Номер цеха	Продуктивність праці, млн.грн./ люд.	Фондомісткість, грн./грн.	Коефіцієнт плинності робочої сили, %
1	32	0,59	10,5
2	29	0,43	15,5
3	30	0,70	13,5
4	31	0,61	9,5
5	25	0,51	2,5
6	34	0,51	1,5
7	29	0,65	17,5
8	24	0,43	14,5
9	20	0,51	14,5
10	35	0,92	7,5

Розв'язання

Нормалізація змінних.

Позначимо вектори незалежних змінних — продуктивності праці, фондомісткості, коефіцієнтів плинності робочої сили — через відповідно X_1, X_2, X_3. Елементи стандартизованих векторів розрахуємо за формулою:

$$X_{ik}^{*} = \frac{X_{ik} - \overline{X}_k}{\sqrt{n\sigma_{X_k}^2}},$$

де n — кількість спостережень, $n = 10$;

m — число незалежних змінних, $m = 3$;

$\overline{X_k}$ — середня арифметична вектора X_k ;

$\sigma_{X_k}^2$ — дисперсія змінної X_k .

$$\overline{X_1} = \frac{\sum_{t=1}^{10} X_{i1}}{n} = \frac{287}{10} = 28{,}7;$$

$$\overline{X}_2 = \frac{\sum_{i=1}^{10} X_{i2}}{n} = \frac{5{,}86}{10} = 0{,}586;$$

$$\overline{X}_3 = \frac{\sum_{i=1}^{10} X_{i3}}{n} = \frac{139}{10} = 13{,}9 .$$

В табл. 4.2 наведені всі розрахунки по стандартизації незалежних змінних X_1, X_2, X_3 згідно з наведеним співвідношенням.

Таблиця 4.2

$X_{i1} - \overline{X_1}$	$X_{i2} - \overline{X_2}$	$X_{i3} - \overline{X_3}$	$\left(X_{i2} - \overline{X_2}\right)^2$	$\left(X_{i2} - \overline{X}_2\right)^2$	$\left(X_{i3} - \overline{X}_3\right)^2$	X_{i1}^{*}	X_{i2}^{*}	X_{i3}^{*}
3.3	0,004	–3,4	10,89	0,000015	11,56	0,2487	0,0091	–0,2518
0.3	–0,156	1,6	0,09	0,024336	2,56	0,0226	–0,2531	0,1185
1.3	0,114	–0,4	1,69	0,012996	0,16	0,0979	0,2580	–0,0296
2.3	0,024	–4,4	5,29	0,000576	19,36	0,1733	0,0543	-0,3258
3.7	–0,076	9,6	13,09	0,005776	92,16	–0,2788	–0,1720	0,7108
5.3	–0,076	–1,4	23,09	0,005776	1,96	0,3994	–0,1720	–0,1037
0.3	0,064	3,5	10,09	0,004095	12,95	0,0226	0,1448	0,2666
–4.7	–0,156	0,6	22,09	0,024336	0,36	–0,3542	–0,3531	0,0444
–8.7	–0,076	0,6	75,69	0,005776	0,36	–0,6556	–0,1720	0,0444
4.3	0,334	–6,4	14,49	0,111556	40,95	0,3240	0,7559	–0,4739
Σ			17,1	0,19524	182,4			

Дисперсії кожної незалежної змінної мають такі значення:

$$\sigma_{X_1}^2 = \frac{\sum_{i=1}^{10} (X_{i1} - \overline{X_1})^2}{n} = \frac{176{,}1}{10} = 17{,}61 ;$$

$$\sigma_{X_2}^2 = \frac{\sum_{i=1}^{10} (X_{i2} - \overline{X_2})^2}{n} = \frac{0{,}19524}{10} = 0{,}0195 ;$$

$$\sigma_{X_3}^2 = \frac{\sum_{i=1}^{10}(X_{i3} - \overline{X_3})^2}{n} = \frac{182,4}{10} = 18,24 \; .$$

Тоді знаменник для стандартизації кожної незалежної змінної буде дорівнювати:

для X_1: $\sqrt{n\sigma_{X\,1}^2} = \sqrt{10 \cdot 19,57} = \sqrt{195,7} = 13,27$;

для X_2: $\sqrt{n\sigma_{X\,2}^2} = \sqrt{10 \cdot 0,0217} = \sqrt{0,217} = 0,44$;

для X_3: $\sqrt{n\sigma_{X\,3}^2} = \sqrt{10 \cdot 20,27} = \sqrt{202,7} = 13,51$.

Знаходження кореляційної матриці R:

$$R = X^{*\,'} X^{*},$$

де $X^{*'}$ — матриця транспонована до матриці X^{*}.

Кожен елемент цієї матриці характеризує тісноту зв’язку однієї незалежної змінної з іншою. Оскільки діагональні елементи характеризують тісноту зв’язку кожної незалежної змінної з цією самою змінною, то вони дорівнюють одиниці.

Інші елементи матриці R трактуються так:

$$r_{X_1 X_2} = 0,494;$$

$$r_{X_1 X_3} = -0,551 \; ;$$

$$r_{X_2 X_3} = -0,517 \; ,$$

тобто вони є парними коефіцієнтами кореляції незалежних змінних. На основі цих коефіцієнтів можна зробити висновок, що між змінними X_1, X_2, X_3 існує зв’язок. Але чи можна стверджувати, що цей зв’язок є явищем мультиколінеарності і він негативно впливатиме на оцінку економетричної моделі?

Щоб відповісти на це запитання, треба продовжити розв’язання на основі алгоритму Феррара—Глобера і в результаті знайти статистичні критерії оцінки мультиколінеарності.

Знайдемо матрицю, обернену до матриці R:

$$C = R^{-1} = (X^{*'} X^{*})^{-1};$$

Використовуючи діагональні елементи матриці C, розрахуємо F-критерії:

$$F_1 = (c_{11}-1)\ \frac{n-m}{m-1} = (1{,}572-1)\frac{7}{2} = 2{,}00\ ;$$

$$F_2 = (c_{22}-1)\ \frac{n-m}{m-1} = (1{,}492-1)\frac{7}{2} = 1{,}72\ ;$$

$$F_3 = (c_{33}-1)\ \frac{n-m}{m-1} = (1{,}620-1)\frac{7}{2} = 2{,}17\ .$$

При рівні значущості $\alpha = 0{,}05$ і ступенях свободи $\gamma_1 = 7$ і $\gamma_2 = 2$ критичне (табличне) значення критерію $F = 4{,}74$.

Через те, що $F_{1\,\text{факт}} < F_{\text{табл}}$,

$F_{2\,\text{факт}} < F_{\text{табл}}$,

$F_{3\,\text{факт}} < F_{\text{табл}}$,

то жодна із незалежних змінних не мультиколінеарна з двома іншими.

Щоб визначити наявність попарної мультиколінеарності, продовжимо розрахунок і перейдемо до кроку 6.

Розрахуємо часткові коефіцієнти кореляції, використавши елементи матриці C:

$$r_{12\cdot 3} = \frac{-c_{12}}{\sqrt{c_{11}c_{22}}} = \frac{0{,}448}{\sqrt{1{,}572\cdot 1{,}492}} = 0{,}293\ ;$$

$$r_{13\cdot 2} = \frac{-c_{13}}{\sqrt{c_{11}c_{33}}} = \frac{-0{,}634}{\sqrt{1{,}572\cdot 1{,}620}} = -0{,}397\ ;$$

$$r_{23\cdot 1} = \frac{-c_{23}}{\sqrt{c_{22}c_{33}}} = \frac{-0{,}524}{\sqrt{1{,}492\cdot 1{,}620}} = -0{,}337\ .$$

Часткові коефіцієнти кореляції характеризують тісноту зв'язку між двома змінними за умови, що третя не впливає на цей зв'язок.

Порівнявши часткові коефіцієнти кореляції з парними, які наведені вище, можна помітити, що часткові коефіцієнти значно менше парних. Це

ще раз підтверджує, що на основі парних коефіцієнтів кореляції не можна зробити висновки про наявність чи відсутність мультиколінеарності.

Визначимо t- критерії на основі часткових коефіцієнтів кореляції:

$$t_{12} = \frac{r_{12\cdot 3}\sqrt{n-m}}{\sqrt{1-r_{12\cdot 3}^2}} = \frac{0{,}293\sqrt{7}}{\sqrt{1-0{,}0858}} = 0{,}811;$$

$$t_{13} = \frac{r_{13\cdot 3}\sqrt{n-m}}{\sqrt{1-r_{13\cdot 3}^2}} = \frac{0{,}39\sqrt{7}}{\sqrt{1-0{,}1521}} = 1{,}146;$$

$$t_{23} = \frac{r_{23\cdot 3}\sqrt{n-m}}{\sqrt{1-r_{23\cdot 3}^2}} = \frac{0{,}34\sqrt{7}}{\sqrt{1-0{,}1156}} = 0{,}947.$$

Табличне значення t- критерію при $n - m = 7$ ступенях свободи і рівні значущості $\alpha = 0{,}05$ дорівнює 1,89. Всі числові значення t- критеріїв, знайдених для кожної пари змінних, менше за їх табличне значення. Звідси робимо висновок, що всі пари незалежних змінних не є мультиколінеарними.

Таким чином, незважаючи на те, що між незалежними змінними, що досліджуються, існує лінійна залежність, але вона не є явищем мультиколінеарності і не буде негативно впливати на кількісні параметри економетричної моделі.

Якщо F- критерій більше табличного значення, а це значить, що k-та змінна залежить від всіх інших в масиві, то необхідно вирішувати питання про її виключення з переліку змінних.

Якщо t_{kj} - критерій більше табличного, то ця пара змінних (k і j) тісно взаємопов’язані. Звідси, аналізуючи рівень обох видів критеріїв F і t, можна зробити обгрунтований висновок про те, яку із змінних необхідно виключити із дослідження чи замінити іншою. Але заміна масиву незалежних змінних завжди повинна узгоджуватись із економічною доцільністю, що випливає з мети дослідження.

Завдання для самостійної роботи

Завдання 5.1. Нехай на витрати обігу впливають: обсяг вантажообороту, запаси по вантажообороту та трудомісткість його одиниці. Щоб побудувати економетричну модель цієї залежності на основі методу 1МНК, необхідно бути впевненим, що між факторами вантажообороту, запасів та трудомісткості не існує мультиколінеарності. Треба дослідити наявність мультиколінеарності між цими факторами на основі даних, що наведені в табл. 4.3 — 4.4.

Дані кожної таблиці є одним із варіантів завдання 4.1.

Таблиця 4.3

№ п / п	Вантажо-оборот	Запаси	Трудоміст-кість
1	15,6	40,4	2,11
2	13,5	38,9	2,78
3	15,3	36,6	2,17
4	14,9	41,4	2,15
5	15,1	32,2	2,11
6	16,1	31,4	1,97
7	16,7	32,6	1,96
8	15,4	38,7	2,12
9	17,1	44,3	2,02
10	16,8	39,3	2,13

Таблиця 4.4

№ п / п	Вантажо-оборот	Запаси	Трудо-місткість
1	15,1	32,0	2,12
2	16,1	32,4	1,98
3	16,7	32,6	1,96
4	15,5	38,7	2,15
5	17,2	44,3	2,02
6	16,9	39,3	2,05
7	17,0	40,4	2,02
8	16,2	41,5	2,13
9	15,0	45,2	2,14
10	18,0	50,2	1,90

Тема 6. Узагальнений метод найменших квадратів

Одним з основних припущень класичної лінійної регресії, яке дозволяє коректно застосувати для оцінювання параметрів моделі МНК, є припущення про сталість дисперсії стохастичної складової ε_i, тобто припущення про гомоскедастичність стохастичної складової економетричної моделі.

В попередніх темах розглядалася класична лінійна модель, яка має відповідати чотирьом умовам Гауса – Маркова. Одним з основних припущень є припущення про сталість дисперсій випадкової величини E_i:

$$Var(E_i) = \sigma_E^2 = const.$$

Ця умова відома як гомоскедастичність, що означає “однакове розкидання”. Іншими словами, імовірність того,що величина Е прийме будь–яке додатнє (чи від’ємне значення), буде однаковою для всіх одиниць спостереження, тобто дисперсія σ_E^2 не є функцією $x_{ij} - \sigma_E^2 \neq f(x_{1i}, x_{2i} \ldots x_{pi})$.

Коли порушується умова сталості дисперсії:

$$Var(E_i) = \sigma_E^2 \neq const$$

Маємо справу з гетероскедатичностю, що означає “неоднакове розкидання”. Іншими словами $\sigma_E^2 = f(x_{1i}, x_{2i} \ldots x_{pi})$.

Якщо має місце гетероскедатичність, то оцінки, отримані за МНК, є неефективними, а точність оцінок зменшується. Гетероскедатичність стає проблемою, коли значення змінних в рівнянні регресії згачно розрізняються в різних спостереженнях. Вона може виникати при аналізі динамічних рядів.

Визначити гетероскедатичність можливо за допомогою відповідних тестів6 рангової кореляції Спірмена, Голфелда-Квандта, Глейзера тощо. Якщо результати тестів підтверджують наявність гетероскедатичності використовувати для оцінки параметрів МНК недоцільно. В даному випадку потрібно скористатися узагальненим методом найменших квадратів (УМНК), який був запропонований А.Ейткеном, для трансформованої моделі. Трансформовану модель отримуємо в результаті ділення всіх складових частин лінійного рівняння регресії на σ_i^2. Для трансформованої моделі за УМНК оцінка α_1 дорівнює:

$$a_1' = \frac{\sum j_i x_i y_i - \sum j_i x_i \sum j_i y_i}{(\sum j_i)(\sum j_i x_i^2) - (\sum j_i x_i)^2},$$ де

$j_i = 1/\sigma_i^2$ - вагові коефіцієнти.

Загальний вигляд УМНК записується таким чином: $\sum j_i E^2 = \sum j_i (y_i - a_0' - a_1' x_i)^2 \to \min.$

Використання в умовах гетероскедатичності МНК приводить до збільшення інтервалів довіри параметрів, і, як наслідок, t-та F- критерії дають неточні результати за рахунок того, що $Var(a_1)$ надмірно велика.

Наявність гетероскедастичності спричиняє порушення властивостей оцінок параметрів моделі при розрахунку їх за методом 1МНК. Тому завжди виникає необхідність вивчати це явище, і, якщо воно існує, для оцінки параметрів моделі використовувати узагальнений метод найменших квадратів (метод Ейткена).

Для визначення гетероскедастичності застосовуються чотири критерії:

1) $F_{\text{критерій}}$ μ;

2) параметричний тест Гольдфельда—Квандта;

3) непараметричний тест Гольдфельда—Квандта;

4) тест Глейсера.

1. Критерій μ

Цей метод застосовується в тих випадках, коли вихідна сукупність спостережень досить велика. Розглянемо цей алгоритм.

Крок 1. Вихідні дані залежної змінної Y розбиваються на k груп $(r=\overline{1,k})$ згідно із зміною рівня величини Y.

Крок 2. По кожній групі даних розраховується сума квадратів відхилень:

$$S_r = \sum_{i=1}^{n_r} (y_{ir} - \overline{y_r})^2 .$$

Крок 3. Розраховується сума квадратів відхилень у цілому по всій сукупності спостережень:

$$\sum_{r=1}^{k} S_r = \sum_{i=1}^{n_r} \sum_{r=1}^{k} (y_{ir} - \bar{y}_r)^2 .$$

Крок 4. Обчислюється параметр λ:

$$\lambda = \prod_{r=1}^{k} \left(\frac{S_r}{n_r} \right)^{n_r/2} \Bigg/ \left(\frac{\sum_{r=1}^{k} S_r}{n} \right)^{n/2} ,$$

де n — загальна сукупність спостережень;

n_r — кількість спостережень r-ї групи.

Крок 5. Розраховується критерій μ:

$$\mu = -2 \ln \lambda ,$$

який наближено буде відповідати розподілу X^2 при ступенях свободи k – 1, коли дисперсія всіх спостережень однорідна. Тобто, якщо значення μ менше табличного значення X^2 при вибраному рівні довіри і ступені свободи k – 1, то явище гетероскедастичності відсутнє.

Параметричний тест Гольдфельда—Квандта

Коли сукупність спостережень невелика, то розглянутий метод 1 застосовувати неможливо.

Тоді Гольдфельд і Квандт розглянули випадок, коли $M(uu') = \sigma^2 x_{ij}^2$, тобто дисперсія залишків зростає пропорційно квадрату однієї із незалежних змінних моделі:

$$Y = XA + u \ .$$

Вони запропонували для виявлення наявності гетероскедастичності параметричний тест, в якому треба виконати наступні кроки.

Крок 1. Упорядкувати спостереження згідно з величиною елементів вектора x_j.

Крок 2. Відкинути c спостережень, які будуть знаходитись у центрі вектора. На основі експериментальних розрахунків автори вирахували оптимальні співвідношення між параметрами c і n, де n — кількість елементів вектора x_j.

$$\frac{c}{n} = \frac{4}{15} \ .$$

Крок 3. Побудувати дві економетричні моделі на основі 1МНК по двох створених сукупностях спостережень $(n_1 - c)/2$ за умови, що $(n_2 - c)/2$ перевищує кількість змінних m.

Крок 4. Знайти суму квадратів залишків за першою (1) і другою (2) моделях S_1 і S_2.

$S_1 = \hat{u}_1' \hat{u}_1$, де $\hat{u}_1$ — залишки по моделі (1) ;

$S_2 = \hat{u}_2' \hat{u}_2$, де $\hat{u}_2$ — залишки по моделі (2).

Крок 5. Розрахувати критерій R:

$$R = \frac{S_2}{S_1} \ ,$$

який при виконанні гіпотези про гомоскедастичність буде відповідати F-розподілу з $(n_1 - c - 2m)/2$, $(n_2 - c - 2m)/2$ ступенями свободи. Це означає, що розраховане значення R^* порівнюється з табличним значенням

F-критерію при ступенях свободи $(n_1 - c - 2m)\big/2$ і $(n_2 - c - 2m)\big/2$ і вибраному рівні довіри. Якщо $R \le F_{\text{табл}}$, то гетероскедастичність відсутня.

Непараметричний тест Гольдфельда—Квандта

Гольфельд і Квандт запропонували також для оцінки наявності гетероскедастичності непараметричний тест. Цей тест базується на числі піків у величині залишків після упорядкування спостережень по x_{ij}.

Тест Глейсера

Ще один тест для перевірки гетероскедастичності запропонував Глейсер: розглядати регресію абсолютних значень залишків $|u_i|$, які відповідають регресії найменших квадратів як деяку функцію від x_j, де x_j є тією незалежною змінною, яка відповідає зміні дисперсії σ_u^2. Для цього використовуються такі види функцій:

1) $|u| = a_0 + a_1 x_j$;

2) $|u| = a_0 + a_1 x_j^{-1}$;

3) $|u| = a_0 + a_1 x_j^{1/2}$ і т.п.

Рішення про відсутність гетероскедастичності залишків приймається на основі статистичної значущості коефіцієнтів a_0 й a_1. Переваги цього тесту визначаються можливістю розрізняти випадок чистої і змішаної гетероскедастичності. Чистій гетероскедастичності відповідають значення параметрів $a_0 = 0$, $a_1 \ne 0$; а змішаній — $a_0 \ne 0$, $a_1 \ne 0$. Залежно від цього треба користуватись різними матрицями S. Нагадаємо, що:

$$M(uu') = \sigma_u^2 S .$$

Якщо при економетричному моделюванні для певних вихідних даних буде виявлено явище гетероскедастичності, то оцінку параметрів моделі

треба виконувати на основі узагальненого методу найменших квадратів. Оператор оцінювання цим методом запишеться:

$$\hat{A} = (X'S^{-1}X)^{-1} X'S^{-1} Y ,$$

де

$$S^{-1}=\begin{pmatrix} \lambda_1 & 0 & 0 & 0 & \cdots & 0 \\ 0 & \lambda_2 & 0 & 0 & \cdots & 0 \\ 0 & 0 & \lambda_3 & 0 & \cdots & 0 \\ \cdots & \cdots & \cdots & \cdots & \cdots & \cdots \\ 0 & 0 & 0 & 0 & \cdots & \lambda_n \end{pmatrix}.$$

В даній матриці залежно від висунутої гіпотези:

або $\lambda_i = \dfrac{1}{x_{ij}}$;

або $\lambda_i = \dfrac{1}{x_{ij}^2}$;

або $\lambda_i = \left\{ \left| \hat{u}_i \right| \right\}$.

Прогноз на основі економетричної моделі, в якій оцінка параметрів виконана узагальненим методом найменших квадратів, можна отримати на основі такого співвідношення:

$$\hat{Y}_{пр} = X_0 A + W' V^{-1} U,$$

де u — вектор залишків, який відповідає оцінці параметрів моделі на основі 1МНК;

W' — транспонований вектор коваріацій поточних і прогнозних значень залишків;

$$V^{-1} = S^{-1}, \text{ а } V = \sigma_U^2 S \ .$$

Приклад 6.1. Нехай треба побудувати економетричну модель, яка характеризує залежність заощаджень від доходів населення. Для побудови цієї моделі використовується вихідна сукупність даних, яка включає 18 спостережень. Ці дані та розрахунки на основі їх наведені в табл. 5.1. Виходячи із сутності взаємозв’язку величини заощаджень та доходу населення, можна припустити, що дисперсія залишків не є постійною для кожного спостереження, тобто тут може існувати явище гетероскедастичності. Тому, щоб правильно вибрати метод для оцінки параметрів моделі, необхідно перевірити, чи властива гетероскедастичність для наведених вихідних даних.

Таблиця 6.1

Рік	Заощадження Y	Дохід* X		X^2	XY	$\hat{Y}$	$Y-\hat{Y}$	$\left(Y-\hat{Y}\right)^2$
1-й	1,36	13,8		190,44	18,768	1,1974	0,1626	0,0264
2-й	1,20	14,4		207,36	17,280	1,2016	–0,0016	0,0000025
3-й	1,08	15,0		225,00	16,200	1,2057	–0,1257	0,0158
4-й	1,20	15,6	n2	243,36	18,720	1,2098	–0,0098	0,000096
5-й	1,10	16,0		256,00	17,600	1,2126	–0,1126	0,0127
6-й	1,12	16,9		285,61	18,928	1,2188	–0,0988	0,0098
7-й	1,41	17,7		313,29	24,957	1,2243	0,1857	0,0345
8-й	1,50	18,5						
9-й	1,43	19,3						
10-й	1,59	20,5						
11-й	1,90	21,7						
12-й	1,95	22,7		515,29	44,265	1,8401	0,1099	0,0121
13-й	1,82	23,6		556,96	42,925	1,9885	–0,1685	0,0284
14-й	2,04	24,7		610,09	50,388	2,1699	–0,1299	0,0169
15-й	2,53	26,1	n1	681,21	66,033	2,4008	0,1292	0,0167
16-й	2,94	27,8		772,84	81,732	2,6811	0,2589	0,0670
17-й	2,75	28,9		835,21	79,475	2,8625	–0,1125	0,0127
18-й	2,99	30,2		912,04	90,298	3,0769	–0,0869	0,0076
$\sum$	31,91	373,4						

Таблиця 6.2

Рік	$\lambda_i = 1/X_i$	$\hat{Y}$	$Y-\hat{Y}$	$\left(Y-\hat{Y}\right)^2$	$Y-\overline{Y}$	$\left(Y-\overline{Y}\right)^2$
1-й	0,0725	0,9865	0,3735	0,1395	–0,4128	0,1704
2-й	0,0694	1,0542	0,1458	0,2126	–0,5728	0,3281
3-й	0,0667	1,1219	–0,0419	0,0176	–0,6928	0,4799
4-й	0,0641	1,1896	0,0104	0,0001	–0,5728	0,3281
5-й	0,0625	1,2347	–0,1347	0,0181	–0,6728	0,4527
6-й	0,0592	1,3362	–0,2162	0,0467	–0,6528	0,4261
7-й	0,0565	1,4265	–0,0165	0,0003	–0,3628	0,1316
8-й	0,0541	1,5167	–0,0167	0,0003	–0,2728	0,0744
9-й	0,0518	1,6069	–0,1769	0,0313	–0,3428	0,1175
10-й	0,0488	1,7423	–0,1523	0,0232	–0,1828	0,0334
11-й	0,0461	1,8777	0,0223	0,0005	0,1272	0,0162
12-й	0,0441	1,9905	–0,0405	0,0016	0,1772	0,0314
13-й	0,0424	2,0919	0,2719	0,0739	0,0472	0,0022
14-й	0,0405	2,2161	–0,1761	0,0310	0,2672	0,0714
15-й	0,0383	2,3739	0,1561	0,0244	0,7572	0,5734
16-й	0,0359	2,5657	0,3743	0,1401	1,1672	1,3624
17-й	0,0346	2,6898	0,0602	0,0036	0,9772	0,9549
18-й	0,0331	2,8365	0,1535	0,0236	1,2172	1,4816
$\sum$	х	х	х	0,7884	х	7,0357

В таблиці дані впорядковані за величиною доходу, починаючи від меншого до більшого значення.

Розв'язання

1. Ідентифікація змінних:

$$Y = f(X, u) \ ,$$

Y — залежна змінна (заощадження);

X — незалежна змінна (дохід);

u — стохастична складова.

2. Специфікація моделі:

$$Y = a_0 + a_1 X + u \ ,$$

$$\hat{Y} = \hat{a}_0 + \hat{a}_1 X \ ,$$

$$u = Y - \hat{Y} \ .$$

3. Визначимо наявність гетероскедастичності. Для цього застосуємо алгоритм Гольдфельда—Квандта. Дану сукупність спостережень

впорядкуємо по X від меншого до більшого значення. Відшукуємо C спостережень, які знаходяться в середині сукупності:

$$\frac{C}{n} = \frac{4}{15} \ , \quad n = 18 \ , \quad \frac{C}{18} = \frac{4}{15} \ , \quad C = \frac{4 \cdot 18}{15} \approx 4 \ .$$

Тоді $n_1 = n_2 = 7$.

3.1. Розрахуємо економетричну модель для сукупності $n_1 = 7$.

Оцінимо кількісно параметри моделі на основі МНК.

$$\begin{cases} n\hat{a}_0 + \hat{a}_1 \sum x = \sum y \\ \hat{a}_0 \sum x + \hat{a}_1 \sum x^2 = \sum xy \end{cases}$$

$$\sum x = 109{,}4$$

$$\sum y = 8{,}47$$

$$\sum x^2 = 1721{,}06$$

$$\sum xy = 132{,}453$$

$$\left\{ \begin{array}{l|l} 7\hat{a}_0 + 109{,}4\hat{a}_1 = 8{,}47 & *109{,}4 \\ 109{,}4\hat{a}_0 + 1721{,}06\hat{a}_1 = 132{,}453 & *7 \end{array} \right.$$

$$\begin{cases} 765{,}8\hat{a}_0 + 11968{,}36\hat{a}_1 = 926{,}618 \\ 765{,}8\hat{a}_0 + 12047{,}42\hat{a}_1 = 927{,}171 \end{cases}$$

$$79{,}06\hat{a}_1 = 0{,}553$$
$$\hat{a}_1 = 0{,}0069$$

$$7\hat{a}_0 + 109{,}4*0{,}0069 = 8{,}47$$
$$7\hat{a}_0 = 8{,}47 - 0{,}7549$$
$$7\hat{a}_0 = 7{,}7151$$
$$\hat{a}_0 = 1{,}1022$$

$\hat{Y}_1 = 1{,}1022 + 0{,}0069X$ — перша економетрична модель.

На основі моделі можна зробити висновок: і якщо дохід виросте на 1, то заощадження збільшаться на 0,0069 одиниці.

3.2 Розрахуємо економетричну модель для сукупності $n_2 = 7$.

$\hat{Y}_2 = -1{,}9031 + 0{,}1649X$ — друга економетрична модель.

На основі моделі можна зробити висновок: і якщо дохід виросте на 1, то заощадження збільшаться на 0,1649 одиниці для даної сукупності спостережень.

3.3. Для кожної моделі знайдемо суму квадратів залишків:

$$S_1 = u_1'u_1 = \sum(Y_1 - \hat{Y}_1)^2 ;$$

$$S_2 = u_2'u_2 = \sum(Y_2 - \hat{Y}_2)^2 ;$$

$$S_1 = 0{,}0993$$
$$S_2 = 0{,}1614$$

3.4. Знаходимо критерій R:

$$R = \frac{S_2}{S_1};\ R = \frac{0{,}1614}{0{,}0993} = 1{,}625 .$$

Порівняємо цей критерій із табличним значенням критерію Фішера при ступенях свободи $\frac{n-c-2m}{2} = 5$ і рівні довіри α = 0,05 $F_{табл}$ = 5,05. Гетероскедастичність відсутня, тому що $R < F_{табл}$.

Приклад 6.2. Необхідно оцінити параметри економетричної моделі, яка характеризує залежність витрат на харчування від загальних затрат на основі даних, що наведені в табл. 5.2.

Таблиця 6.2

Номер спостереження	Витрати на харчування	Загальні затрати	$\hat{Y}$	u	u^2
1	*2*	*3*	*4*	*5*	*6*
1	2,30	15	2,16	0,14	0,020
2	2,20	15	2,16	0,04	0,002
3	2,08	16	2,20	–0,12	0,015
4	2,20	17	2,25	–0,05	0,002
5	2,10	17	2,25	–0,15	0,022
6	2,32	18	2,29	0,26	0,0007
7	2,45	19	2,34	0,11	0,012
8	2,50	20			
9	2,20	20			
10	2,50	22			
11	3,10	64			
12	2,40	68	2,37	0,13	0,016
13	2,82	72	2,52	0,29	0,085
14	3,04	80	2,68	0,36	0,128
15	2,70	85	2,99	–0,29	0,084
16	3,91	90	3,18	0,76	0,573
17	3,10	95	3,38	–0,28	0,076
18	3,99	100	3,57	0,42	0,178

Виходячи з особливостей вихідної інформації, можна припустити, що порушується гіпотеза про незмінність дисперсії.

Розв’язання

1. Ідентифікуємо змінні моделі:

Y — витрати на харчування, залежна змінна;

X — загальні затрати, незалежна змінна;

$$Y = f(X, u) .$$

2. Перевіримо наявність гетероскедастичності для наведених вихідних даних на основі параметричного тесту Гольдфельда—Квандта.

2.1. Впорядкуємо значення незалежної змінної X від меншого до більшого і відкинемо C значень, які знаходяться всередині впорядкованого ряду:

$$\frac{C}{18} = \frac{4}{15}\ ; \qquad C \approx 4\ .$$

2.2. На основі отриманих двох сукупностей спостережень (від першого до сьомого включно і від одинадцятого до вісімнадцятого значення) побудуємо дві економетричні моделі за методом 1 МНК.

1-ша модель: $\hat{Y}_1 = 1{,}475 + 0{,}046\,X$;

2-га модель: $\hat{Y}_2 = -0{,}093 + 0{,}039X$.

2.3. Визначимо залишки по цих двох моделях:

$$u_1 = Y_1 - \hat{Y}_1\ ;$$

$$u_2 = Y_2 - \hat{Y}_2\ .$$

Залишки та квадрати залишків наведені в табл. 5.2.

2.4. Розрахуємо залишкові дисперсії та знайдемо їх співвідношення R:

$$R = \frac{\sigma^2_{u_2}}{\sigma^2_{u_1}} = \frac{1{,}14/7}{0{,}074/7} = 15{,}41\ .$$

2.5. Порівняємо критерій R з критичним значенням F- критерію при $\gamma_1 = 5$ і $\gamma_2 = 5$ ступенях свободи і рівні довіри α = 0,01 F = 11. Оскільки $R > F_{\text{крит}}$, вихідні дані мають гетероскедастичність.

3. При наявності гетероскедастичності оцінку параметрів моделі виконаємо методом Ейткена:

$$\hat{A} = (X'S^{-1}X)^{-1}X'S^{-1}Y\ .$$

3.1. Запишемо матриці змінних, які входять в оператор Ейткена:

$$X' = \begin{pmatrix} 1 & 1 & 1 & 1 & 1 & 1 & 1 & 1 & 1 & 1 & 1 & 1 & 1 & 1 \\ 15 & 15 & 16 & 17 & 17 & 18 & 19 & 68 & 72 & 80 & 85 & 90 & 95 & 100 \end{pmatrix}.$$

3.2. Визначимо добутки матриць:

$$X'S^{-1} = \begin{pmatrix} 0{,}0667 & 0{,}0667 & 0{,}0625 & 0{,}0589 & 0{,}0589 & 0{,}0555 & 0{,}0526 & 0{,}05 & 0{,}05 & 0{,}0154 & 0{,}0156 & 0{,}0148 & 0{,}0139 & 0{,}0125 \\ 1 & 1 & 1 & 1 & 1 & 1 & 1 & 1 & 1 & 1 & 1 & 1 & 1 & 1 \end{pmatrix}$$

$$X'S^{-1}X = \begin{pmatrix} 0{,}6672 & 18 \\ 18 & 833 \end{pmatrix};$$

3.3. Знайдемо обернену матрицю:

$$(X'S^{-1}X)^{-1} = \begin{pmatrix} 3{,}5934 & -0{,}0776 \\ -0{,}0776 & 0{,}0029 \end{pmatrix};$$

і вектор:

$$X'S^{-1}Y = \begin{pmatrix} 1{,}5998 \\ 48{,}04 \end{pmatrix}.$$

3.4. Обчислимо вектор оцінок параметрів моделі:

$$\hat{A} = \begin{pmatrix} 3,5934 & -0,0776 \\ -0,0776 & 0,0029 \end{pmatrix} * \begin{pmatrix} 1,5998 \\ 48,04 \end{pmatrix} = \begin{pmatrix} 2,0187 \\ 0,0141 \end{pmatrix}.$$

Звідси $\hat{a}_0 = 2,0187$;

$$\hat{a}_1 = 0,0141 .$$

Економетрична модель витрат на харчування запишеться так:

$$\hat{Y} = 2,0187 + 0,0141X .$$

4. Економічний аналіз характеристик економетричної моделі.

4.1. Коефіцієнт детермінації $R^2 = 0,722$. Це означає, що на 72,2% варіація витрат на харчування залежить від варіації загальних затрат.

4.2. Коефіцієнт кореляції: $R = \sqrt{R^2} = 0,85$ свідчить про досить тісний зв'язок витрат на харчування і загальних затрат.

4.3. Залишкова дисперсія $\sigma_U^2 = 0,083$ показує, що розрахункові значення витрат на харчування дуже близькі до фактичних.

4.4. Параметр моделі $\hat{a}_1 = 0,0141$ свідчить про те, що збільшення загальних затрат на одиницю сприятиме граничному зростанню витрат на харчування на 0,014 одиниць.

5. Розрахуємо матрицю коваріацій оцінок параметрів моделі:

$$\operatorname{var}(\hat{A}) = \sigma_U^2 (X' S^{-1} X)^{-1} = 0,083 \begin{pmatrix} 3,5934 & -0,0776 \\ -0,0776 & 0,0029 \end{pmatrix} = \begin{pmatrix} 0,298 & -0,0064 \\ -0,0064 & 0,0002 \end{pmatrix}$$

Діагональні елементи цієї матриці є дисперсіями оцінок параметрів моделі, інші елементи характеризують коваріацію між оцінками.

6. Визначимо стандартні помилки оцінок параметрів і знайдемо їх довірчі інтервали

$$S_{\hat{a}_j} = \sqrt{\operatorname{var}(\hat{A})} = \sigma_U \sqrt{c_{jj}};$$
$$S_{\hat{a}_0} = \sqrt{0,298} = 0,546 ;$$
$$S_{\hat{a}_1} = \sqrt{0,0002} = 0,014 .$$

Для побудови довірчих інтервалів оцінок параметрів моделі знайдемо t-критерій при ступенях свободи $n - m = 12$ і рівні довіри $\alpha = 0,05$: $t_{\text{крит}} = 2,179$.

Довірчі інтервали оцінок:

$$\hat{a}_0 - t_{0,05} S_{\hat{a}_0} \le \hat{a}_0 \le \hat{a}_0 + t_{0,05} S_{\hat{a}_0}$$

$$2{,}0187 - 1{,}102 \le a_0 \le 2{,}0187 + 1{,}102$$

$$0{,}917 \le a_0 \le 3{,}121$$

$$\hat{a}_1 - t_{0,05} S_{\hat{a}_1} \le \hat{a}_1 \le \hat{a}_1 + t_{0,05} S_{\hat{a}_1}$$

$$0{,}016 \le a_1 \le 0{,}045$$

Рівень стандартних помилок та довірчі інтервали оцінок параметрів моделі свідчать про те, що отримані оцінки є неефективними та зміщеними.

Завдання для самостійної та лабораторної роботи

Завдання 6.1. Для побудови економетричної моделі, що характеризує залежність між затратами на реалізацію продукції, обсягом товарообігу та середнім рівнем товарних запасів необхідно перевірити гіпотезу про наявність гетероскедастичності для вихідних даних, які наведені в табл. 6.1 — 6.2.

Таблиця 6.1

№ складу	Затрати на реалізацію продукції, млн. грн.	Обсяг товарообігу, млн.т.	Середній рівень товарних запасів, млн. т
1	300	25	5
2	280	20	4
3	350	30	6
4	340	30	7
5	330	28	7
6	320	28	5
7	310	25	6
8	300	24	4
9	320	27	5
10	280	22	4
11	340	35	6
12	360	30	7
13	320	29	7
14	300	28	5
15	310	25	6
16	350	26	4

Таблиця 6.2

№ складу	Затрати на реалізацію продукції, млн. грн.	Обсяг товарообігу, млн.т	Середній рівень товарних запасів, млн.т
1	350	26	5
2	280	22	4
3	350	30	6
4	340	30	7
5	300	29	7
6	320	28	5
7	320	25	6
8	280	24	4
9	300	23	6
10	380	21	5
11	340	30	6
12	360	32	7
13	330	28	8
14	320	29	5
15	340	25	6
16	300	24	9

Список літератури

Основна література:

1.Економетрія / В. В. Здрок, Т. Я. Лагоцький [+компакт диск]. – К. : Знання, 2010. – 118 с.
2. Економетрія: [навч. посіб.] / за ред.. О. А. Корольова. – К. : Книга, 2005. – 164 с.
3. Кремер Н. Ш. Эконометрика: [Учебник для вузов.] / Н. Ш. Кремер, Б. А. Путко. – М.: ЮНИТИ-ДАНА, 2005. – 311 с.
4. Кузьмичов А. І. Економетрія. Моделювання засобами MS Excel: [навчальний посібник] / А. І. Кузьмичов. – К. : ЦУЛ, 2011. – 214 с.
5. Лещинський О. Л. Економетрія / О. Л. Ліщинський. – К. : МАУП, 2003. – 208 с.
6. Лугінін В. М. Економетрія: [навч. посіб.] / В. М. Лугінін. – К. : ЦНЛ, 2008. – 312 с.
7. Наконечний С. І. Економетрія. / С. І. Наконечний, Т. О. Терещенко. – К. : КНЕУ, 2006. – 528 с
8. Скоков Б. Г. Конспект лекцій до курсу «Економетрія" / Б. Г. Скоков, К. А. Мамонов. – Харків : ХНАМГ, 2006 –105 с.
9. Толбатов Ю. А. Економетрика: [Підручник для студентів] / Ю. А. Толбатов. – Тернопіль: Підручники і посібники, 2008. – 288 с.

Додаткова література:

1. Бредюк В. І. Курс лекцій з дисципліни „Економетрія" / В. І. Бредюк. – Рівне: НУВГП, 2006- 154с.
2. Винн Р. Введение в прикладной економетрический анализ. / Р. Винн, К. Холден. – М. : Финансы и статистика, 1981. – 268 с.
3. Джонстон Дж. Економетрические методы. / Дж. Джонстон. – М. : Статистика, 1980. – 312 с.
4. Доугерти К. Введение в економетрику. / К. Доугерти. – М. : Статистика, 1997. – 402 с.

5. Елисеева И. И. Практикум по эконометрике. / И. И. Елисеева. – М. : Финансы и Статистика, 2002. – 192 с.
6. Елисеева И. И. Эконометрика. / И. И. Елисеева. – М.: Финансы и Статистика, – 2004. – 344 с.
7. Лизер С. Эконометрические методы и задачи. / С. Лизер. – М. : Дело, 1997. – 248 с.
8. Магнус Я. Р. Эконометрика. Начальный курс: [учеб.] / Я. Р. Магнус, П. К. Катышев, А. А. Пересецкий. – М. : Дело, 2005.- 576 с.
9. Наконечний С. І. Математичне програмування: [навч. посіб.] / С. І. Наконечний, С. С. Савіна. – К.: КНЕУ, 2003. – 198 с.
10. Наконечний С. І., Практикум з економетрії: [Навч. посібник] / С. І. Наконечний, Т. О. Терещенко, Н. К. Водзянова, О. С. Роскач. – К.: КНЕУ, 1998. —176с.
11. Практичні заняття з економетрії в EXCEL: [навч. посіб.] / О. О. Кубайчук, С. А. Теренчук. – К. : Вид-во Європейського ун-ту, 2007. – 212 с.

Навчально-методичне видання

Автор: Бегун Світлана Іванівна

ЕКОНОМЕТРИКА

Методичні вказівки

Друкується в авторській редакції