

## ТЕМА 6

### Емпіричні методи кількісного аналізу на основі статистичних рівнянь

Виявлення основних закономірностей розвитку економіки неможливе без дослідження взаємозв'язку між економічними процесами та явищами. Останні підлягають моделюванню, а на підставі моделей можливо з'ясувати із певною імовірністю, чи відбулися якісні та кількісні зміни в процесах, в їх взаємовпливі один на одного тощо. Тему 6 також присвячено вивченню причинно-наслідкового зв'язку між процесами та методів, якими його можна дослідити.

#### Основні питання, що розглядаються:

1. Специфіка статистичної оцінки взаємозв'язку двох часових рядів
2. Коінтеграція та хибна (уявна) регресія. Тест Гренджера
3. Методи виключення тенденції
4. Бінарні змінні. Порівняння двох регресій

#### Основні терміни:

Хибна кореляція; автокореляція в залишках; коінтеграція часових рядів; хибна (уявна) регресія; тест Гренджера; метод послідовних різниць; метод відхилень від трендів; метод включення в регресію фактору часу; дихотомічні (бінарні, dummy) змінні; Чоу-тест.

#### 6.1. СПЕЦИФІКА СТАТИСТИЧНОЇ ОЦІНКИ ВЗАЄМОЗВ'ЯЗКУ ДВОХ ЧАСОВИХ РЯДІВ

Вивчення причинно-наслідкових зв'язків між змінними, представлених у формі часових рядів, є однією з найбільш складних задач еконо-метричного моделювання. Застосування в цих цілях традиційних методів кореляційно-регресійного аналізу, може призвести до ряду серйозних проблем, які в першу чергу пов'язані із специфікою часових рядів як джерела даних в економетричному моделюванні. В попередніх темах було розглянуто, що динамічний ряд містить три основних компоненти: тенденцію, циклічні, або сезонні, коливання і випадкову компоненту. Розглянемо більш детально, яким чином наявність цих компонент відображається на результатах кореляційно-регресійного аналізу часових рядів даних.

Попередній етап такого аналізу зосереджується на виявленні сезонних, циклічних коливань, тому зусилля при моделюванні зводяться до усунення впливу зазначених коливань, оскільки їх наявність призведе до завищення реальних показників сили та тісноти зв'язку часових рядів, що вивчаються у випадку, якщо обидва ряди містять циклічні коливання однакової періодичності, або до заниження цих показників у випадку, коли сезонні, або циклічні коливання містять тільки один з рядів або періодичність коливань у часових рядах різна.

Усунення сезонної компоненти можна проводити, виходячи з запропонованою в попередніх темах методикою.

При вивченні двох рядів  $x$  та  $y$  виходимо з того, що сезонні коливання відсутні. Проте, якщо в цьому випадку ми отримаємо високий коефіцієнт кореляції, з цього неможливо буде визначити, який ряд є причиною, а який наслідком (наприклад, інфляція – зарплата).

Для того, щоб отримати коефіцієнти кореляції, які характеризують причинно-наслідковий зв'язок між рядами, слід позбавитися так званої **хибної кореляції**, що викликається наявністю тенденції в кожному ряді. Зазвичай це здійснюється за допомогою одного з методів виключення тенденції.

Припустимо, що по двох часових рядах  $x_t$  та  $y_t$  будується рівняння парної регресії виду:

$$y_t = a + b \cdot x_t + \varepsilon_t \quad (6.1)$$

Наявність тенденції в кожному з цих часових рядів означає, що на залежну  $y_t$  і незалежну  $x_t$  змінні моделі впливає фактор часу, який безпосередньо в моделі не врахований.

Вплив фактору часу буде виражено в кореляційній залежності між значеннями залишків  $\varepsilon_t$  за поточний і попередній моменти часу, яка отримала назву **автокореляція в залишках**.

Автокореляція в залишках – це порушення одного з основних припущень МНК – припущення про випадковість залишків, отриманих по рівнянню регресії. Один з можливих шляхів усунення цієї проблеми полягає в застосуванні узагальненого МНК до оцінки параметрів моделі, застосування якого ми показували в попередніх темах. Крім того, може бути наявна мультиколінеарність у тому випадку, коли фактори, що пояснюють залежну змінну, містять тенденцію.

#### 6.2. КОІНТЕГРАЦІЯ ТА ХИБНА (УЯВНА) РЕГРЕСІЯ.

##### ТЕСТ ГРЕНДЖЕРА

Розглянемо два часових ряди  $y_t$  та  $x_t$ . Припустимо, що обидва ряди мають одиничні корені, тобто є нестационарними. Припустимо далі, що дослідник не знає механізмів, що породжують  $y_t$  та  $x_t$ , і оцінює регресію:

$$y_t = \beta x_t + \varepsilon_t, \quad t=1, \dots, n.$$

Якщо  $\varepsilon_t = y_t - \beta x_t$ ,  $t=1, \dots, n$  є стаціонарним часовим рядом, то часові ряди  $y_t$  та  $x_t$  називаються коінтегрованими, а вектор  $(1 - \beta)$  називається коінтегруючим вектором.

Прикладами коінтегрованих рядів є:

1. Довга ставка відсотку  $R$ , коротка ставка відсотку  $r$ :  $\varepsilon_t = R_t - r_t$ , вектор коінтеграції (1-1).
2. Логарифм споживання  $C_t$ , логарифм доходу  $y_t$ :  $\varepsilon_t = C_t - y_t$ , вектор коінтеграції (1-1).
3. Логарифм обмінного курсу  $D_t$ , логарифм внутрішньої ціни  $P_t$ , логарифм цін світового ринку  $P_t^*$ :  $\varepsilon_t = D_t - P_t + P_t^*$ , вектор коінтеграції (1 - 1 1).

У випадку коінтегрованості часових рядів говорять про *довгострокову динамічну рівновагу*. Якщо  $y_t$  та  $x_t$  коінтегровані, то  $y_t$  та  $\beta x_t$  містять спільну нестационарну компоненту – довготривалу тенденцію, а різниця  $y_t - \beta x_t$  стаціонарна і здійснює коливання біля нуля.

Таким чином, *коінтеграція часових рядів* – причинно-наслідкова залежність у рівнях часових рядів, яка виражається у співпадінні або протилежній спрямованості їх тенденцій і випадковими коливаннями.

Можливий випадок, коли похибка  $\varepsilon_t = y_t - \beta x_t$ ,  $t = 1, \dots, n$  у регресії є нестационарним часовим рядом. Тоді умови класичної регресійної моделі не виконуються, зокрема дисперсія  $\varepsilon_t$  не є постійною. Крім того, МНК оцінка параметра  $\beta$  не ефективна, тому з ростом обсягу вибірки збільшуються шанси отримання хибних висновків про взаємозв'язок  $y_t$  та  $x_t$ . Така ситуація називається *хибною (уявною) регресією*. На практиці ознаками хибної регресії є високе значення  $R^2$  і мале значення статистики Дарбіна-Уотсона.

Для перевірки рядів на коінтеграцію використовуються тести Енгеля-Гренджера, Йохансена, на причинність – тест Гренджера.

*Тест Енгеля-Гренджера* проводиться за таким алгоритмом:

1. Висувається нульова гіпотеза про відсутність коінтеграції між рядами  $y_t$  та  $x_t$ .
2. Розраховуються параметри рівняння регресії виду

$$\Delta \varepsilon_t = a + b \cdot \varepsilon_{t-1}, \quad (6.2)$$

де  $\Delta \varepsilon_t$  – перші різниці залишків, отримані зі співвідношення  $\varepsilon_t = y_t - a - b \cdot x_t$ .

3. Визначають фактичне значення  $t$ -критерію в рівнянні (6.2) для коефіцієнта  $a$ .
4. Порівнюють отримане значення з критичним значенням статистики  $t$ , обчисленим Інглом та Гренджером для рівня значущості 1; 5 і 10 %. Критичні значення статистики  $t$  складають відповідно 2,5899; 1,9439 та 1,6177. Якщо фактичне значення  $t$  більше критичного значення  $t$  для заданого рівня значущості  $\alpha$ , нульову гіпотезу про відсутність коінтеграції досліджуваних часових рядів відхиляють і з імовірністю  $(1-\alpha)$  приймають альтернативну гіпотезу про те, що між рядами  $y_t$  та  $x_t$  є коінтеграція. В іншому випадку гіпотеза про відсутність коінтеграції між досліджуваними рядами не відхиляється.

*Тест причинності Гренджера* припускає, що інформація, потрібна для прогнозування відповідних змінних, повністю міститься у часових рядах. Тест включає оцінювання двох регресій:

$$y_t = \sum_{i=1}^n \alpha_i x_{t-i} + \sum_{j=1}^n \beta_j y_{t-j} + \varepsilon_{1t} \quad (6.3)$$

$$x_t = \sum_{i=1}^m \lambda_i x_{t-i} + \sum_{j=1}^m \delta_j y_{t-j} + \varepsilon_{2t}, \quad (6.4)$$

де припускається, що випадкові величини  $\varepsilon_{1t}$  і  $\varepsilon_{2t}$  некорельовані між собою.

Перше рівняння показує, що поточне значення  $y_t$  залежать як від попередніх значень самого  $y_t$ , так і від  $x_t$ , а друге рівняння визначає подібну поведінку для  $x_t$ . Розглянемо чотири випадки.

1. *Однобічна причинність від  $x$  до  $y$  наявна*, якщо оцінені коефіцієнти при  $x$  в (6.3) статистично відрізняються від нуля як група (тобто  $\sum \alpha_i \neq 0$ ), а множина оцінених коефіцієнтів при  $y$  в (6.4) статистично не відрізняються від нуля ( $\sum \delta_i = 0$ ).

2. *Навпаки, однобічна причинність від  $y$  до  $x$  наявна*, коли множина коефіцієнтів при  $x$  в (6.3) статистично не відрізняється від нуля (тобто  $\sum \alpha_i = 0$ ), а множина коефіцієнтів при  $y$  в (6.4) статистично відрізняються від нуля (тобто  $\sum \delta_i \neq 0$ ).

3. *Зворотний зв'язок, або двобічна причинність існує*, коли множини коефіцієнтів при  $x$  та  $y$  статистично відрізняються від нуля в обох регресіях.

4. *Нарешті, між  $x$  та  $y$  причинного зв'язку немає*, тобто вони незалежні, коли множини коефіцієнтів при  $x$  і  $y$  статистично не відрізняються від нуля в обох регресіях.

Для оцінки статистичної значущості коефіцієнтів в тесті Гренджера застосовується тест Фішера. Зокрема, розраховується  $F$ -критерій з параметрами  $m$ ,  $(n-2m)$ , де  $m$  – кількість затриманих значень (тобто лагових значень кожної змінної),  $n$  – кількість спостережень.

При вивченні двох взаємопов'язаних часових рядів на попередній стадії регресійного аналізу рекомендується усунути сезонні або циклічні коливання, якщо вони наявні в досліджуваних часових рядах, у відповідності з прийнятою адитивною чи мультиплікативною моделями рядів.

Якщо часові ряди  $y_t$  та  $x_t$ , що розглядаються, містять тенденцію, то коефіцієнт кореляції, що характеризує ступінь залежності між  $y_t$  та  $x_t$  буде мати велике значення. Така ж ситуація буде мати місце тоді, коли  $y_t$  та  $x_t$  залежать від змінної часу  $t$ . Як у першому, так і в другому випадку має місце хибна кореляція, яка призводить при побудові регресії  $y_t$  та  $x_t$  виду (6.1) до автокореляції в залишках і нестационарності ряду залишків регресії (хибна регресія), тобто до порушення припущень МНК.

Для отримання регресії із стаціонарним часовим рядом залишків  $\varepsilon_t$  може бути застосовано метод послідовних різниць, метод відхилення від тренду та метод включення фактору часу.

### 6.3. МЕТОДИ ВИКЛЮЧЕННЯ ТЕНДЕНЦІЇ

Сутність усіх методів виключення тенденції полягає в тому, щоб усунути або зафіксувати вплив фактору часу на формування рівнів ряду. Основні методи усунення тенденцій можна поділити на 2 групи:

1) методи, засновані на перетворенні рівнів вихідного ряду в нові змінні, що не містять тенденції. Отримані змінні використовуються для аналізу взаємозв'язків часових рядів. Ці методи передбачають безпосереднє усунення трендової компоненти  $T$  з кожного рівня часового ряду. Два основних методи в цій групі – **метод послідовних різниць** і **метод відхилень від трендів**;

2) методи, засновані на вивченні взаємозв'язку вихідних рівнів часових рядів при елімінаванні впливу фактору часу на залежну і незалежну змінні моделі. В першу чергу – це **метод включення в модель регресії за часовими рядами фактору часу**.

*Метод відхилення від тренду*

Нехай є два часових ряди  $x_t$  та  $y_t$ , кожен з яких містить трендову компоненту  $T$  і випадкову компоненту  $\varepsilon$ . Аналітичне вирівнювання кожного з цих рядів дозволяє знайти параметри відповідних рівнянь трендів і визначити розрахункові за трендом рівні  $\xi_t$  та  $\zeta_t$  відповідно. Ці розрахункові значення можна прийняти за оцінку трендової компоненти  $T$  кожного ряду. Тому вплив тенденції можна усунути шляхом вирахування розрахованих значень рівнів ряду з фактичних. Цю процедуру здійснюють для кожного часового ряду в моделі. Отримані відхилення піддають аналізу на тісноту зв'язку (автокореляцію) і якщо така не виявлена, то прогнознi значення коригують на відхилення.

Подальший аналіз взаємозв'язку рядів проводять з використанням не вихідних рівнів, а відхилень від тренду  $x_t - \xi_t$  та  $y_t - \zeta_t$  за умови, що останнє не містить тенденції.

*Метод послідовних різниць*

У ряді випадків замість аналітичного вирівнювання часового ряду з метою усунення тенденцій можна застосувати більш простий метод – метод послідовних різниць.

Якщо часовий ряд містить яскраво виражену лінійну тенденцію, її можна усунути шляхом заміни вихідних рівнів ряду ланцюговими абсолютними приростами (першими різницями).

Нехай

$$y_t = \xi_t + \varepsilon_t$$

$$\xi_t = a + b \cdot t.$$

Тоді

$$\Delta_t = y_t - y_{t-1} = a + b \cdot t + \varepsilon_t - (a + b \cdot (t-1) + \varepsilon_{t-1}) = b + (\varepsilon_t - \varepsilon_{t-1}).$$

Коефіцієнт  $b$  – константа, яка не залежить від часу. При наявності сильної лінійної тенденції залишки  $\varepsilon_t$  достатньо малі і у відповідності з припущеннями МНК носять випадковий характер. Тому перші різниці рівнів ряду  $\Delta_t$  не залежать від змінної часу, їх можна використовувати для подальшого аналізу.

Якщо часовий ряд містить тенденцію у формі параболи другого порядку, то для її усунення можна замінити вихідні рівні ряду на другі різниці.

Нехай має місце співвідношення  $y_t = \xi_t + \varepsilon_t$ , однак

$$\xi_t = a + b_1 \cdot t + b_2 \cdot t^2.$$

Тоді:

$$\Delta_t = y_t - y_{t-1} = a + b_1 \cdot t + b_2 \cdot t^2 + \varepsilon_t - (a + b_1 \cdot (t-1) + b_2 \cdot (t-1)^2 + \varepsilon_{t-1}) =$$

$$= b_1 - b_2 + 2 \cdot b_2 \cdot t + (\varepsilon_t - \varepsilon_{t-1}).$$

Як показує це співвідношення, перші різниці  $\Delta_t$  безпосередньо залежать від фактору часу і містять тенденцію.

Визначимо другі різниці:

$$\Delta_t'' = \Delta_t - \Delta_{t-1} = 2 \cdot b_2 + (\varepsilon_t - 2 \cdot \varepsilon_{t-1} + \varepsilon_{t-2}).$$

Очевидно, що другі різниці  $\Delta_t''$  не містять тенденції, тому при наявності у вихідних рівнях тренду у формі параболи другого порядку їх можна використовувати для подальшого аналізу. Якщо тенденції часового ряду відповідає експоненційний, або степеневий, тренд, метод послідовних різниць слід застосовувати не до вихідних рівнів ряду, а до їх логарифмів.

*Включення в модель регресії фактору часу*

У кореляційно-регресійному аналізі можна усунути дію будь-якого фактору, якщо зафіксувати дію цього фактору на результат та інші включені в модель фактори. Даний прийом широко застосовується в аналізі часових рядів, коли тенденція фіксується через включення фактору часу в модель в якості незалежної змінної.

Модель виду:

$$y_t = a + b_1 \cdot x_t + b_2 \cdot t + \varepsilon_t$$

відноситься до групи моделей, що включають фактор часу. Очевидно, що число незалежних змінних у такій моделі може бути більше одиниці. Крім того, це можуть бути не тільки поточні, але й лагові значення незалежної і результативної змінних.

Перевага даної моделі перед методами відхилення від трендів і послідовних різниць полягає в тому, що вона дозволяє врахувати всю інформацію, що міститься у вихідних даних, оскільки значення у та  $x$  – це рівні вихідних часових рядів.

Крім того, модель будується по всій сукупності даних за період, що розглядається на відміну від методу послідовних різниць, який призводить до втрати числа спостережень. Параметри  $a$  та  $b$  моделі з включенням фактору часу визначається звичайним МНК.

#### 6.4. БІНАРНІ ЗМІННІ. ПОРІВНЯННЯ ДВОХ РЕГРЕСІЙ

Економічні явища є багатоманітними, не завжди кількісно вимірюваними. Часто вони приймають вид бінарних значень: 0 та 1. Такі змінні називаються *dummy (дихотомічні)* або *фіктивні* змінні. До них можуть бути віднесені: стать, релігія, якість продукції, змінні в економічній політиці тощо.

Пара  $(0,1)$  може легко трансформуватися у будь-яку іншу з лінійними перетворенням  $y = a + bz$  ( $b \neq 0$ ), де  $a$  та  $b$  – константи,  $z = 1$  або  $0$ . Наприклад, коли  $z = 1$ ,  $y = a + b$ , а коли  $z = 0$ ,  $y = a$ .

Dummy-змінні можуть використовуватися як окремі фактори в моделях або формувати цілі моделі з dummy-змінними. Такі моделі називаються AOV-моделями (analysis – of – variance).

Більш поширеним є випадок змішаних факторів ASCOV (analysis of covariance).

Які особливості додає до моделі включення dummy-змінної?

1. Dummy-змінні відокремлюють різні класи або різні категорії. Вони можуть ускладнювати явище, тобто поділяти на більшу кількість категорій.

**Приклад:**  $d$  – dummy-змінна в певній моделі  $y = a + b_1x_1 + b_2d + e$ , де  $d = [0; 1]$ , наприклад, студент – відмінник або невідмінник.

Можна ускладнити модель:  $y = a + b_1d_1 + b_2d_2 + b_3x_1 + e$ ,

де  $d_1 = 1$  – відмінник;

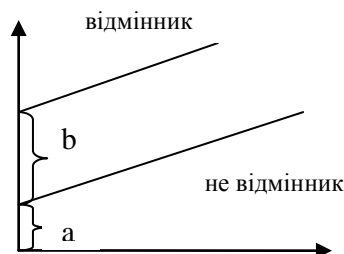
= 0 – усі інші випадки;

$d_2 = 0$  – не відмінник;

= 1 – усі інші випадки.

2. Категорія 0 – є контрольною, вихідною категорією. Всі порівняння робляться для неї.

3. Коефіцієнт біля dummy-змінної називається диференціальним коефіцієнтом перетину, тому що він показує, наскільки значення перетину першої категорії відрізняється від значення перетину базової категорії.



4. Якісні змінні впливають лише на перетин, а не на нахил. А якщо нахили також відрізняються? Тоді важливим постає визначення відмінностей у двох і більше регресіях. Відмінність може бути у перетинах, нахилах або в обох випадках.

**Приклад.**

Досліджується залежність розміру заощаджень від доходу у Великобританії в 1946-1963 роках.

Період 1	Заощадження	Доходи	Період 2	Заощадження	Доходи
1946	0,36	8,8	1955	0,59	15,5
1947	0,21	9,4	1956	0,90	16,7
1948	0,08	10,0	1957	0,95	17,7
1949	0,20	10,6	1958	0,82	18,6
1950	0,10	11,0	1959	1,004	19,7
1951	0,12	11,9	1960	1,53	21,1
1952	0,41	12,7	1961	1,94	22,8
1953	0,50	13,5	1962	1,75	23,9
1954	0,43	14,3	1963	1,99	25,2

Інформація поділена в таблиці на два періоди (період I – після II світової війни, період реконструкції) і 1955-1963 ( після реконструкції). Хочемо дослідити, чи зміниться взаємозв'язок агрегованих доходів та заощаджень між двома періодами.

$$\text{Період I} \quad y_i = \alpha_0 + \alpha_1 x_i + \varepsilon_{1i} \quad (6.5)$$

$$\text{Період II} \quad y_i = \gamma_0 + \gamma_1 x_i + \varepsilon_{2i}, \quad (6.6)$$

де  $y$  – заощадження;

$x$  – доходи;

$\varepsilon_{1i}, \varepsilon_{2i}$  – відхилення в обох регресіях.

Необхідно, щоб кількість спостережень в обох групах була однаковою.

Між регресіями (6.5) і (6.6) можливі такі чотири варіанти співвідношень:

- 1)  $\alpha_0 = \gamma_0, \alpha_1 = \gamma_1$  – тобто дві регресії ідентичні;
- 2)  $\alpha_0 \neq \gamma_0, \alpha_1 = \gamma_1$  – розрізняються лише за розміщенням;
- 3)  $\alpha_0 = \gamma_0, \alpha_1 \neq \gamma_1$  – розрізняються за нахилами;
- 4)  $\alpha_0 \neq \gamma_0, \alpha_1 \neq \gamma_1$  – обидві регресії абсолютно різні.

Будуються всі (6.6) регресії, а потім використовуються спеціальні статистичні засоби для перевірки всіх попередніх варіантів. Найбільш поширені: Чоу (Chow)-тест і dummy-змінні.

Для Чоу-теста використовуються 2 припущення:

- 1) випадкові величини  $e_{1i}, e_{2i}$  – нормально розподілені;
- 2) випадкові величини незалежно розподілені.

Крок 1.  $N_1 + N_2$  (об'єднуємо спостереження) і будуємо «збірну» регресію  $y = \alpha + \beta x_i + \varepsilon$ .

Знаходимо SSE з  $df = N_1 + N_2 - k, k = 2$  у прикладі.

Крок 2. Розглянемо окремо дві регресії (1) і (2). Для кожної теж отримуємо SSE з  $df = N_{(2)} - k$ .

Складаємо:  $S_4 = SSE_1 + SSE_2$  з  $df = N_1 + N_2 - 2k$ .

Крок 3.  $S_5 = SSE_{заг} - S_4$ .

Крок 4. F-тест:  $F = \frac{\frac{S_5}{k}}{\frac{S_4}{(N_1 + N_2 - 2k) - 2}}$  – якщо  $F > F_{кр}$ , відхиляємо гіпотезу про те, що дві регресії однакові.

Dummy-тест реалізується через такий алгоритм.

1) Об'єднуються  $N_1$  та  $N_2$  і оцінюється регресія:

$$y_i = \alpha_1 + \alpha_2 D_i + \beta_1 x_i + \beta_2 (D_i x_i) + \varepsilon_i,$$

де  $y_i, x_i$  – заощадження, дохід;

$D_i = 1$  – 1-й період спостереження;

$D_i = 0$  – 2-й період спостереження.

2) Оцінюємо модель:

$$E(y_i / D_i = 0, x_i) = \alpha_1 + \beta_1 x_i$$

$$E(y_i / D_i = 1, x_i) = (\alpha_1 + \alpha_2) + (\beta_1 + \beta_2) x_i,$$

які є функціями середніх заощаджень для другого та першого періодів.

$$\gamma_0 = \alpha_1, \gamma_1 = \beta_1, \alpha_0 = (\alpha_1 + \alpha_2), \alpha_1 = (\beta_1 + \beta_2).$$

$\alpha_2$  – диференціальним перетином;

$\beta_2$  – диференціальним нахилом.

Якщо вони за t-статистикою значущі, то регресії для двох періодів різні, тож можна побудувати дві регресії, тобто рівень залежності в обох періодах відрізняється.

Переваги методу dummy-тесту перед тестом Чоу:

- a) тільки одна регресія;
- b) один тест;
- c) Chow-тест чітко не показує, який коефіцієнт – перетину чи нахилу – відмінний.

### ПИТАННЯ ДЛЯ САМОКОНТРОЛЮ

1. Назвіть основні випадки наявності взаємозв'язку між часовими рядами.
2. У чому сутність явища коінтеграції?
3. Що являє собою хибна кореляція? Хибна регресія?
4. З якою метою проводиться тест Гренджера? Які його недоліки?
5. Які методи усунення тенденції вам відомі? В чому їх сутність?
6. Охарактеризуйте бінарні (дихотомічні) змінні. Яка їх роль при дослідженні взаємозв'язку двох часових рядів?
7. Назвіть основні переваги і недоліки dummy-тесту та тесту Чоу.

### ПРАКТИЧНІ ЗАВДАННЯ

1. Чим відрізняється стохастичний тренд від звичайного лінійного тренду з точки зору усунення проблеми хибної регресії?

2. Які з перерахованих факторів враховуються в регресії за допомогою фіктивних змінних: а) професія; б) курс долара; в) чисельність населення; г) розмір середньомісячних споживчих витрат?

3. У рівнянні регресії для доходів населення вводяться три якісних фактори: стать («чол», «жін»), освіта («початкова», «середня», «вища») і місце проживання («місто», «село»). Скільки фіктивних змінних (з урахуванням усіх взаємодій факторів) у вихідній та перетвореній (після усунення лінійних залежностей) формі рівняння?

4. Розглянемо регресійну модель  $x_t = \alpha_1 z_{t1} + \alpha_2 z_{t2} + \beta_0 + \varepsilon_t$ ,  $t = 1, \dots, T$ . Нехай для спостережень  $t = 1$  і 2 параметри  $\alpha_1$ ,  $\alpha_2$  та  $\beta_0$  відрізняються від інших  $(T-2)$  спостережень. Запишіть регресійну модель з фіктивними змінними і опишіть проблеми оцінювання, що виникли.

5. На основі даних про витрати на автомобілі ( $x$ ) та дохід у розпорядженні ( $Z$ ) за період з 1963 до 1982 рр. отримана модель:  $x = 0,77 + 0,035 \cdot Z - 4,7 \cdot Z_F$ , де  $Z_F$  – фіктивна змінна, що враховує нафтову кризу 1974 року, і яка дорівнює 0 для періоду з 1963 до 1973 рр. та дорівнює 1 для періоду з 1974 до 1982 рр.

Схематично побудуйте графік регресійної функції та дайте повну інтерпретацію. Запишіть модель, в якій якісний фактор  $Z_F$  не впливає на перетин, але впливає на нахил лінії регресії. Схематично побудуйте графік такої регресійної функції.

6. У таблиці наведено значення біржового індексу на момент закриття на Нью-Йоркській фондовій біржі ( $y$ ) та населення Шрі-Ланки в тис. осіб ( $x$ ) за період 1966-1992 роки.

**Вихідні дані до задачі 6\***

Рік	1966	1967	1968	1969	1970	1971	1972	1973	1974
$x$	43,72	53,83	58,9	51,53	50,23	54,63	64,48	51,82	36,13
$y$	11 440	11 702	11 992	12 252	12 516	12 608	12 861	13 091	13 284
Рік	1975	1976	1977	1978	1979	1980	1981	1982	1983
$x$	47,64	57,88	52,5	53,62	61,95	77,86	71,11	81,03	95,18
$y$	13 496	13 717	13 942	14 184	14 471	14 738	14 988	15 189	15 417
Рік	1984	1985	1986	1987	1988	1989	1990	1991	1992
$x$	96,38	121,58	138,58	138,23	156,26	195,01	180,49	229,44	240,21
$y$	15 599	15 837	16 117	16 361	16 587	16 806	16 993	17 190	17 405

\*Джерело: Salvatore D., Reagle D. Theory and Problems of Statistics and Econometrics. 2<sup>nd</sup> edition. Schaum's Outline Series, McGraw-Hill, 2002. – P. 256.

Необхідно:

- 1) Побудувати графіки наведених часових рядів. Зробити висновок про наявність спільної тенденції.
  - 2) Побудувати просту лінійну регресію залежності  $y$  від  $x$  та перевірити коефіцієнт біля змінної  $x$  при 5 %-му рівні значущості тестом Ст'юдента. Інтерпретуйте результати.
  - 3) Побудувати залежність  $\Delta y$  від  $\Delta x$ , перевірити тестом Ст'юдента коефіцієнт біля змінної  $\Delta x$  та зробити висновок про наявність хибної кореляції в попередньому випадку. Зробити висновок, чи дійсно кількість населення Шрі-Ланки визначає рівень біржового індексу на Нью-Йоркській фондовій біржі.
7. У таблиці наведено логарифми споживання ( $y$ ) та ВВП ( $x$ ) в Кореї за період 1953-1991 роки.

**Вихідні дані до задачі 7\***

Рік	$y$	$x$	Рік	$y$	$x$	Рік	$y$	$x$
1953	2,5291	2,8062	1966	3,2181	3,5233	1979	4,3523	4,8255
1954	2,5861	2,8423	1967	3,2817	3,5896	1980	4,3252	4,7699
1955	2,6923	2,9255	1968	3,3758	3,7099	1981	4,3691	4,8233
1956	2,7412	2,9618	1969	3,4438	3,8178	1982	4,4447	4,8941
1957	2,8026	3,0434	1970	3,6410	3,9821	1983	4,5313	4,9981
1958	2,8311	3,0732	1971	3,7531	4,0802	1984	4,6068	5,0865
1959	2,8772	3,0966	1972	3,7829	4,1170	1985	4,6650	5,1479
1960	2,8899	3,1081	1973	3,9005	4,2500	1986	4,7397	5,2488
1961	2,9091	3,1499	1974	3,9615	4,3487	1987	4,8115	5,3528
1962	2,9689	3,1870	1975	4,0161	4,4062	1988	4,9042	5,4610
1963	2,9921	3,2771	1976	4,1105	4,5185	1989	5,0142	5,5538
1964	3,0988	3,3384	1977	4,1584	4,6173	1990	5,1114	5,6562
1965	3,1563	3,4035	1978	4,2465	4,2465	1991	5,1934	5,7485

\*Джерело: Salvatore D., Reagle D. Theory and Problems of Statistics and Econometrics. 2<sup>nd</sup> edition. Schaum's Outline Series, McGraw-Hill, 2002. – P. 248.

Необхідно:

- 1) Побудувати графічно обидва ряди. Зробити висновок про наявність спільної тенденції.
- 2) Перевірити ряди на коінтеграцію тестом Енгла-Гренджера.
- 3) Визначити, яка змінна є причиною, а яка наслідком, користуючись тестом Гренджера, приймаючи для порівняння максимальний лаг, який допускає модель.

## ТЕСТИ

1. Фіктивні змінні – це:

- а) атрибутивні ознаки (наприклад, як професія, стать, освіта), яким надали кількісні значення;
- б) економічні змінні, що мають кількісні значення в деякому інтервалі;
- в) значення залежної змінної за попередній період часу.

2. Dummy-змінна може вживатися, коли:
  - а) незалежна змінна кількісна;
  - б) незалежна змінна якісна;
  - в) незалежна змінна гетероскедастична;
  - г) немає автокореляції;
  - д) проблемою є мультиколінеарність.
3. Якщо ми зацікавлені у використанні dummy-змінних для відображення ефекту різних місяців на віддачу від цінних паперів, ми повинні використати:
  - а) 12 dummy-змінних (атрибутивних змінних), одну для кожного місяця року;
  - б) 1 атрибутивну змінну;
  - в) 11 атрибутивних змінних;
  - г) стільки атрибутивних змінних, скільки ми захочемо;
  - д) не можна використовувати атрибутивну змінну для відображення цього ефекту.
4. Використання атрибутивних змінних для вимірювання різниці в оплаті праці чоловіків та жінок означає, що:
  - а) нахил у моделі для чоловіків та жінок буде однаковим;
  - б) перетин у моделі для чоловіків та жінок буде однаковим;
  - в) нахил та перетин будуть однаковими;
  - г) нахил та перетин будуть різними;
  - д) наявна мультиколінеарність.
5. Хибна кореляція викликається:
  - а) чітко визначеними сезонними коливаннями в часовому ряді;
  - б) однаковою тенденцією в декількох часових рядах;
  - в) впливом випадкової компоненти;
  - г) вірні відповіді а) та б);
  - д) вірні відповіді б) та в).
6. Причинність між двома рядами можна визначити за допомогою:
  - а) теста Бокса-Дженкінса;
  - б) теста Гренджера;
  - в) теста Енгеля;
  - г) методу Монте-Карло;
  - д) будь-якого з вищенаведених.
7. До методів виключення тенденції, що ґрунтуються на перетворенні рівнів вихідного ряду відносяться:
  - а) узагальнений метод найменших квадратів;
  - б) двокроковий метод найменших квадратів;
  - в) метод послідовних різниць;
  - г) метод включення в модель регресії фактору часу;
  - д) жоден не підходить.
8. До методів, заснованих на перетворенні рівнів вихідного ряду в нові змінні, що не містять тенденції, відносять:
  - а) узагальнений метод найменших квадратів;
  - б) двокроковий метод найменших квадратів;
  - в) метод відхилень від трендів;
  - г) метод включення в модель регресії фактору часу;
  - д) жоден не підходить.
9. До методів, заснованих на вивченні взаємозв'язку вихідних рівнів часових рядів при елімінаванні впливу фактору часу відносять:
  - а) метод послідовних різниць;
  - б) метод відхилень від трендів;
  - в) двокроковий метод найменших квадратів;
  - г) трикроковий метод найменших квадратів;
  - д) метод включення в модель регресії за часовими рядами фактору часу.
10. Бінарні змінні впливають:
  - а) як на нахил, так і на перетин регресії;
  - б) тільки на нахил регресії;
  - в) тільки на перетин регресії.
11. Dummy-тест має такі переваги перед Чоу-тестом:
  - а) будується дві регресії;
  - б) показує тільки відмінність коефіцієнту перетину;
  - в) показує тільки відмінність коефіцієнту нахилу;
  - г) будується тільки одна регресія;
  - д) проводиться тільки 2 тести.