

Міністерство освіти і науки України
Харківський національний університет імені В. Н. Каразіна

Кафедра статистики, обліку та аудиту

«ЗАТВЕРДЖУЮ»

Завідувач кафедри

_____ Володимир СОБОЛЄВ

Протокол № 10 від “22” червня 2020 р.

НАВЧАЛЬНО-МЕТОДИЧНИЙ КОМПЛЕКС

дисципліни «Сучасні методи прогнозування»

для студентів денної (заочної) форми навчання

рівень вищої освіти другий (магістерський)

галузь знань 05 «Соціальні та поведінкові науки»

спеціальність 051 «Економіка»

освітня програма «Бізнес-аналітика та міжнародна статистика»,

«Економічна аналітика та статистика»

Розроблено:

д.е.н., доцент, доцент кафедри статистики, обліку та аудиту

Корепанов Олексій Сергійович

2020 / 2021 навчальний рік

ЗМІСТ

1. Робоча програма навчальної дисципліни
2. Навчальний контент (розширений план лекцій, конспект лекцій)
3. Плани практичних (семінарських) занять, самостійної роботи (у т.ч. індивідуальні завдання)
4. Питання, задачі, завдання або кейси для поточного контролю знань і вмінь здобувачів вищої освіти, для контрольних робіт, передбачених навчальним планом, післятестаційного моніторингу набутих знань і вмінь з навчальної дисципліни
 - 4.1. Питання, задачі, завдання або кейси для поточного контролю знань і вмінь здобувачів вищої освіти
 - 4.2. Контрольні роботи, передбачені навчальним планом
5. Завдання підсумкового контролю (семестрових екзаменів, письмових залікових робіт)
6. Критерії оцінювання знань студентів та розподіл балів

1. РОБОЧА ПРОГРАМА НАВЧАЛЬНОЇ ДИСЦИПЛІНИ

Міністерство освіти і науки України

Харківський національний університет імені В. Н. Каразіна

Кафедра статистики, обліку та аудиту

“ЗАТВЕРДЖУЮ”

Проректор з науково-педагогічної
роботи

_____ Антон ПАНТЕЛЕЙМОНОВ

“ _____ ” _____ 2020 р.

Робоча програма навчальної дисципліни

“СУЧАСНІ МЕТОДИ ПРОГНОЗУВАННЯ”

(шифр і назва навчальної дисципліни)

рівень вищої освіти _____ другий (магістерський) _____

галузь знань _____ 05 «Соціальні та поведінкові науки» _____

спеціальність _____ 051 «Економіка» _____

освітня програма _____ «Бізнес-аналітика та міжнародна статистика» _____

спеціалізація _____

вид дисципліни _____ обов'язкова _____

факультет _____ економічний _____

2020 / 2021 навчальний рік

Міністерство освіти і науки України
Харківський національний університет імені В. Н. Каразіна

Кафедра статистики, обліку та аудиту

“ЗАТВЕРДЖУЮ”

Проректор з науково-педагогічної
роботи

_____ Антон ПАНТЕЛЕЙМОНОВ

“ _____ ” _____ 2020 р.

Робоча програма навчальної дисципліни

“СУЧАСНІ МЕТОДИ ПРОГНОЗУВАННЯ”

(шифр і назва навчальної дисципліни)

рівень вищої освіти _____ другий (магістерський) _____

галузь знань _____ 05 «Соціальні та поведінкові науки» _____

спеціальність _____ 051 «Економіка» _____

освітня програма _____ «Економічна аналітика та статистика» _____

спеціалізація _____

вид дисципліни _____ обов’язкова _____

факультет _____ економічний _____

2020 / 2021 навчальний рік

2. НАВЧАЛЬНИЙ КОНТЕНТ

ТЕМАТИЧНИЙ ПЛАН НАВЧАЛЬНОЇ ДИСЦИПЛІНИ

Розділ 1. Прогнозування динаміки соціально-економічних явищ із використанням сучасних методів

Тема 1. Аналіз одновимірних часових рядів

- 1.1. Предмет, метод, завдання курсу
- 1.2. Базовий алгоритм методу «Гусениця»
- 1.3. Формальний опис базового алгоритму методу «Гусениця»

Тема 2. Вибір параметрів при застосуванні методу «Гусениця»

- 2.1. Вибір довжини гусениці
- 2.2. Відбір головних компонент

Тема 3. Опис програми «Caterpillar» для реалізації методу «Гусениця»

- 3.1. Загальна структура програми
- 3.2. Графічне представлення результатів
- 3.3. Керівництво користувача
 - 3.3.1. Технічна інформація
 - 3.3.2. Загальні положення
 - 3.3.3. Формати файлів вихідних даних
 - 3.3.4. Параметри етапів
- 3.4. Робота з графікою

Розділ 2. Приклади реалізації сучасних методів прогнозування

Тема 4. Приклад реалізації методу «Гусениця» для аналізу одновимірних часових рядів

- 4.1. Аналіз динаміки виробництва м'яса усіх видів (у живій вазі) в Україні
- 4.2. Оцінка внеску складових досліджуваного часового ряду
- 4.3. Прогнозування обсягів виробництва м'яса усіх видів (у живій вазі) в Україні

Тема 5. Визначення наявності тренда

- 5.1. Сутність процедури вирівнювання динамічних рядів та визначення тренду
- 5.2. Побудова трендових моделей МНК
- 5.3. Екстраполяція тренда

Тема 6. Статистичний аналіз сезонних та циклічних коливань

- 6.1. Індекси сезонності
- 6.2. Сезонно-декомпозиційна модель Холта-Вінтера
- 6.3. Моделювання сезонної хвилі на основі гармонійного аналізу

КОРОТКИЙ КОНСПЕКТ ЛЕКЦІЙ

Розділ 1. Прогнозування динаміки соціально-економічних явищ із використанням сучасних методів

Тема 1. Аналіз одновимірних часових рядів

1.1. Предмет, метод і завдання курсу

1.2. Базовий алгоритм методу «Гусениця»

1.3. Формальний опис базового алгоритму методу «Гусениця»

1.1. Предмет, метод і завдання курсу

Дослідження часових рядів здійснюється з використанням різних статистичних методів, зокрема різних моделей динаміки, а саме трендових, авторегресії і ковзних середніх, сезонної декомпозиції, спектрального аналізу Фур'є, моделей багатофакторного прогнозування.

Необхідність виділення окремих складових часових рядів, таких як повільний тренд, періодичні та сезонні коливання обумовлюють особливу увагу до курсу «Сучасні методи прогнозування».

Для задач аналізу та прогнозування одновимірних часових рядів доцільно застосовувати метод сингулярного спектрального аналізу, або метод «Гусениця» у вітчизняній літературі.

Метод «Гусениця» об'єднує елементи класичного аналізу, методи нелінійної динаміки та обробки сигналів. Відмінною рисою методу є те, що він дозволяє працювати з широким спектром нестационарних часових рядів та не вимагає попереднього завдання моделі ряду, але він дозволяє розкласти часовий ряд на адитивні складові, що можна інтерпретувати, як тренд (зокрема, лінійний або експоненціальний), періодичні та сезонні компоненти й шум. При цьому не треба заздалегідь знати параметричний вид тренду, а також наявність коливальних компонентів та їхніх періодів.

Технічну основу методу становить сингулярне розкладання траєкторної матриці, рядками якої є послідовно взяті відрізки ряду довжини M – основного параметра методу, названого довжиною “гусениці” (довжиною “вікна”). На основі такого розкладання одержують власні трійки, кожна з яких складається з власного числа, власного вектору та факторного вектору, які далі групуються та відновлюється адитивна складова вихідного ряду.

Сутність методу полягає в:

- перетворенні одновимірного ряду в багатовимірний,
- дослідженні отриманої багатовимірної траєкторії методом головних компонент і
- наступному відновленні багатовимірного ряду.

При цьому часто виявляється можливим виділити окремі складові вихідного ряду, такі як повільний тренд загального вигляду, повільні сезонні складові, періодичні складові (якщо вони є) і випадкові варіації. Це, в свою чергу, дозволяє прогнозувати як сам часовий ряд, так і тенденції розвитку різних його складових.

Метод допускає природне узагальнення на випадок аналізу багатовимірних часових рядів, і навіть на випадок аналізу зображень.

Під часовим рядом ми розуміємо набір $\{f(t_i)\}_{i=1}^N$ значень, взагалі кажучи, довільної функції однієї змінної $f(t)$, $t > 0$, в рівновіддалених точках t_i ($i = 1, \dots, N$). При цьому змінна t не обов'язково має сенс часу.

У багатьох природних науках є уявлення про можливості опису природних процесів за допомогою функцій, що складаються з декількох складових:

$$f(t) = f_T(t) + f_n(t) + f_r(t) + \varepsilon(t), \quad t = 0, \dots, T,$$

де $f_T(t)$ – повільна нерегулярна складова, так званий тренд, часто її намагаються описувати алгебраїчними поліномами невисоких порядків;

$f_n(t)$ – періодична складова або сума періодичних складових; в залежності від області додатків вони називаються сезонними, добовими і т.п. коливаннями;

$f_r(t)$ – швидкі нерегулярні малі коливання, в які зазвичай включають все, що не входить в формальну модель, іноді сюди включають і випадкові шуми;

$\varepsilon(t)$ – чисто випадкова складова, описувана випадковим процесом певного типу.

У багатьох випадках у теперішній час створені потужні теорії з розвиненим апаратом додатків і комп'ютерними реалізаціями у вигляді бібліотек і пакетів програм. Так, для функцій виду $f(t) = f_T(t) + \varepsilon(t)$ такою теорією є теорія апроксимації (при малих $\varepsilon(t)$) або метод найменших квадратів математичної статистики (при великих $\varepsilon(t)$), для функцій виду $f(t) = f_n(t)$ добре працює теорія гармонійних рядів Фур'є.

Однак, у багатьох ситуаціях виникають чималі складнощі ефективного дослідження функцій. Прикладом може бути випадок $f(t) = f_T(t) + f_n(t)$, де при відсутності апріорної інформації про частоти компонент періодичної складової не працюють ні теорія апроксимації, ні теорія рядів Фур'є. У різних конкретних додатках придумано багато евристичних прийомів, але вони, як правило, погано теоретично обґрунтовані. При використанні методу «Гусениця» часто виявляється можливим (груба результати) виділити окремі складові вихідного ряду.

1.2. Базовий алгоритм методу «Гусениця»

Базовий алгоритм методу «Гусениця» можна розбити на чотири етапи.

Етап 1. Розгортка одновимірною ряду в багатовимірний.

Для дослідження обирається часовий ряд $\{x_i\}_{i=1}^N$, який утворений послідовністю N рівновіддалених значень деякої функції $f(t)$:

$$x_i = f[i] = f((i-1)\Delta t), \text{ де } i = 1, 2, \dots, N.$$

Виберемо деяке число $M < N$, що має назву «довжина гусениці», і представимо перші M значень послідовності f в якості першого рядка матриці X . Як другий рядок матриці беремо значення часової послідовності з x_2 до x_{M+1} . Останнім рядком з номером $k = N - M + 1$ будуть останні M елементів послідовності. Цю матрицю, елементи якої дорівнюють $x_{ij} = x_{i+j-1}$, можна

розглядати як M – мірну вибірку обсягу k або M – мірний часовий ряд, якому відповідала б M – мірна траєкторія (ломана в M – мірному просторі з $k-1$ ланки.

$$X = (x_{ij})_{i,j=1}^{k,M} = \begin{pmatrix} x_1 & x_2 & x_3 & \dots & x_M \\ x_2 & x_3 & x_4 & \dots & x_{M+1} \\ x_3 & x_4 & x_5 & \dots & x_{M+2} \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ x_k & x_{k+1} & x_{k+2} & \dots & x_N \end{pmatrix} \quad (1)$$

Відзначимо, що матриця X (будемо називати її матрицею ряду) представлена в традиційному для статистики вигляді «рядок - індивід, стовпець – ознака». При викладі теоретичних аспектів методу природним є відповідність індивіду стовпчика.

Далі за звичайною схемою (за винятком стандартизації ознак) проводиться аналіз головних компонент (АГК).

Етап 2. Аналіз головних компонент: сингулярне розкладання вибіркової коваріаційної матриці.

Спочатку обчислюються середнє арифметичне значення і стандартні відхилення по стовпцях матриці X .

$$\bar{x}_j = \frac{1}{k} \sum_{i=1}^k x_{i+j-1}; \quad s_j = \sqrt{\frac{1}{k} \sum_{i=1}^k (x_{i+j-1} - \bar{x}_j)^2} \quad (2)$$

Позначимо через $X^* = (x_{ij}^*)_{i,j=1}^{k,M}$ матрицю, яка була отримана з матриці X в результаті центрування по стовпцям і нормування на середні квадратичні відхилення s_j :

$$x_{ij}^* = (x_{ij} - \bar{x}_j) / s_j, \quad i = 1, \dots, k, \quad j = 1, \dots, M, \quad (3)$$

Слід зазначити, що операції віднімання вибірових середніх (центрування) і ділення на середні квадратичні відхилення (нормування) не є обов'язковими.

Можуть бути відсутніми або операція нормування, або обидві операції. До того ж може бути зроблено подвійне центрування - центрування елементів матриці як по стовпцям, так і по рядкам.

Далі розрахуємо матрицю других моментів:

$$R = \frac{1}{k} X^* (X^*)^T \quad (4)$$

Для обчислення власних чисел і власних векторів матриці, розкладемо її в такий спосіб:

$$R = P \Lambda P^T \quad (5)$$

де $\Lambda = \begin{pmatrix} \lambda_1 & 0 & \dots & 0 \\ 0 & \lambda_2 & \dots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \dots & \lambda_M \end{pmatrix}$ – діагональна матриця власних чисел,

$P = (p_1, p_2, \dots, p_M) = \begin{pmatrix} p_{11} & p_{21} & \dots & p_{M1} \\ p_{12} & p_{22} & \dots & p_{M2} \\ \vdots & \vdots & \ddots & \vdots \\ p_{1M} & p_{2M} & \dots & p_{MM} \end{pmatrix}$ – ортогональна матриця

власних векторів матриці. При цьому виконуються такі співвідношення:

$$P^T = P^{-1}; \quad P^T P = P P^T = I_M, \quad \Lambda = P^T R P, \quad \sum_{i=1}^M \lambda_i = M, \quad \prod_{i=1}^M \lambda_i = \det R.$$

Матриця P може бути розглянута як матриця переходу до головних компонентів:

$$X^* P = Y = (y_1, y_2, \dots, y_M). \quad (6)$$

Власні числа матриці R є вибірковими дисперсіями відповідних головних компонент, а квадратні корені з них - вибірковими стандартами, якщо маємо вибірку з випадкової сукупності. В аналізі головних компонент (АГК) графічне представлення власних чисел і деяких функцій від них використовується для виявлення структури досліджуваної сукупності і вибору та інтерпретації головних компонент.

Зауважимо також, що при виборі довжини гусениці, що дорівнює $N-M+1$, власні вектора і головні компоненти (з точністю до нормування) просто міняються місцями.

Етап 3. Вибір головних компонент.

В силу властивостей матриці P ми можемо уявити матрицю ряду X як $X = Y P^T$. Таким чином, ми отримуємо розкладання матриці ряду по ортогональним складовим (головним компонентам).

Пронормуємо значення головних компонент на вибіркові стандарти (при $\lambda_M \neq 0$):

$$Y^* = Y \Lambda^{-1/2} = (y_1^*, y_2^*, \dots, y_M^*).$$

Головні компоненти виявляються ортонормованими:

$$Y^{*T} Y^* = I_M.$$

Таким чином, маємо розкладання вихідного M -мірного процесу на ортогональні компоненти. Вектор y_j може бути розглянуто як результат проектування вихідної M -мірної нормованої і центрованої сукупності в напрямках, які визначаються відповідними власними векторами p_j . Тому майже завжди можна проінтерпретувати кожен з векторів y_j і відповідний власний вектор p_j .

Розглянувши перетворення $y_j = X^* p_j$, можна відзначити, що воно дуже близько до лінійного перетворення вихідного процесу за допомогою дискретного оператора згортки, тобто

$$y_j[l] = \sum_{q=1}^M X_{lq}^* p_{jq} = \sum_{q=1}^M (x_{l+q-1} - \bar{x}_q) \frac{p_{jq}}{s_q} = \sum_{q=1}^M x_{l+q-1} \frac{p_{jq}}{s_q} - \sum_{q=1}^M \bar{x}_q \frac{p_{jq}}{s_q} .$$

Таким чином, процедура «Гусениця» породжує набір лінійних фільтрів, які налаштовані на складові вихідного процесу. При цьому власні вектори матриці R є перехідними функціями відповідних фільтрів. При відсутності нормування і центрування вони точно відповідають перехідним функціям лінійних фільтрів.

Візуальне і аналітичне вивчення як власних векторів, так і головних компонент, отриманих в результаті лінійної фільтрації, може дати багато цікавої інформації про структуру досліджуваного процесу і властивості складових його доданків.

Зокрема, серед головних компонент можна виділити такі:

1. що відносяться до тренду (повільно мінливі);
2. періодичні;
3. шумові.

Для знаходження періодичних складових надзвичайно велику візуальну інформацію дає вивчення двовимірних графіків, аналогічних фігур Ліссажу, коли по осях x та y відкладаються різні пари власних векторів або головних компонент. Відомо, що, якщо по осях відкладати значення синусоїди однієї і тієї ж частоти, але з різними фазами, то на площині виходить еліпс. З ортогональності власних векторів і головних компонент виходить, що зрушення фаз між такими парами обов'язково буде дорівнює $\pi R / 2$ і еліпс переходить в окружність.

Етап 4. Відновлення одновимірного ряду.

Ключовим елементом методу «Гусениця» є процедура відновлення. Ця процедура заснована на розкладанні $X = Y P^T$.

Матриця X^* може бути відновлена в результаті перемноження матриці головних компонент Y та P^T , що виходить з ортогональності матриці P . При цьому отримано розкладення нормованої та центрованої матриці X^* у суму матриць X_l^* , кожна з котрих породжена одним власним вектором матриці R :

$$X^* = Y P^T = (y_1, y_2, \dots, y_M) \begin{pmatrix} p_1^T \\ p_2^T \\ \vdots \\ p_M^T \end{pmatrix} = \sum_{l=1}^M y_l p_l^T = \sum_{l=1}^M X_l^* .$$

Ця процедура є розкладанням нормованої і центрованої матриці X^* в суму матриць, кожна з яких породжена одним власним вектором матриці.

Будемо говорити, що відновлення проводиться з даного набору головних компонент, якщо при застосуванні формули відновлення $X = Y^* P^T$ матриця Y^* отримана з матриці Y обнуленням всіх головних компонент, які не входять в

набір. Таким чином, ми можемо отримати наближення матриці ряду, що нас цікавить, нас або частину цієї матриці, що може бути проінтерпритована.

Перехід до вихідного ряду формально може бути здійснений усередненням матриці ряду по побічним діагоналям і може привести до деякого спотворення отриманої структури.

Наступним кроком є денормування.

Денормування відбувається за допомогою перемноження цієї матриці та діагональної матриці S , що складається з вибірових стандартів, а децентрування шляхом додавання до елементів кожного стовбця відповідних вибірових середніх:

$$X = \bar{x}1_k^T + X^*S = X_0^* + \sum_{l=1}^M X_l^*S = \sum_{l=0}^M X_l^*S.$$

Отримавши в результаті вихідну матрицю діагональної структури у вигляді суми $(M+1)$ матриць знову переходимо до вихідного ряду, здійснивши усереднення по побічним діагоналям:

$$x = A(X) = \sum_{l=0}^M A(X_l^*S). \quad (7)$$

де A - оператор усереднення. Таким чином, маємо розкладення вихідного часового ряду на суму $(M+1)$ рядів.

Відзначимо відразу одну дуже важливу особливість описаного методу - його інтерактивність, тобто використання діалогу дослідника і ЕОМ в процесі застосування методу. Причому ця інтерактивність не повинна розглядатися як наслідок загальної тенденції розвитку сучасного програмного забезпечення персональних ЕОМ. Швидше, навпаки, ефективна реалізація алгоритму стала можливою тільки завдяки можливостям сучасних ПК. Його інтерактивність пов'язана з типово статистичною властивістю алгоритму - необхідністю інтерпретації проміжних результатів і управлінням роботою алгоритму в процесі багатоетапної процедури обробки. Досвід багаторічного використання різних реалізацій алгоритму на ЕОМ різних поколінь показав, що чим більше проміжних результатів вдається «побачити», тим повнішим виявляється рішення поставленого завдання.

Зауважимо, що у разі програмної реалізації етапи наведеного вище алгоритму не відображаються безпосередньо. Наприклад, етап побудови матриці X відсутній, а формули перетворені до виду, зручного для проведення прискорених обчислень.

1.3. Формальний опис базового алгоритму методу «Гусениця»

Наведемо формальний опис базового алгоритму методу «Гусениця» для аналізу одновимірною часового ряду $\{x_i\}_{i=1}^N$.

I. Розгортка одновимірною ряду в багатовимірний.

Виберемо деякий число $M < N$ (довжину гусениці), припустимо, що $k = N - M + 1$, утворюємо матрицю $X = (x_{ij})_{i,j=1}^{k,M}$ з елементами $x_{ij} = x_{i+j-1}$.

II. Обчислення матриці X^* .

По матриці X обчислимо матрицю $X^* = (x_{ij}^*)_{i,j=1}^{k,M}$. Загальна формула для елементів цієї матриці така:

$$x_{ij}^* = (x_{ij} - \mu_j) / \delta_j, \quad i=1, \dots, k; \quad j=1, \dots, M. \quad (8)$$

$$\mu_j = \begin{cases} 0 \\ \bar{x}_j \end{cases}; \quad \delta_j = \begin{cases} 1 \\ s_j \end{cases}, \quad (9)$$

де

$$x_j = \frac{1}{k} \sum_{i=1}^k x_{ij}, \quad s_j = \sqrt{\frac{1}{k} \sum_{i=1}^k (x_{ij} - x_j)^2}.$$

III. Обчислення матриці R із використанням формули (4).

IV. Сингулярне розкладання матриці R у вигляді (5).

V. Обчислення головних компонент згідно (6).

VI. Відбір головних компонент: вибираємо кілька r головних компонент з номерами $i_1, \dots, i_r$.

VII. Відновлення багатовимірного ряду:

$$\tilde{X}^* = \sum_{l=1}^r y_{i_l} p_{i_l}^T.$$

$$\tilde{x}_s = \begin{cases} \frac{1}{s} \sum_{i=1}^s \tilde{x}_{s-i+1,i} & 1 \leq s \leq M; \\ \frac{1}{M} \sum_{i=1}^M \tilde{x}_{s-i+1,i} & M \leq s \leq k; \\ \frac{1}{N-s+1} \sum_{i=1}^{N-s+1} \tilde{x}_{k-i+1,i+s-k} & k \leq s \leq N. \end{cases}$$

VIII. Децентрування и денормування (у разі необхідності) матриці $\tilde{X}^*$, результат – матриця $\tilde{X}$.

IX. Ганкелізація (відновлення одновимірного ряду шляхом усереднення по побічним діагоналям $\tilde{X}$).

Спосіб завдання μ_j и δ_j в формулах (9) буде визначати назву методу:

$\mu_j = 0$ – «Гусениця» без центрування;

$\mu_j = \bar{x}_j$ – «Гусениця» з центруванням;

$\delta_j = s_j$ – «Гусениця» з нормуванням;

$\delta_j = 1$ – «Гусениця» без нормування.

Так, «Гусениця» з центруванням і нормуванням (кореляційна) визначається вибором $\delta_j = s_j$, $\mu_j = \bar{x}_j$. «Гусеницю» з центруванням, але без нормування ($\delta_j = 1$, $\mu_j = \bar{x}_j$) будемо називати коваріаційною.

До чотирьох варіантів алгоритму, який визначається різними комбінаціями формули (9), можна додати ще один – «Гусеницю» з подвійним центруванням, в якій

$$x_{ij}^* = x_{ij} - (\tilde{x}_i + \bar{x}_j - \mu); \quad \tilde{x}_i = \frac{1}{M} \sum_{j=1}^M x_{ij}; \quad \mu = \frac{1}{kM} \sum_{i=1}^k \sum_{j=1}^M x_{ij}.$$

Зазвичай з контексту ясно, про який з варіантів «Гусениці» йдеться.

На рис. 1.1 зображена блок-схема алгоритму проведення сингулярного спектрального аналізу досліджуваних часових рядів та їх прогнозування.

При програмній реалізації пункти наведеного вище алгоритму не відображаються безпосередньо. Наприклад, етап побудови матриць X і X^* відсутній, а формули перетворені до вигляду, зручного для проведення прискорених обчислень.

Метод «Гусениця» об'єднує елементи класичного аналізу, методи нелінійної динаміки і обробки сигналів. Відмінною рисою методу є те, що він дозволяє працювати з широким спектром нестационарних часових рядів і не вимагає попереднього завдання моделі ряду, але він дозволяє розкладати часовий ряд на адитивні складові, які можна інтерпретувати, як тренд (зокрема, лінійний або експоненціальне), періодичні і сезонні компоненти і шум. При цьому не треба заздалегідь знати параметричний вид тренда, а також наявність коливальних компонентів і їх періодів.

Рекомендована література

1. Голяндина Н. Э. Метод «Гусеница»-SSA : анализ временных рядов : Учеб. пособие. – СПб : Изд-во СПбГУ, 2004. – 76 с.
2. Голяндина Н. Э. Метод «Гусеница»-SSA : прогноз временных рядов : Учеб. пособие. – СПб : Изд-во СПбГУ, 2004. – 52 с.
3. Главные компоненты временных рядов: метод «Гусеница» / Под ред. Д.Л.Данилова и А.А.Жиглявского // Санкт-Петербургский университет. URL: <http://www.gistatgroup.com/gus/book1/index.html>
4. Корепанов О. С. Статистичний аналіз тенденцій зовнішньоекономічної діяльності України: сектор ІКТ // Соціальна економіка. 2018. № 1, Вип. 55. С. 18–26.
5. Корепанов О. С., Прус Ю.І., Чала Т.Г. Короткострокове прогнозування пропозиції робочої сили на ринку праці України // Бізнес Інформ. 2017. № 4. С. 216–222.

6. Hyndman R. J. Forecasting: principles and practice / R. J. Hyndman, G. Athanasopoulos. – Otexts, 2018. – 504 p.

Тема 2. Вибір параметрів при застосуванні методу «Гусениця»

2.1. Вибір довжини гусениці

2. Відбір головних компонент

1. Вибір довжини гусениці

Як видно з опису методу «Гусениця», основним параметром, необхідним для його реалізації, є M - довжина гусениці.

При геометричній інтерпретації цей параметр є розмірністю простору, в якому досліджується траєкторія багатовимірної ламаної лінії, в яку переводиться вихідний часовий ряд процедурою «гусениця».

Природною умовою є $M < N / 2$, так як розмірність множини k точок (вершин ламаної) в M -вимірному просторі не перевищує $\min(M, k-1)$. Цей підхід тісно пов'язаний з аналітичною інтерпретацією методу гусениця як апроксимації рішення рівняння в кінцевих різницях з l коефіцієнтами.

Можна сказати, що l - це число ступенів свободи функції $f(t)$, а отже, й відповідного їй часового ряду. У процедурі «Гусениця» це виразиться в тому, що при $M > l$ в кореляційній матриці виявиться тільки l ненульових власних чисел. (Для визначеності розглядається кореляційна «Гусениця»).

Реально вихідні дані виражаються числами з кінцевим (і зазвичай невеликим) числом цифр, що відповідає наявності помилок округлення. Тому замість $M-1$ нульових власних чисел виходять маленькі власні числа, але в АГК відпрацьовані прийоми відсікання малих власних чисел.

Отже, якщо виникає ситуація, коли частина власних чисел дуже близька до нуля і при цьому може бути відокремлена від решти, то може бути висунута гіпотеза про детерміновану структуру часового ряду. Однак, ця ситуація досить рідкісна і при дослідженні природних процесів може бути постульована тільки для відносно коротких (в порівнянні з довжиною гусениці) часових рядах.

У загальному випадку вибір довжини гусениці істотно залежить від завдання, розв'язуваного цим методом.

Розглянемо три випадку, найбільш типових і досліджених розробниками.

а) якщо вирішується завдання аналізу вихідного часового ряду, наприклад, з метою пошуку прихованих періодичностей з невідомими періодами, то брати дуже коротку гусеницю ризиковано, так як може виявитися $M < l$ (в таких ситуаціях l зазвичай невідомо і може бути досить великим). Тут можна рекомендувати виконати два проходи. Спочатку здійснюється обчислення власних чисел при максимальній довжині гусениці, яку допускає наявна у дослідника програмна реалізація методу, і оцінюється l . Потім проводиться повне дослідження ряду при довжині гусениці M , що дорівнює або трохи більшою оціненого l . Цей «запас» необхідний, оскільки в методі «гусениця» в загальному випадку не виконується правило АГК, що інформативність головної компоненти

пропорційна величині відповідного власного числа. Бувають випадки, коли деякі періодичності проявляються при досить малих власних числах.

б) Якщо вирішується вужче завдання, наприклад, згладжування вихідного ряду або виділення періодичності з відомим періодом, то працює інший механізм, пов'язаний з «Гусеницею» - механізм фільтрації. Як уже зазначалося, виділення деякої головної компоненти в M -вимірному поданні вихідного часового ряду еквівалентно пропусканню ряду через фільтр, перехідна функція якого збігається з власним вектором цієї компоненти. Спектральна функція фільтра дорівнює перетворенню Фур'є цього власного вектора. Ширина смуги пропускання залежить від форми перехідної функції фільтра і визначається як видом власного вектора, так і довжиною інтервалу усереднення, тобто довжиною гусениці M . Чим більше M , тим вужчою може бути смуга фільтра. Вибір кількох головних компонент еквівалентний паралельному з'єднанню відповідних фільтрів, що дозволяє управляти формою спектральної характеристики.

в) ще один випадок вибору M - при необхідності найкращого виділення (або виключення) періодичного, не обов'язково гармонійного, коливання з певною частотою i , отже, періодом. У цьому випадку буває необхідно для кращого вирішення завдання змінити довжину часового ряду N (у бік зменшення). З теорії методу виходить, що періодичні коливання найкращим чином виділяються, якщо M дорівнює, а N - є кратним довжині періоду коливання, що виділяється.

Слід пам'ятати, однак, що метод надзвичайно стійкий щодо зміни довжини гусениці, і, тому, коли мова йде про вибір M , то слід розуміти, що проявляється резонансний ефект щодо довжини гусениці не стільки в кількісному, скільки в якісному сенсі.

2.2. Відбір головних компонент

Наступним елементом методики проведення аналізу методом «Гусениця», який не може бути виконаний апіорі, є відбір головних компонент, інформативних в тому чи іншому сенсі. Нагадаємо, що маємо справу з двома паралельними уявленнями вихідного ряду:

- поданням ряду у вигляді суми доданків, які не зобов'язані бути ортогональними;

- поданням ряду у вигляді траєкторії в M -вимірному просторі, яка в ортонормального базисі головних компонент (власних векторів кореляційної матриці) розкладається на ортогональні в сукупності компоненти.

Для цих двох уявлень є 4 набори інтерпретованих об'єктів:

- власні числа кореляційної матриці M -мірного подання вихідного одновимірного ряду. Їх інтерпретація багато в чому схожа з АГК.

- набір власних векторів кореляційної матриці. Оскільки їх елементи впорядковані оператором формування гусениці, їх можна вивчати як функції часу (точніше, часові ряди довжини M).

- набір головних компонент M -мірного подання. Вони, як і відповідні їм власні вектори, утворюють ортогональну систему і також можуть бути представлені як функції від номера елемента.

– усі можливі часові ряди, відновлені за різними множинами головних компонент, одержувані в результаті послідовного застосування двох зворотних операторів (оператора переходу від головних компонент до вихідного M -мірного подання і оператора A переходу від M -мірного до одновимірного подання часового ряду).

Слід розділити два крайні випадки, де логіка відбору дещо відрізняється:

I. При $M \ll N$ ми маємо справу з довгими (з точки зору гусениці) рядами. У цьому випадку більш природним здається інтерпретація власних векторів як перехідних функцій лінійних фільтрів, а відповідних головних компонент як результатів дії цих фільтрів. Тут може бути застосована термінологія, пов'язана з лінійною фільтрацією така, як «ширина смуги пропускання», фільтри «високих і низьких частот» і тому подібні терміни спектрального підходу.

II. При $M \approx N / 2$ метод «Гусениця» можна інтерпретувати як метод апроксимації вихідного часового ряду рядами кінцевого рангу. Тут більш доречний геометричний підхід до інтерпретації окремих кроків і результатів застосування методу. Найбільш простим є пошук гармонійних компонент досліджуваного процесу. Як вже зазначалося вище, кожному синусоїдальному доданку ряду відповідають дві головні компоненти, що мають вид відрізків синуса і косинуса однієї і тієї ж частоти. Їх легко виявляти по двовимірним графіками для пар власних векторів кореляційної матриці або пар відповідних головних компонент. Крім того, для таких пар власні числа зазвичай виявляються досить близькими, тому виявляється можливим визначати ці пари за графіком власних чисел або квадратних коренів або логарифмів з них.

Рекомендована література

1. Голяндина Н. Э. Метод «Гусеница»-SSA : анализ временных рядов : Учеб. пособие. – СПб : Изд-во СПбГУ, 2004. – 76 с.
2. Главные компоненты временных рядов: метод «Гусеница» / Под ред. Д.Л.Данилова и А.А.Жиглявского // Санкт-Петербургский университет. URL: <http://www.gistatgroup.com/gus/book1/index.html>
3. Корепанов О. С. Статистичний аналіз тенденцій зовнішньоекономічної діяльності України: сектор ІКТ // Соціальна економіка. 2018. № 1, Вип. 55. С. 18–26.
4. Корепанов О. С., Прус Ю.І., Чала Т.Г. Короткострокове прогнозування пропозиції робочої сили на ринку праці України // Бізнес Інформ. 2017. № 4. С. 216–222.

Тема 3. Опис програми «Caterpillar», що реалізує метод «Гусениця»

3.1. Загальна структура програми

3.2. Графічне представлення результатів

3.3. Керівництво користувача

3.4. Робота з графікою

3.1. Загальна структура програми

При величезному різноманітті реальних даних математичне обґрунтування методу може дати лише деякі базові уявлення. По-справжньому навчитися застосовувати метод можна лише, набравши великий досвід при обробці реальних даних. Програма «Гусениця» представляє собою зручний засіб для досягнення цієї мети.

Робота програми розділена на три послідовні етапи (не рахуючи читання даних з файлу), кожен з яких виробляє обчислювальні дії:

1. Перетворення ряду. До них відноситься лінійне перетворення, а також виділення частини ряду і його проріджування.

2. Розкладання. На цьому етапі проводиться вибір параметра «довжина гусениці» і аналіз головних компонент (АГК) матриці спостережень, побудованої із значень ряду відповідно до цього параметру.

3. Відновлення. На цьому етапі проводиться відновлення ряду за обраними головними компонентами.

Перерахуємо графічні вікна, що з'являються в результаті роботи програми:

1. Часовий ряд.
2. Інформація на рівні перших двох моментів (середні, дисперсії, кореляційний (або ковариационная) функція).
3. Власні числа у відсотках (самі числа і накопичені).
4. Логарифми і коріння з власних чисел.
5. Власні вектора.
6. Головні компоненти.
7. Відновлений ряд і ряд залишків (тобто ряд різниць між вихідним рядом і відновленим).

Зауважимо, що графічні вікна з номерами 2-6 з'являються в результаті роботи другого етапу і відносяться до результатів аналізу головних компонент матриці спостережень, тому в їхніх назвах присутня термінологія АГК.

Для того, щоб полегшити роботу з програмою для початківців, в програму був введений Навігатор, що дозволяє послідовно рухатися вперед (задаючи параметри і переглядаючи результати, корисні при завданні параметрів на наступному етапі). У той же час, щоб жорстка послідовність не заважала досвідченому користувачеві, передбачена можливість переглядати результати за вибором.

3.2. Графічне представлення результатів

Важливою частиною реалізації методу є візуальне представлення результатів. Можливості сучасного графічного інтерфейсу дозволяють приділити графічним уявленням велику увагу.

Відзначимо ряд понять і можливостей розробленої програми.

Так як ми повинні аналізувати одночасно велику кількість головних компонент, щоб відібрати потрібні для відновлення, було введено поняття Мультиграфіка, що дозволяє зображати одночасно потрібну кількість власних векторів або головних компонент.

Так як однієї гармонійної складової ряду (періодиці) відповідають дві головні компоненти, то було введено поняття парної графіки, що допомагає знаходити ці пари ГК. Для знаходження відповідних один одному головних компонент корисним виявилось їх двомірне зображення.

Так як пара головних компонент, відповідна одній періодиці, повинна бути пов'язана з близькими за значенням власними числами (в ідеальному випадку з одним і тим же власним числом), то в програмі пропонується графік логарифмів і коренів з власних чисел, що дозволяє краще побачити «сходінки» в поведінці власних чисел.

Інформація на рівні перших двох моментів дає можливість побачити, по-перше, динаміку середніх характеристик ряду, а, по-друге, побачити по кореляційної функції основні періодици ряду.

Формат даних і видача результатів

При написанні програми «Гусениця» велику увагу було приділено способам друку і збереження графічних і текстових результатів роботи програми.

Одним з реалізованих принципів є те, що графік повинен містити всю інформацію, за якою його можна повторити. Це дуже важливо для роботи з програмою, так як в процесі роботи виявляється необхідним роздруковувати велику кількість графіків, досліджуючи результати при різних значеннях параметрів (довжини гусениці, номерів ГК, обраних для відновлення). Для цього при побудові графіка вся інформація, що відноситься до нього, в скороченому вигляді записується в рядки коментарів до графіка. У той же час, для чистової друку є можливість змінювати підписи, заголовки і коментарі.

При збереженні даних та ж ідея реалізована шляхом збереження разом з даними інформації про те, в результаті яких дій цей ряд був отриманий.

Для того, щоб збережені дані можна було знову обробляти за допомогою програми, є можливість зберігати дані в так званому внутрішньому форматі програми, який дозволяє зберігати текстову інформацію разом з числової, а також кілька рядів однакової довжини в одному файлі.

Для отримання чистових графіків в програмі передбачені три шляхи. Перший - це друк графіка зсередини програми з можливістю змінювати розташування його на аркуші. Другий - копіювання графіка в буфер обміну (це не відноситься до Мультиграфікам) і потім друк його через, наприклад, WORD або EXCEL. Третя можливість - це збереження результатів у вигляді чисел і потім побудова графіка за допомогою будь-якої графічної програми.

3.3. Керівництво користувача

Технічна інформація

Вимоги до комп'ютера:

- Оперативна пам'ять - не менше 2 Мб.
- Процесор - 386 і вище.
- Операційна система - Windows 3.1 і вище.

Необхідно також, щоб серед шрифтів наявної версії Windows був шрифт з кирилицею. Якщо програма не правильно визначить цей шрифт, потрібно використовувати файл ініціалізації. У ньому змінної Name потрібно присвоїти ім'я наявного шрифту, що відображає російські літери.

Вимог до віртуальної пам'яті немає, але чим більше її розмір, тим довше ряд, який програма може аналізувати (незалежно від розміру пам'яті максимальна довжина ряду становить 8000 точок).

Загальні положення

Програма «Гусениця» має головне вікно, яке складається з меню, панелі гарячих кнопок, робочої області і рядки підказок (рис. 3.1).

У процесі роботи програми в робочій області головного вікна з'являються додаткові вікна, що містять графічну інформацію, а в меню додаються нові пункти, відповідні активного вікна. Є три типи додаткових вікон:

- Вікно, що містить один графік.
- Вікно, що містить верхній і нижній графік.
- Вікно, що містить мультиграфік.

Як уже згадувалося, робота програми полягає в послідовному виконанні одного з чотирьох етапів: читання тимчасового ряду з файлу і трьох етапів, які виробляють обчислювальні дії:

1. Перетворення ряду.
2. Розкладання ряду.
3. Відновлення ряду.

В майбутньому передбачається додати етап - Прогнозування ряду.

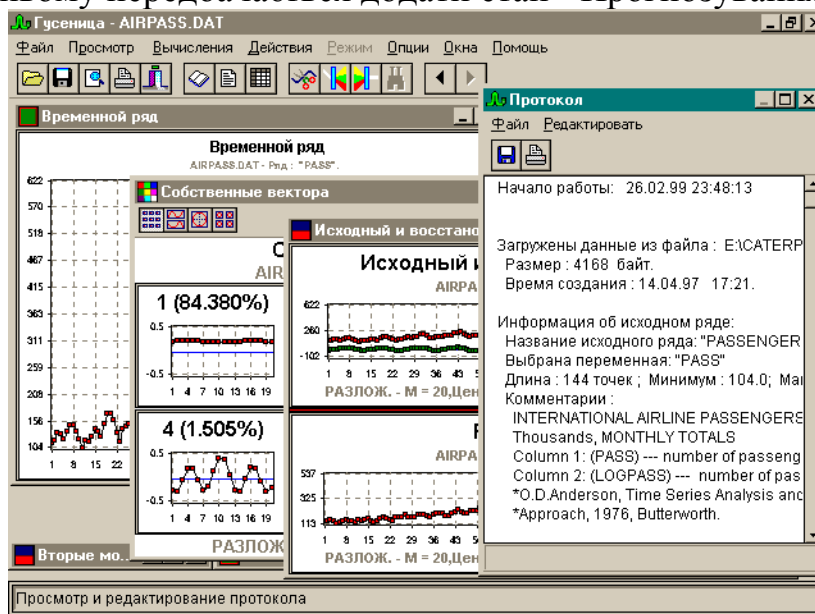


Рисунок 3.1 – Головне вікно програми «Гусениця».

При просуванні по етапах можна пропустити етап перетворення ряду, всі інші - обов'язкові (наприклад, не можна відновлювати ряд, не виконавши його розкладання). Повертатися можна на будь-який етап. При цьому результати наступних етапів, якщо їх не зберегти, будуть втрачені (тобто, якщо після відновлення ряду виконати перетворення, то результати розкладання і відновлення пропадуть).

Перед виконанням кожного етапу, програма пропонує налаштувати його параметри. Після успішного виконання етапу, в головне вікно додаються додаткові вікна (Кількість цих вікон залежить від виконуваного етапу. Наприклад, після розкладання ряду додається п'ять вікон, а після відновлення - одне).

3.4. Робота з графікою

Як було сказано раніше, в програмі існує три типи представлення графіки в додаткових вікнах. При активізації додаткового вікна будь-якого типу в меню головного вікна додається кілька пунктів, за допомогою яких можна змінити опції графіки, зберегти їх або завантажити нові.

В цьому місці необхідно зробити деякий відступ і пояснити, звідки програма бере опції за замовчуванням. Коли виконаний будь-якої етап (див. Вище) і створюється чергове вікно з графікою, програма шукає файл DEFLT.CAT.OPT, що містить опції всіх типів графіки і, якщо він є, завантажує опції з нього. Якщо такого файлу немає, беруться опції, «зашиті» всередину програми. Якщо зберегти нові опції в файл DEFLT.CAT.OPT, то при створенні наступних вікон з графікою будуть використовуватися нові опції (в уже створених вікнах нічого не зміниться). Збереження опцій в файл з іншим ім'ям дасть можливість згодом використовувати їх.

Склад файлу, що зберігає опції, наступний:

- Опції верхнього графіка.
- Опції нижнього графіка.
- Опції двовимірного графіка.
- Опції індивідуальної мультіграфіки.
- Опції парної мультіграфіки.

Слід зазначити, що «Опції верхнього графіка» використовуються в усіх типах вікон:

1. У вікнах, що містять один графік, як опції цього графіка.
2. У вікнах, що містять верхній і нижній графіки, як опції верхнього графіка.
3. У вікнах, що містять мультіграфіки, як опції одновимірної графіки.

Вікна, що містять один графік

У програмі створюються такі представники цього типу:

- Часовий ряд
- Власні числа у відсотках
- Власні числа

Буфер обміну (т.зв. ClipBoard) використовується багатьма програмами як універсальний спосіб передачі різної інформації іншим програмам. Дана програма не є винятком. Вона дозволяє передавати іншим програмам графіки в двох форматах: BitMap (.bmp) і Windows MetaFile (.wmf). З основних відмінностей між ними можна виділити наступні. Графік в форматі Windows MetaFile не спотворюється при стисканні / розтягуванні як BitMap і займає в пам'яті зазвичай

набагато менше місця. У свою чергу, формат BitMap підтримується великою кількістю програм, як більш простий.

Щоб скопіювати графік в буфер обміну, виберіть пункт контекстного меню Копіювати зображення ... і уточніть формат, в якому потрібно передати графік.

Графіки можна зберігати в файл двома способами. Перший - вибрати пункт меню головного вікна Файл | Зберегти зображення Тоді будуть збережені всі графіки активного вікна в форматі BitMap. Другий спосіб зберегти графік - скористатися контекстним меню і його пунктом Зберегти зображення В цьому випадку буде збережений тільки один графік і формат файлу можна вибирати.

Для перегляду і збереження даних існує два способи. За допомогою першого способу можна переглянути і зберегти всі дані, що належать активного вікна (команда меню головного вікна Перегляд | Дані). За допомогою другого - дані одного графіка (через пункт контекстного меню Дані та Зберегти дані ...). При спробі зберегти дані в файл незалежно від способу виникає вікно Параметри збереження в файл. У ньому можна встановити формат, точність виведення значень в файл, кодування символів кирилиці і, якщо обраний внутрішній формат, редагувати коментарі до даних.

Використані джерела:

1. Главные компоненты временных рядов: метод «Гусеница» / Под ред. Д.Л.Данилова и А.А.Жиглявского // Санкт-Петербургский университет. URL: <http://www.gistatgroup.com/gus/book1/index.html>

Розділ 2. Приклади реалізації сучасних методів прогнозування

Тема 4. Приклад реалізації методу "Гусениця" для аналізу одновимірних часових рядів

4.1. Аналіз динаміки виробництва м'яса всіх видів (у живій вазі) в Україні

4.2. Оцінка внеску складових досліджуваного часового ряду

4.3. Прогнозування обсягів виробництва м'яса всіх видів (у живій вазі) в Україні

4.1. Аналіз динаміки виробництва м'яса всіх видів (у живій вазі) в Україні

У якості прикладу, для детального розгляду переваг даного методу проведено аналіз та прогнозування динаміки виробництва м'яса усіх видів (у живій вазі) в Україні. Надалі для скорочення назви ряду опустимо уточнення «у живій вазі». Розрахунки були проведені з використанням програми «Caterpillar на базі часового ряду, що містить в собі помісячні дані за період з січня 2011 до червня 2017 року. Дані за липень, серпень, вересень 2017р. використовувались для параметризації методу та оцінки адекватності моделі. Довжина «гусениці» була обрана кратна 12, а саме $M=24$, оскільки у ряді присутня очевидна річна періодичність.

Для реалізації алгоритму методу «Гусениця» пройдено чотири етапи. На першому етапі проведена розгортка одномірного вихідного ряду, графічне зображення якого наведено на рис.4.1, в багатомірний. Обраний часовий ряд

виробництва м'яса $\{x_i\}_{i=1}^{N=78}$ утворений послідовністю 78 рівновіддалених значень функції $f(t)$: $x_i = f[i] = f((i-1)\Delta t)$, $i = 1, 2, \dots, 78$.

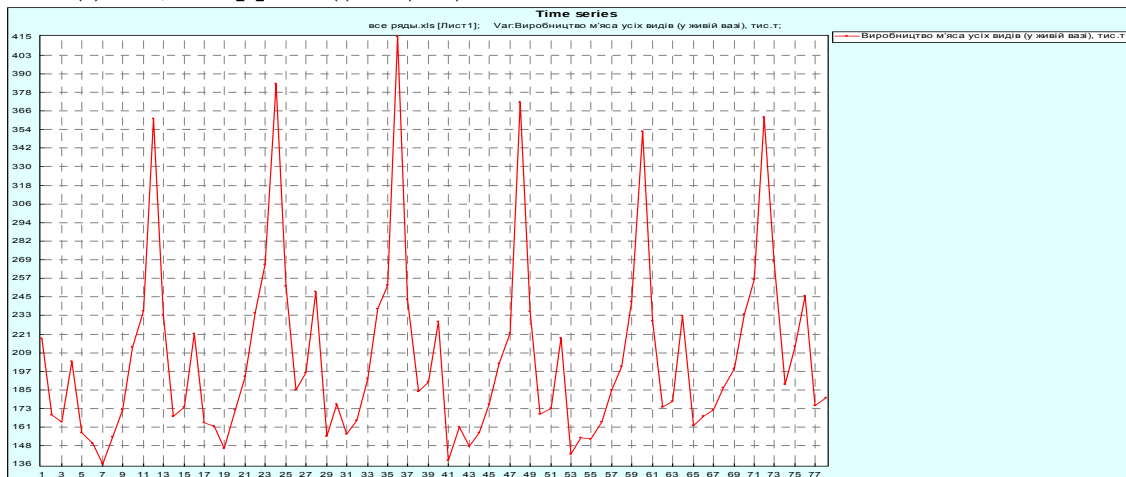


Рис. 4.1. Динаміка виробництва м'яса усіх видів в Україні у січні 2011р. – червні 2017 р.

Перший рядок матриці спостережень X складається з перших $M=24$ значень послідовності f . Другий рядок матриці складається зі значень послідовності від x_2 до $x_{M+1}=x_{25}$ і так далі до останнього рядка із номером $k = N - M + 1 = 78 - 24 + 1 = 55$: від значення x_k до x_N . Елементи матриці дорівнюють $x_{ij} = x_{i+j-1}$.

На другому етапі проведено аналіз головних компонент: сингулярний розклад вибіркової матриці.

Алгоритм реалізації методу передбачає розрахунок матриці других моментів: $R = \frac{1}{k} X^* (X^*)^T$. Слід відмітити, якщо матриця X^* була отримана в результаті центрування та нормування, то матриця R буде представляти собою вибірку кореляційну матрицю з елементами $r_{ij} = \frac{1}{k} \sum_{l=1}^k \frac{1}{s_i s_j} (x_{i+l-1} - \bar{x}_i)(x_{j+l-1} - \bar{x}_j)$.

Кореляційна матриця багатомірного ряду спостережень зображена на рис. 4.2.

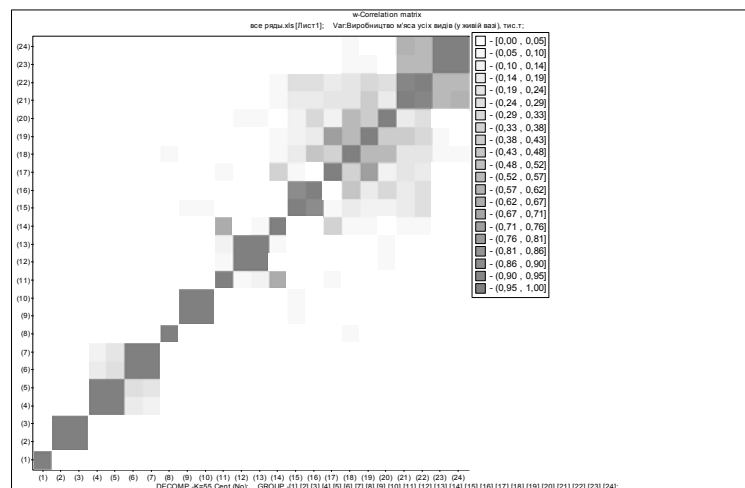


Рис. 4.2. Кореляційна матриця багатомірного ряду спостережень.

Для її зображення обрано візуальний спосіб, у якому більші значення виділені темнішим кольором.

На третьому етапі здійснено вибір головних компонент, що є інформативними для виділення складових часового ряду виробництва м'яса усіх видів в Україні у січні 2011р. – червні 2017 р.

На підставі візуального аналізу діаграм власних функцій і факторних векторів була виявлена структура досліджуваної сукупності, а також проведені вибір та інтерпретація головних компонент.

Для виділення пар власних чисел, що ідентифікують одну періодику, на рис 4.3 приведені логарифми та корні з власних чисел.

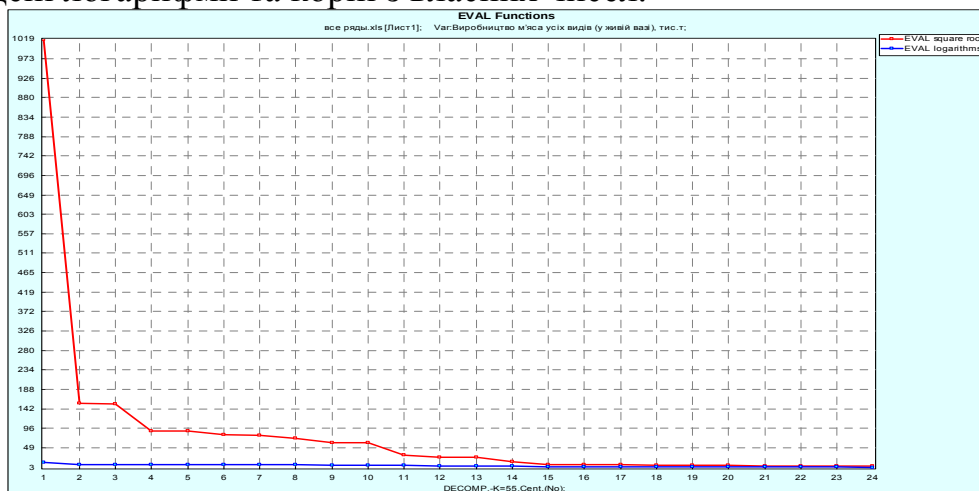


Рис 4.3. Корні з власних чисел та логарифми власних чисел.

На рис. 4.3 помітні «сходишки», що відповідають 2-3, 4-5, 6-7, 9-10 головним компонентам. Це свідчить про те, що кожному з означених пар головних компонент слід віднести до окремої періодики.

Проаналізувавши рис. 4.3 зроблено припущення, що приблизно після 12-ї головної компоненти спостерігається рівномірне убуття дуже малих власних чисел, тобто знаходиться шум. Отже, для відновлення вихідного ряду достатньо перших 12 головних компонент.

Сформульоване припущення підтверджено одномірними графіками головних компонент, що приведені на рис. 4.4.

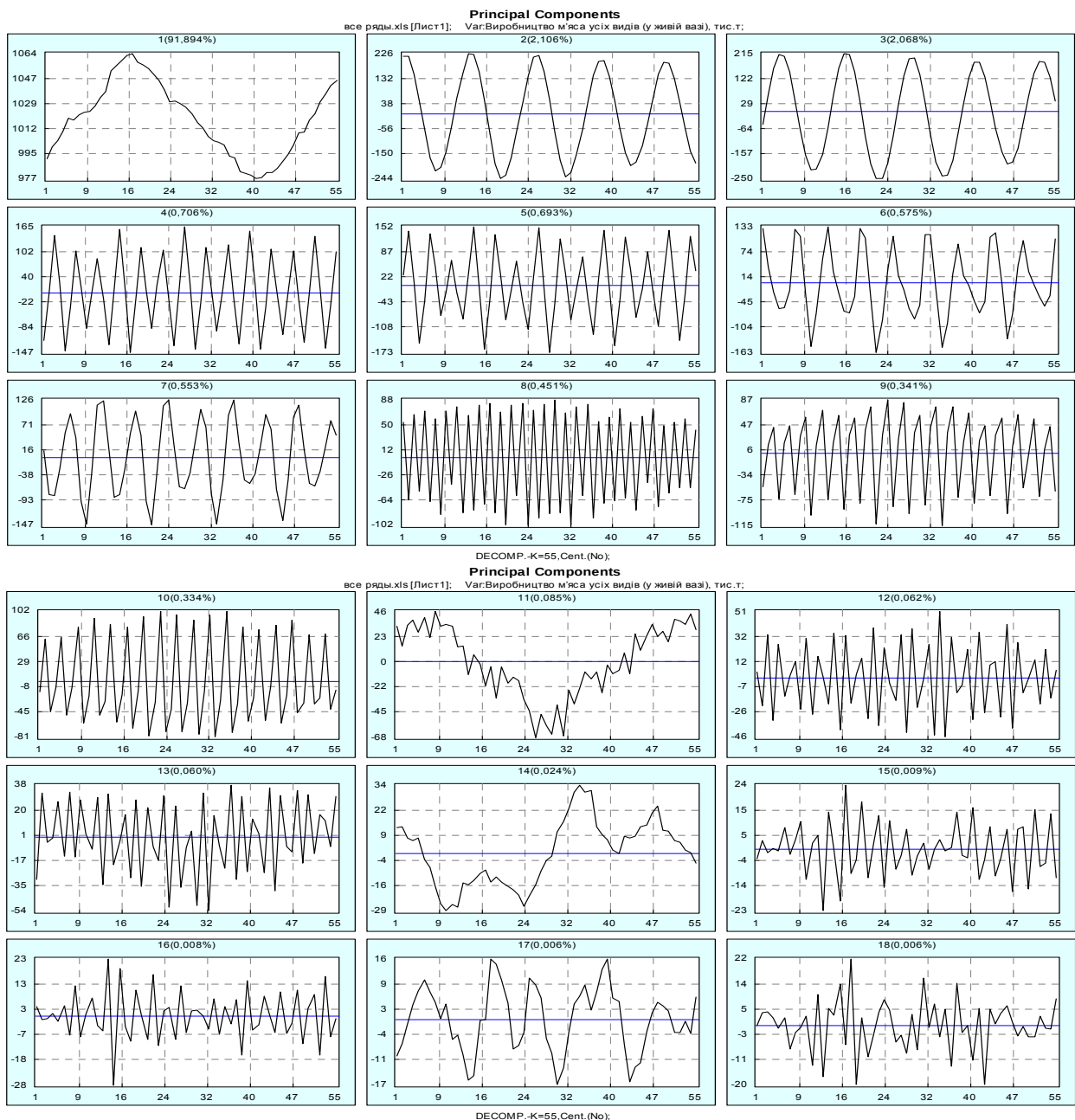
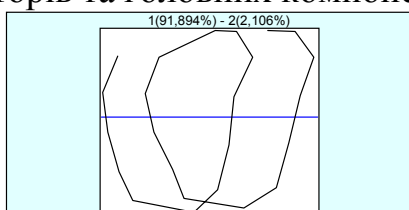


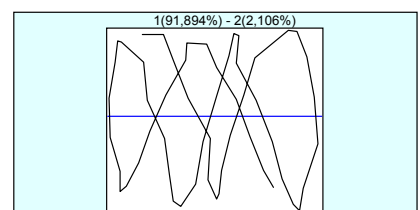
Рис. 4.4. Одномірні графіки головних компонент.

Після проведеного аналізу графіків, які зображені на рис. 3.6, зроблено припущення про те, що перша головна компонента є складовою тренда. Другу та третю головні компоненти можна інтерпретувати як річну періодику. Четверту та п'яту головні компоненти можна інтерпретувати як чотиримісячну періодику. Також можна помітити піврічну та трьохмісячну періодичність (головні компоненти 6-7 та 9-10 відповідно). Для підтвердження зроблених припущень щодо об'єднання головних компонент у пари побудовані двомірні графіки власних векторів та головних компонент (рис. 4.5).

1 та 2
власних
векторів



1 та 2
головної
компоненти



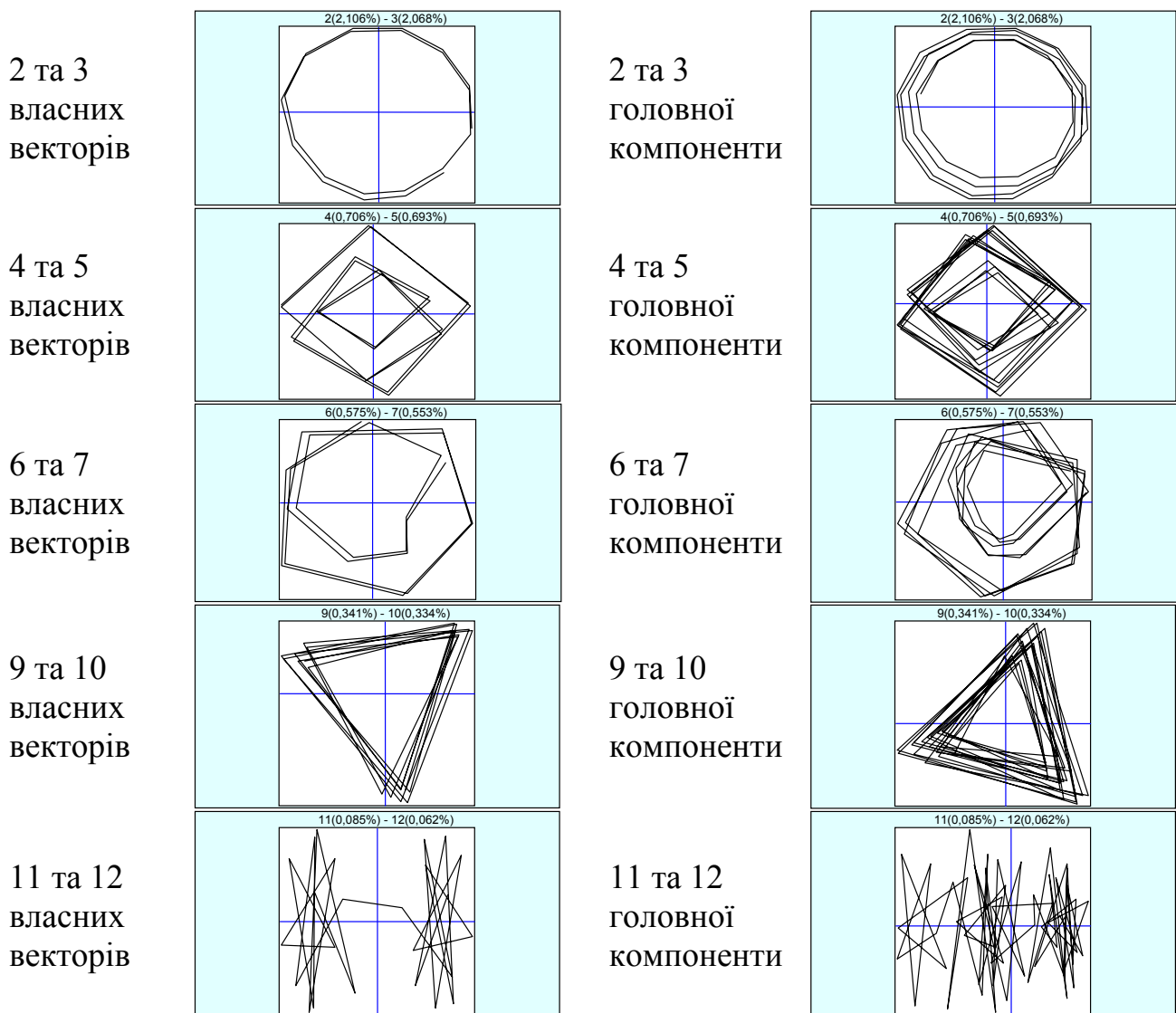


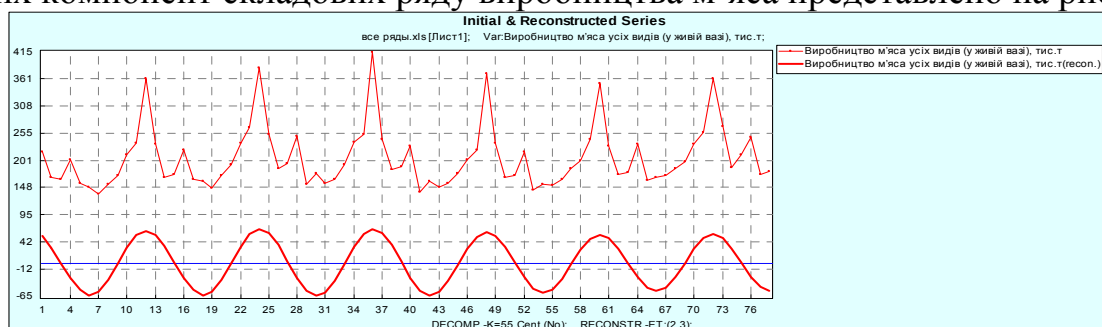
Рис. 4.5. Двомірні графіки власних векторів та головних компонент.

Відомо, що двомірне зображення синуса та косинуса створює одиничне коло. Отже, гармонічна складова із цілим періодом зображується у вигляді правильного багатокутника з числом вершин, відповідним довжині періоду. При зміні амплітуди багатокутник перетворюється у спіраль. Зіркоподібне зображення свідчить про наявність періодики з дрібним періодом.

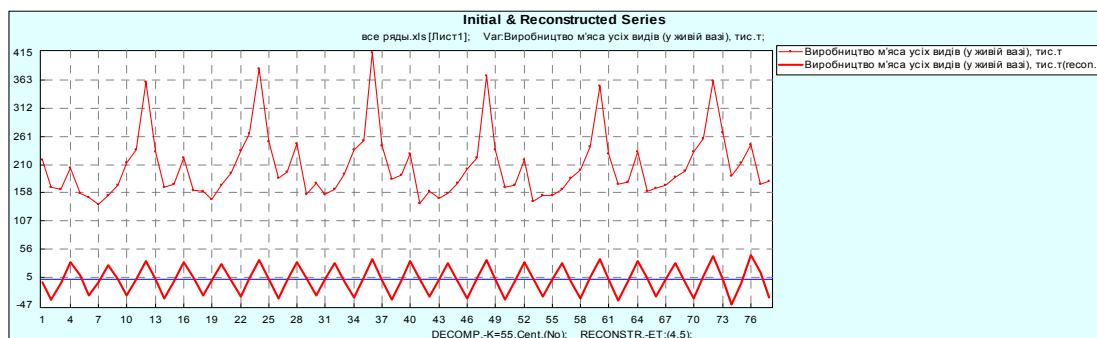
На четвертому етапі проведено відновлення одномірного ряду.

Графічне зображення вихідного та відновлених по виділеним парам головних компонент складових ряду виробництва м'яса представлено на рис. 4.6.

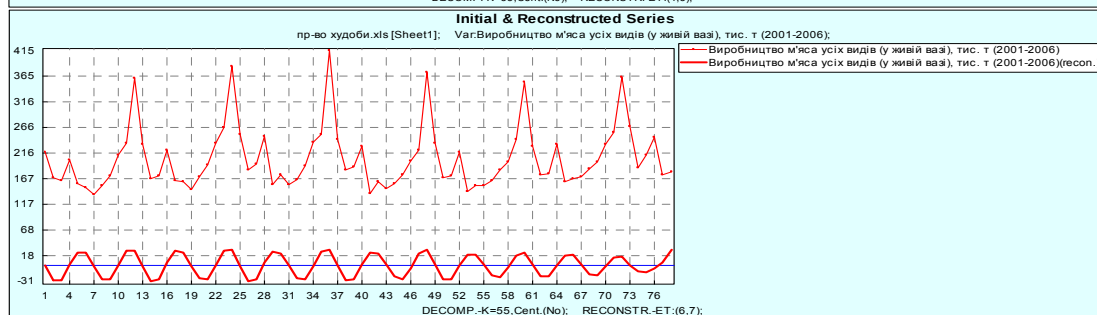
по
2-3-й
ГК



по
4-5-й
ГК



по
6-7-й
ГК



по
9-10-й
ГК

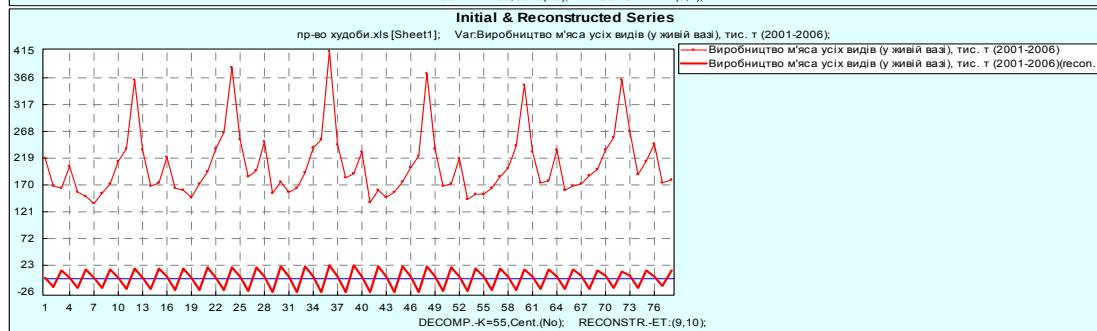


Рис. 4.6. Вихідні та відновлені по виділеним головним компонентам значення часового ряду виробництва м'яса усіх видів в Україні в січні 2011р. - червні 2017р.

Підсумки аналізу та інтерпретація головних компонент представлені у табл. 4.1.

Таблиця 4.1 – Інтерпретація головних компонент

№№ головних компонент	Інтерпретація
1	Нелінійний тренд
2-3	Річна періодика
4-5	Чотиримісячна періодика
6-7	Піврічна періодика
9-10	Тримісячна періодика

Підкреслимо той факт, що наявність періодик з кратними періодами означає, що у даному випадку, аналізуючи часовий ряд виробництва м'яса в Україні, було знайдено річну періодику і її розклад на синусоїдальні складові.

Часові ряди, що характеризують динаміку елементів економічного, внутрішнього виробничого потенціалу та цін основних видів тваринницької продукції аграрної галузі, після застосування методу «Гусениця» були розкладені на адитивні складові (табл. 4.2).

Таблиця 4.2 – Виділені адитивні складові часових рядів

Часовий ряд	Виділені складові											
	тренд	4-річ-на	3-річ-на	2-річ-на	річ-на	пів-річ-на	9-мі-сяч-на	8-мі-сяч-на	4-мі-сяч-на	3-мі-сяч-на	2,5-мі-сячна	інші
	відповідні головні компоненти											
Поголів'я овець та кіз	1				2-3	4-5, 6-7				8-9		10,11,12
Поголів'я ВРХ	1		4-5		2-3	6-7				9-10		8,11,12
Поголів'я свиней	1	2-3			4-5	6-7			8-9	11-12		10
Виробництво молока та молочних продуктів	1				2-3	4-5				9-10	6-7	8,11,12
Виробництво м'яса усіх видів	1				2-3	6-7			4-5	9-10		8,11,12
Виробництво яєць	1				2-3	4-5			6-7		9-10	8,11,12
Середні ціни на молоко та молочні продукти	1			4-5	2-3			6-7	8-9			10-14
Середні ціни на худобу та птицю (у живій вазі)	1	2-3			4-5	6-7				8-9		10,11,12
Середні ціни на ВРХ	1		2-3			6-7	4-5			8-9		10,11
Середні ціни на птицю	1		2-3				4-5			7-8	10-11	6,9,12, 13,14
Середні ціни на свиней	1		2-3		4-5				6-7	8-9, 10-11		12,13,14
Середні ціни на яйця	1			4-5	2-3			10-11	8-9	6-7		12-20

Проведене розкладення одномірних часових рядів, що характеризують елементи кон'юнктури аграрного ринку України, свідчить про наявність у кожному ряді періодичної складової. Циклічні складові (4-річна, 3-річна, 2-річна) присутні у рядах динаміки поголів'я ВРХ та свиней, а також у рядах цін на м'ясо, яйця, молоко та молочні продукти.

4.2. Оцінка внеску складових досліджуваного часового ряду

Для оцінки внеску кожної складової рівня ряду необхідно розкласти ряд на складові таким чином, щоб кожна складова була додатною. Загальновживаним за наявності періодичних коливань часових рядів є використання моделей сезонної хвилі на основі гармонійного аналізу. Оцінкою вкладу гармонік, що характеризують сезонні хвилі з урахуванням розкладення на ряд Фур'є, є дисперсійне відношення.

Недостатньо характеризувати внесок коливань сезонної складової лише в загальну дисперсію процесу, тому що це положення більш коректно використовувати для стаціонарних процесів. Для нестаціонарних процесів, якими в основному є часові ряди, що характеризують різні елементи ринкової системи, це положення трактується неоднозначно.

Доцільною є побудова адитивної моделі часових рядів у вигляді суми відповідних складових. Цю модель можна зобразити таким чином:

$$Y_t = U_t + V_t + \sum_{i=1}^m S_{ti} + \xi_t, \quad (3.2)$$

де Y_t – рівні часового ряду;

U_t – трендова складова; V_t – циклічна складова;

$\sum_{i=1}^m S_{ti}$ – сума m сезонних складових; ξ_t – випадкова складова; t – час.

Запропоновано перенести систему координат таким чином, щоб мінімальне значення кожної сезонної складової дорівнювало нулю. При переносі системи координат для кожної з сезонних складових виникає різниця, яка дорівнює відхиленню найменшого від’ємного значення складової від нуля. Для збереження тотожності суму модулів цих різниць необхідно додати до значень ряду, відновленого по обраним компонентам.

З огляду на викладене вище запропоновано наступний алгоритм виділення окремих складових вихідного ряду та оцінки внеску кожної складової:

- 1) реалізація методу Singular Spectrum Analysis (SSA);
- 2) аналіз головних компонент та їх інтерпретація;
- 3) виділення окремих складових вихідного ряду;
- 4) відновлення значень окремих складових у відповідності з виділеними головними компонентами;
- 5) знаходження мінімальних значень кожної з сезонних складових;
- 6) додавання модуля цих значень до відповідного рівня відновленого по складовій ряду;
- 7) розрахунок модуля суми мінімальних значень всіх сезонних складових;
- 8) збільшення відповідних рівнів ряду, відновленого по обраним головним компонентам, на модуль суми мінімальних значень всіх сезонних складових;
- 9) розрахунок частки складових у загальному значенні відновленого ряду.

З урахуванням запропонованого алгоритму та формули (3.2) ряд представлено у наступному вигляді:

$$\hat{Y}_t + \sum_{i=1}^m |\min S_{ti}| + |\min V_t| = U_t + V_t + |\min V_t| + \sum_{i=1}^m (S_{ti} + |\min S_{ti}|) + \xi_t.$$

Внесок i -ї сезонної складової у t -й рівень ряду дорівнює:

$$w_{S_{ti}} = \frac{S_{ti} + |\min S_{ti}|}{\hat{Y}_t + \sum_{i=1}^m |\min S_{ti}| + |\min V_t|}.$$

Середній внесок i -ї сезонної складової дорівнює:

$$\overline{w_{S_i}} = \frac{\sum_{t=1}^n \frac{S_{ti} + |\min S_{ti}|}{\hat{Y}_t + \sum_{i=1}^m |\min S_{ti}| + |\min V_t|}}{n}.$$

Середній внесок трендового компоненту дорівнює:

$$\overline{w_U} = \frac{\sum_{t=1}^n \frac{U_t}{\hat{Y}_t + \sum_{i=1}^m |\min S_{ti}| + |\min V_t|}}{n}.$$

Середній внесок циклічної складової дорівнює:

$$\overline{w_V} = \frac{\sum_{t=1}^n \frac{V_t + |\min V_t|}{\hat{Y}_t + \sum_{i=1}^m |\min S_{ti}| + |\min V_t|}}{n}.$$

Розрахована структура складових часових рядів виробництва м'яса усіх видів, а також інших продуктів аграрної галузі та поголів'я сільськогосподарських тварин в Україні, наведена у табл. 4.3.

Таблиця 4.3

Структура складових часових рядів

Часовий ряд	Внесок складових, %											
	трен- ду	4- річ- ної	3- річ- ної	2- річ- ної	річ- ної	пів- річ- ної	9-мі- сяч- ної	8-мі- сяч- ної	4-мі- сяч- ної	3-мі- сяч- ної	2,5-мі- сяч- ної	ін- ших
Поголів'я овець та кіз	82,0				7,0	9,0				0,9		1,1
Поголів'я ВРХ	86,0		2,9		8,2	1,9				0,4		0,6
Поголів'я свиней	81,7	11,6			4,2	1,3			0,6	0,4		0,2
Виробництво молока та молочних продуктів	64,4				24,8	4,3				1,7	2,5	2,3
Виробництво м'яса усіх видів	51,2				14,9	7,5			11,1	6,2		9,1
Виробництво яєць	67,6				17,0	5,3			4,5		2,4	3,2
Середні ціни на молоко та молочні продукти	71,1			5,3	15,7			3,0	2,1			2,9
Середні ціни на худобу та птицю (у живій вазі)	80,6	8,2			4,2	2,1				2,1		2,8
Середні ціни на ВРХ	82,3		9,3			2,1	3,2			1,7		1,4
Середні ціни на птицю	76,1		6,8				3,9			3,9	3,3	6,0
Середні ціни на свиней	73,6		13,0		4,6				2,9	3,0		2,9
Середні ціни на яйця	59,5			8,3	14,1			3,3	2,6	7,5		4,7

Аналізуючи одержані результати, можна зробити висновки про великий внесок сезонних складових у рівні всіх розглянутих часових рядів.

Аналіз вихідних одномірних часових рядів елементів економічного, виробничого потенціалу та цін продукції ринку тваринництва, перетворення їх в багатомірні ряди, проведення їх дослідження із застосуванням методу головних

компонент та подальше відтворення одномірного ряду дають можливість моделювання виділених окремих складових вихідних рядів.

4.3. Прогнозування обсягів виробництва м'яса всіх видів (у живій вазі) в Україні

Проведене дослідження окремих складових часових рядів, таких як повільний тренд, періодичні та сезонні коливання свідчить про необхідність та доцільність застосування їх статистичної оцінки при прогнозуванні кон'юнктури аграрного ринку в Україні.

Для визначення міри точності одержаного прогнозу використана статистична оцінка:

$$\text{MAPE} = \frac{100}{n} \sum_t \left| \frac{y_t - \hat{y}_t}{y_t} \right|,$$

де $\hat{y}_t$ – прогнозне значення часового ряду у t -му періоді.

Розраховані прогнозні значення в липні-листопаді 2017 р. були порівняні з оновленими офіційними статистичними даними (табл. 4.4).

Розраховані статистичні оцінки MAPE дозволяють зробити висновок, що побудовані за допомогою даного методу моделі є високо адекватними.

Достовірність прогнозу також була підтверджена розрахунком коефіцієнта Генрі Тейла, так званого коефіцієнта невідповідності (розбіжності), згідно з формулою (1.1). Для ряду виробництва м'яса усіх видів в Україні було отримано значення цього коефіцієнта $U=0,0515$. Як бачимо, це значення близьке до нуля. За теорією Г.Тейла це означає, що прогноз майже ідеальний.

Таким чином, розрахований коефіцієнт невідповідності підтвердив висновок про дуже високу точність одержаних прогнозних значень.

Розрахована статистична оцінка адекватності й прогностичних властивостей одержаних моделей дозволила обґрунтувати висновок, що при короткостроковому прогнозуванні кон'юнктури аграрного ринку високий рівень достовірності прогнозу можна одержати застосовуючи метод «Гусениця» майже для всіх часових рядів.

Можна стверджувати, що метод аналізу сингулярного спектру обґрунтовано займе почесне місце серед інструментів аналізу та прогнозу часових рядів.

Аналізуючи статистичні оцінки MAPE, зауважимо, що найбільші розбіжності між прогнозними та фактичними значеннями притаманні цінам на яйця. Зважаючи на це, для підвищення прогностичних властивостей моделей особливу увагу автором приділено з'ясуванню закономірностей одночасних змін взаємопов'язаних рядів, таких як ціни та обсяги пропозиції. Для цього метод одномірного SSA було поширено на випадок декількох часових рядів. У даному розумінні очікуваним результатом є одночасне розкладання декількох рядів на погоджені компоненти.

Таблиця 4.4 – Прогнозні та фактичні значення часових рядів у липні – листопаді 2017р.

Часовий період	Елементи економічного потенціалу								
	Поголів'я ВРХ			Поголів'я свиней			Поголів'я овець та кіз		
	Значення ряду, тис. гол.		Помилка прогнозу, у, %	Значення ряду, тис. гол.		Помилка прогнозу, у, %	Значення ряду, тис. гол.		Помилка прогнозу, у, %
	прогнознi	фактичнi		прогнознi	фактичнi		прогнознi	фактичнi	
Липень	6691,8	6762,7	1,048	8487,9	8445,5	0,502	2074,8	2048,8	1,269
Серпень	6499,1	6672,2	2,594	8530,0	8375,5	1,845	2014,2	2016,0	0,089
Вересень	6388,8	6540,9	2,325	8519,8	8281,7	2,875	1956,8	1966,2	0,478
Жовтень	6305,2	6399,0	1,466	8461,1	8250,8	2,549	1956,2	1908,9	2,478
Листопад	6055,1	6223,6	2,707	8374,6	8096,5	3,435	1867,0	1853,9	0,707
МАРЕ			2,028			2,241			1,004
Часовий період	Внутрішній виробничий потенціал								
	Виробництво м'яса усіх видів (у живій вазі)			Виробництво молока			Виробництво яєць		
	Значення ряду, тис. т		Помилка прогнозу, %	Значення ряду, тис. т		Помилка прогнозу, у, %	Значення ряду, млн. шт.		Помилка прогнозу, у, %
	прогнознi	фактичнi		прогнознi	фактичнi		прогнознi	фактичнi	
Липень	184,0	189,6	2,980	1392,7	1364,4	2,074	1158,8	1244,6	6,897
Серпень	213,0	200,9	6,013	1354,8	1277,3	6,067	1260,3	1226,9	2,726
Вересень	216,1	211,4	2,237	1177,1	1109,5	6,093	1148,0	1106,6	3,740
Жовтень	239,5	256,1	6,486	1116,4	997,2	11,953	1159,2	1075,3	7,801
Листопад	280,6	266,2	5,398	835,2	827,4	0,943	1072,9	984,0	9,039
МАРЕ			4,623			5,426			6,041
Часовий період	Середньомісячні ціни								
	худоби та птиці (у живій вазі)			молока та молочних продуктів			яєць		
	Значення ряду, грн./т		Помилка прогнозу, %	Значення ряду, грн./т		Помилка прогнозу, у, %	Значення ряду, грн./тис. шт.		Помилка прогнозу, у, %
	прогнознi	фактичнi		прогнознi	фактичнi		прогнознi	фактичнi	
Липень	6434,0	6068,1	6,031	1539,7	1471,2	4,654	224,8	288,4	22,048
Серпень	6767,5	7011,5	3,480	1557,3	1573,5	1,028	244,7	257,0	4,782
Вересень	6765,4	7536,2	10,228	1721,8	1777,1	3,112	247,6	311,0	20,373
Жовтень	6744,9	6720,1	0,369	1678,4	2032,2	17,410	295,2	361,0	18,217
Листопад	6947,9	6995,6	0,682	1741,9	2191,1	20,499	312,1	360,0	13,308
МАРЕ			4,158			9,341			15,745

Використані джерела:

1. Главные компоненты временных рядов: метод «Гусеница» / Под ред. Д.Л.Данилова и А.А.Жиглявского // Санкт-Петербургский университет. URL: <http://www.gistatgroup.com/gus/book1/index.html>
2. Голяндина Н. Э. Метод «Гусеница»-SSA : прогноз временных рядов : Учеб. пособие. – СПб : Изд-во СПбГУ, 2004. – 52 с.
3. Корепанов О. С. Статистичний аналіз тенденцій зовнішньоекономічної діяльності України: сектор ІКТ // Соціальна економіка. 2018. № 1, Вип. 55. С. 18–26.
4. Корепанов О. С., Прус Ю.І., Чала Т.Г. Короткострокове прогнозування пропозиції робочої сили на ринку праці України // Бізнес Інформ. 2017. № 4. С. 216–222.

Тема 5. Визначення наявності тренда

5.1. Сутність процедури вирівнювання динамічних рядів та визначення тренду

5.2. Побудова трендових моделей МНК

5.3. Екстраполяція тренда

5.1. Сутність процедури вирівнювання динамічних рядів та визначення тренда

Важливою складовою динамічних процесів є тенденція середньої, тобто основний напрям розвитку. В аналізі динамічних рядів тенденцію представляють у вигляді плавної траєкторії та описують певною функцією, яку називають *трендом* $Y_t = f(t)$, де $t = 1, 2, \dots, n$ — змінна часу. На основі такої функції здійснюється вирівнювання динамічного ряду і прогнозування подальшого розвитку процесу.

Процедура вирівнювання динамічних рядів включає два етапи: обґрунтування (вибір) типу функції, яка б адекватно описувала характер динаміки, та оцінювання параметрів функції. На практиці переважно використовують функції, параметри яких мають конкретну інтерпретацію залежно від характеру динаміки. Найбільш поширені поліноми (многочлени), різного роду експоненти та логістичні криві. Так, параметри полінома p -го ступеня $Y_t = a + bt + ct^2 + dt^3 \dots$ характеризують:

a — рівень динамічного ряду при $t = 0$;

b — абсолютну швидкість зміни рівнів ряду (ординат);

$2c$ — прискорення (прирошення абсолютної швидкості);

d — зміну прирощення тощо.

Поліном 1-го ступеня, тобто лінійний тренд $Y_t = a + bt$, описує процеси, які рівномірно змінюються в часі і мають стабільні прирости ординат. Поліном 2-го ступеня (парабола) $Y_t = a + bt + ct^2$ здатний описати процес, характерною особливістю якого є рівноприскорене зростання або зменшення ординат. Форма параболи визначається параметром c : при $c > 0$ гілки параболи спрямовані вгору — парабола має мінімум, при $c < 0$ гілки параболи спрямовані вниз — парабола має максимум. При визначенні екстремуму (max, min) похідну параболи прирівнюють до нуля і розв'язують систему рівнянь відносно t .

Наприклад, динаміка захворювань при епідемії грипу (чол.) описується параболою

$$Y_t = 264 + 45t - 1,5t^2.$$

Похідна параболі $45 - 2,25t = 0$, а $t = 20$. Максимум захворювань буде зафіксовано через 20 днів від початку відліку часу ($t = 0$) і становитиме

$$Y_{\max} = 264 + 45 \cdot 20 - 1,5 \cdot 20^2 = 564 \text{ чол.}$$

У полінома 3-го ступеня $Y_t = a + bt + ct^2 + dt^3$ знак прирощення ординати може змінюватися один чи два рази.

Якщо характерною властивістю процесу є стабільна відносна швидкість (темпи приросту), такий процес описується експонентою, яка може набувати різних еквівалентних форм. Основна (показникова) форма експоненти

$$Y_t = ab^t,$$

де b — середня відносна швидкість зміни ординати: при $b > 1$ ордината зростає з постійним темпом, при $b < 1$, навпаки, зменшується. Абсолютний приріст пропорційний досягнутому рівню.

Експоненту можна представити у формі

$$Y_t = ae^{\lambda t} \text{ або } Y_t = e^{a + bt},$$

де $\lambda = \ln b$, $e = 2,718$ — основа натурального логарифма, $\ln e = 1$.

Експоненти приводяться до лінійного виду заміною y_t десятковими або натуральними логарифмами:

$$\lg Y = \lg a + t \lg b,$$

$$\ln Y = \ln a + \lambda t \ln e = \ln a + \lambda t,$$

$$\ln Y = \ln e^a + \ln e^{bt} = \ln a + \ln bt = \ln a + \lambda t.$$

5.2. Побудова трендових моделей МНК

Оцінювання параметрів трендових рівнянь найчастіше здійснюється *методом найменших квадратів* (МНК), основною умовою якого є мінімізація суми квадратів відхилень фактичних значень y_t від теоретичних Y_t , визначених за трендовим рівнянням

$$\sum_1^n (y_t - Y_t)^2 = \min.$$

Параметри поліноміального тренда визначаються безпосередньо розв'язуванням систем $p + 1$ нормальних рівнянь. Експонента, як показано вище, приводиться до лінійного виду логарифмуванням; розраховані параметри підлягають потенціюванню.

Побудова трендових моделей МНК є ефективною в модулі *Multiple Regression*. Стартова панель модуля дає можливість відкрити необхідний файл даних і вибрати ознаки для аналізу, а за необхідності трансформувати значення цих ознак.

Як приклад, розглянемо динаміку обороту універсальної біржі VAR1 за 18 місяців.

Інтенсивне нарощення біржового обороту можна описати експонентою. Для цього обсяги біржового обороту через опцію *Current Specs* (див. 2.1) замінимо десятковими логарифмами.

У модулі *Multiple Regression* здійснимо селекцію ознак: незалежна — порядковий номер місяця t , залежна — $\lg y_t$ (LOGV1). Після команди на виконання програми *OK* вибираємо опцію *Regression summary*. Вихідні дані та параметри моделі наведено в табл. 5.1. та 5.2.

За даними розрахунку експонента має вигляд $\lg Y_t = 1,867 + 0,0229t$, стандартна похибка $s_e = 0,00585$. Після потенціювання $Y_t = 70,6 \cdot 1,054^t$. Тобто, біржовий оборот щомісячно зростає у середньому на 5,4%, $s_e = 0,0386$.

Таблиця 5.1 – Вихідні дані

Порядковий номер місяця, t	Біржовий оборот, млн. грн., VAR1	LOG V1
1	78,4	1,894316
2	81,2	1,909556
3	85,6	1,932474
4	89,4	1,949878
5	93,8	1,972203
6	102,5	2,011147
7	106,2	2,026125
8	112,8	2,052309
9	120,4	2,080626
10	126,3	2,101403
11	130,7	2,116276
12	140,9	2,148911
13	146,7	2,166430
14	153,6	2,186391
15	163,5	2,213518
16	170,6	2,231979
17	179,5	2,254064
18	185,2	2,267641

Таблиця 5.2 – Параметри моделі

Regression Summary for Dependent Variable: LOGV1						
Continue...	R = ,9989 RI = ,9978 Adjusted RI = ,9977					
	F (1,16) = 7384,6 p < ,000 Std. Error of estimate: ,00585					
N = 18	BETA	St. Err. of BETA	B	St. Err. of B	t(16)	p-level
Intercept			1,8670	0,00287	648,5	8,63E-37
LOGV1	0,9989	0,0116	0,0229	0,00027	85,9	9,38E-23

5.3. Екстраполяція тренда

Виявлену тенденцію можна продовжити за межі динамічного ряду. Така процедура називається *екстраполяцією* тренда. Принципова можливість екстраполяції ґрунтується на припущенні, що умови, які визначали тенденцію у минулому, не зазнають істотних змін у майбутньому. Формально операцію екстраполяції можна представити як визначення функції

$$Y_{t+v} = f(Y_t^*, v),$$

де Y_{t+v} — прогнозне значення на період упередження v ;

Y_t^* — база екстраполяції, найчастіше це останній, визначений за трендом рівень ряду.

Якщо, скажімо, в червні ($t_n = 18$) теоретичний рівень біржового обороту становив (млн. грн): $Y_t = 70,6 \cdot 1,054^{18} = 181,9$, то в липні можна очікувати $Y_{18+1} = 181,9 \cdot 1,054 = 191,7$.

У модулі *Multiple Regression* екстраполяція здійснюється за опцією *Predict dependent var.* Для цього у вікні *Specify dependent for indep. vars.* треба вказати значення v .

Екстраполяція тренда дає точковий прогноз. Очевидно, що «влучення в точку» малоімовірне. Адже тренду властива невизначеність, передусім через похибки параметрів. Джерелом цих похибок є обмежена сукупність спостережень y_t , кожне з яких містить випадкову компоненту e_t . Зсунення періоду спостереження лише на один крок веде до зсунення оцінок параметрів. Випадкова компонента буде присутня і за межами динамічного ряду, а отже, її необхідно врахувати. Для цього визначають довірчий інтервал, який би з певною ймовірністю окреслив межі можливих значень Y_{t+v} . Точковий інтервал перетворюється в інтервальний. Ширина інтервалу залежить від варіації рівнів динамічного ряду навколо тренда та ймовірності висновку $(1 - \alpha)$:

$$Y_{t+v} \pm t_{1-\alpha} s_p,$$

де s_p — середня квадратична похибка прогнозу, значення якої залежить від дисперсії тренда s_Y^2 та дисперсії відхилень від тренда s_e^2 .

Зокрема, для лінійного тренда

$$s_Y^2 = \frac{s_e^2}{n} + \frac{s_e^2}{\sum (t - \bar{t})^2} (t - \bar{t})^2.$$

Якщо база прогнозування — останній рівень ряду, то $\sum (t - \bar{t})^2 = \frac{n(n^2 - 1)}{12}$, а $(t - \bar{t})^2$ замінюється на $(2v + n - 1)^2$. Після нескладних алгебраїчних перетворень похибку прогнозу за лінійним трендом можна представити так:

$$s_p = s_e \sqrt{1 + \frac{1}{n} + \frac{3(2v + n - 1)^2}{n(n^2 - 1)}}$$

або, позначивши підкореневий вираз символом z , $s_p = s_e z$.

Тобто похибка прогнозу залежить від залишкової дисперсії s_e^2 , довжини динамічного ряду (передісторії) n та періоду упередження v . Чим довший період передісторії, тим похибка менша, а збільшення періоду упередження, навпаки, веде до зростання похибки прогнозу.

При визначенні інтервального прогнозу для лінійного тренда (експоненти) можна скористатися таблицями Є. Четиркіна. У додатку 6 наведено значення $Z^* = t_{0,90} z$ для $n = 7 \dots 25$ та $v = 1 \dots 5$. Як бачимо, вони прямо пропорційні періоду упередження v та обернено пропорційні довжині динамічного ряду n .

Скористаємося значеннями Z^* для визначення довірчих меж прогнозу біржового обороту. При $n = 18$ і $v = 1$ значення $Z^* = 1,9455$. Якщо $s_e = 0,0384$, то похибка прогнозу біржового обороту становить $s_p = 0,0384 \cdot 1,9455 = 0,075$, а довірчі межі прогнозу на липень $191,7 \pm 0,075$. Отже, з імовірністю 0,90 можна

стверджувати, що в липні оборот біржі буде щонайменше 191,6 млн. і не перевищить 191,8 млн. грн.

Рекомендована література

1. Єріна А. М., Єрін Д. Л. Статистичне моделювання та прогнозування : навч. посіб. К. : КНЕУ, 2014. 348 с.
2. Корепанов О. С. Статистичний аналіз тенденцій зовнішньоекономічної діяльності України: сектор ІКТ // Соціальна економіка. 2018. № 1, Вип. 55. С. 18–26.
3. Hyndman R. J. Forecasting: principles and practice / R. J. Hyndman, G. Athanasopoulos. – Otexts, 2018. – 504 p.
4. Ord, J. K., Fildes, R., & Kourentzes, N. Principles of business forecasting (2nd ed.). Wessex Press Publishing Co. 2017.

Тема 6. Статистичний аналіз сезонних та циклічних коливань

6.1. Індекси сезонності

6.2. Сезонно-декомпозиційна модель Холта-Вінтера

6.3. Моделювання сезонної хвилі на основі гармонійного аналізу

6.1. Індекси сезонності

Сезонні коливання формуються під впливом не лише природно-кліматичних, але й соціально-економічних факторів. Сила і напрям дії окремих факторів формує різну конфігурацію сезонної хвилі. За своїм характером сезонна компонента може бути адитивною або мультиплікативною. Для адитивної компоненти характерні сталі коливання навколо середнього рівня чи тренда, для мультиплікативної — зростання амплітуди коливань з часом.

Кожний рівень ряду y_t належить до певного сезонного циклу s , довжина якого становить 12 місяців, або 4 квартали. Відношення y_t до середнього рівня за цикл називається *індексом сезонності*:

$$I_s = \frac{y_t}{\bar{y}}.$$

За умови, що вплив несезонних факторів еліміновано, середня з індексів j -го циклу становить 1, або 100 %.

У нестационарних рядах замість середньої використовують лінію тренда $Y_t = y(t)$, яка плавно проходить через ряд динаміки i , як і середня, елімінує його нерівномірності. Сукупність індексів сезонності в межах циклу характеризує сезонний ритм.

Прогнозування сезонних процесів ґрунтується на декомпозиції динамічного ряду. Припускають, що у майбутньому збережеться тенденція і такий же характер коливань. За таких умов прогноз на будь-який місяць (квартал), визначений методом екстраполяції тренда, коригується індексом сезонності: $Y_{t+v}^* = I_t \cdot Y_{t+v}$, де v — період упередження. Скажімо, поквартальна динаміка обсягів імпорту пального (тис. т) за два роки ($n = 8$, $t_1 = -3,5$, $t_n = 3,5$) описується трендом $Y_t = 923,7 + 33,8t$, за яким теоретичний обсяг імпорту у восьмому кварталі становить 1042,0 тис. т, а в 1-му кварталі наступного року ($v = 1$) передбачається

$$Y_{t+v} = 1042,0 + 33,8 \cdot 1 = 1075,8.$$

Якщо середній індекс сезонності 1-го кварталу $I_1 = 1,34$, то скоригований на сезонність прогнозований рівень дорівнює $Y_{t+1}^* = 1,34 \cdot 1075,8 = 1441,6$ тис. т.

Динаміка більшості показників не виявляє чітко вираженої тенденції розвитку. Через постійний перерозподіл впливу факторів, які формують динаміку процесу, змінюється інтенсивність динаміки, частота та амплітуда коливань. До таких фактичних даних більш еластичною виявляється ковзна середня, інтервал згладжування якої дорівнює сезонному циклу (4 або 12). Коригування ковзної середньої на сезонність здійснюється так само, як коригування лінійного тренда.

6.2. Сезонно-декомпозиційна модель Холта-Вінтера

На використанні експоненційної середньої *ґрунтується сезонно-декомпозиційна модель Холта-Вінтера*, в якій поєднуються моделі стаціонарності, лінійності та сезонності. Послідовність операцій така:

1. Визначаються індекси сезонності I_t .
2. Ряд динаміки фільтрується від сезонних коливань діленням y_t на коефіцієнт сезонності з лагом s ; ряд $u_t = y_t : I_{t-s}$ називається декомпозиційним.
3. Перші різниці декомпозиційного ряду $b_t = (u_t - u_{t-1})$ розглядаються як характеристики лінійного тренда.

Кожна з компонент моделі згладжується за допомогою експоненційної середньої. При комбінації лінійної та сезонно-адитивної моделей тренда:

$$u_t = A \frac{y_t}{I_{t-s}} + (1 - A)(u_{t-1} + b_{t-1});$$

$$b_t = B(u_t - u_{t-1}) + (1 - B)b_{t-1};$$

$$I_t = C \frac{y_t}{u_t} + (1 - C)I_{t-s}.$$

Значення параметрів згладжування A (*Alpha*), B (*Delta*) і C (*Gamma*) в системі Statistica за умовчужання визначаються на рівні 0,1, в [9] рекомендуються: $A = 0,2$; $B = 0,2$; $C = 0,5$.

За умови ізольованої оцінки трьох факторів прогноз на період упередження v визначається як скоригована на сезонність сума прогнозного рівня u_t і лінійного тренда:

$$Y_{t+v}^* = (u_t + b_t v) I_{t-s+v}.$$

При комбінації лінійного та сезонно-мультиплікативного трендів кінцевий прогноз визначається за формулою

$$Y_{t+v}^* = u_t (1 + b_t)^v I_{t-s+v},$$

$$\text{де } b_t = B \frac{u_t - u_{t-1}}{u_{t-1}} + (1 - B)b_{t-1}.$$

6.3. Моделювання сезонної хвилі на основі гармонійного аналізу

За наявності періодичних коливань ряду помісячної динаміки використовують також моделі сезонної хвилі на основі *гармонійного аналізу*. Основними її характеристиками є: амплітуда, фаза, період і частота коливань.

Амплітуда A характеризує відстань від середнього рівня максимуму (мінімуму) сезонної хвилі, *період коливань* T — тривалість циклу, *частота* f — кількість циклів в одиницю часу, тобто $f = 1/T$. Якщо $T = 12$ місяців, то $f = 1/12$ циклу в місяць. Відстань між початком відліку часу з точкою $t = 0$ і найближчим піком називають *фазою* Θ . Сезонну хвилю з періодом T можна описати функцією:

$$Y = a + b \cos \omega t + d \sin \omega t,$$

де ω — кутова частота гармоніки; вимірюється радіанами в одиницю часу $\omega = 2\pi f = 2\pi/T$ і змінюється в інтервалі $0 \leq \omega \leq 2\pi$;

b, d — коефіцієнти гармоніки, функціонально зв'язані з амплітудою: $A = \sqrt{b^2 + d^2}$.

Коефіцієнти гармоніки визначаються методом найменших квадратів. Завдяки властивостям ортогональності функцій синуса і косинуса система нормальних рівнянь приводиться до тотожностей

$$\sum y = an$$

$$\sum y \cos \omega t = \frac{1}{2} nb$$

$$\sum y \sin \omega t = \frac{1}{2} nd$$

$$\text{Звідси: для } n = 12 \quad a = \frac{\sum y}{12}; \quad b = \frac{\sum y \cos \omega t}{6}; \quad d = \frac{\sum y \sin \omega t}{6}.$$

Отже, a — це не що інше, як середньомісячний рівень ряду. Коефіцієнти b і d визначають амплітуду коливань навколо середнього рівня.

Очевидно, що чим більша амплітуда коливань, тим вагоміший вклад гармоніки в загальну дисперсію процесу. Оцінкою такого вкладу слугує дисперсійне відношення

$$R^2 = \frac{\delta^2}{\sigma^2},$$

де $\delta^2 = 0,5A^2$ — дисперсія гармоніки,

$$\sigma^2 = \frac{\sum (y - \hat{y})^2}{n} \text{ — загальна дисперсія процесу.}$$

У модель гармонійного аналізу можна включити декілька гармонік з різними періодами коливань. Скажімо, перша гармоніка з періодом 12, друга — з періодом 6, третя — з періодом 4 і т. д.

Гармонійна функція розкладає часовий ряд на правильні періодичні хвилі — синусоїди. Адекватність її реальному процесу залежить від того, наскільки сталими є частота й амплітуда коливань. Відносно сталий характер внутрішньорічної динаміки притаманний ринку сезонних товарів. У табл. 6.1 наведено помісячну динаміку середньої ціни свіжих огірків (грн.).

Таблиця 6.1 – Помісячна динаміка середньої ціни свіжих огірків

Місяць	Ціна, грн.	Місяць	Ціна, грн.	Місяць	Ціна, грн.
1	5,56	5	2,60	9	0,76
2	5,70	6	1,38	10	1,43
3	4,72	7	0,70	11	4,36
4	3,68	8	0,57	12	5,89

Для побудови гармонійної функції на стартовій панелі модуля ініціюємо кнопку *Spectral (Fourier) analysis* — Спектральний аналіз. У діалоговому вікні *Fourier (Spectral) analysis* виберемо опцію *Single series Fourier analysis* — Аналіз Фур'є одиничного ряду. Визначені за опцією *Summary* коефіцієнти гармонік наведено в табл. 6.2.

Таблиця 6.2 - коефіцієнти гармонік Фур'є

Spectral analysis: VAR3 (.sta)				
No. Of cases: 12				
Continue...	Frequency	Period	Cosine Coeffs	Sine Coeffs
0	0		−3,7E-17	−0
1	0,083	12	2,493	0,055
2	0,167	6	0,098	−0,678
3	0,25	4	−0,328	−0,456
4	0,333	3	−0,455	−0,037
5	0,417	2,4	−0,289	0,060
6	0,5	2	−0,176	−0

Найвагомішою виявилася перша гармоніка, яка з амплітудою $A = \sqrt{2,493^2 + 0,055^2} = 2,494$ пояснює 76,5% варіації ряду. Друга гармоніка пояснює 5,8% варіації ряду. Внесок решти гармонік — 17,7%.

Визначену амплітуду коливань можна використати при прогнозуванні сезонного процесу.

Рекомендована література

1. Єріна А. М., Єрін Д. Л. Статистичне моделювання та прогнозування : навч. посіб. К. : КНЕУ, 2014. 348 с.
2. Корепанов О. С. Статистичний аналіз тенденцій зовнішньоекономічної діяльності України: сектор ІКТ // Соціальна економіка. 2018. № 1, Вип. 55. С. 18–26.
3. Hyndman R. J. Forecasting: principles and practice / R. J. Hyndman, G. Athanasopoulos. – Otexts, 2018. – 504 p.
4. Ord, J. K., Fildes, R., & Kourentzes, N. Principles of business forecasting (2nd ed.). Wessex Press Publishing Co. 2017.

**3. ПЛАНИ ПРАКТИЧНИХ
(СЕМІНАРСЬКИХ) ЗАНЯТЬ,
САМОСТІЙНОЇ РОБОТИ
(у т.ч. ІНДИВІДУАЛЬНІ ЗАВДАННЯ)**

План проведення практичних занять

Практичне заняття 1. «Аналіз одномірних часових рядів»

Питання для дискусії:

1. Предмет, метод, завдання курсу
2. Базовий алгоритм методу «Гусениця»
3. Формальний опис базового алгоритму методу "Гусениця"

Завдання на практику (Початковий аналіз часового ряду)

1. Створити базу даних основних макроекономічних показників України за 1996-2016 роки: ВВП, експорт та імпорт товарів та послуг, грошова маса, інфляція, курс долара США до гривні, обсяг вкладів населення в банках тощо. Для роботи вибрати 5 часових рядів (не можуть повторюватися з іншими студентами). Кожен з рядів має містити не менше 30 спостережень. Для визначення потрібної інформації скористатися сайтами:

- www.me.gov.ua,
- www.bank.gov.ua

2. Імпортувати Вашу базу даних до MS Excel.

3. Провести графічний аналіз рядів даних. Визначити наявність трендового, сезонного компонентів.

4. Для обраних рядів даних визначити основні числові характеристики: вибіркове середнє, вибіркєву дисперсію, коваріацію перших 10 порядків.

5. Провести агрегацію Ваших часових рядів (квартальні дані перевести у річні, місячні – у квартальні). З отриманими рядами провести дезагрегацію даних за допомогою методів:

- пропорційного розбиття;
- поліноміальної інтерполяції. Порівняти отримані оцінки з фактичними.

Практичне заняття 2. «Вибір параметрів при застосуванні методу "Гусениця"»

Питання для дискусії:

1. Вибір довжини гусениці
2. Відбір головних компонентів

Практичне заняття 3. «Опис програми "Caterpillar", що реалізує метод "Гусениця"»

Питання для дискусії:

1. Загальна структура програми
2. Графічне представлення результатів
3. Посібник користувача
 - 3.1. Технічна інформація
 - 3.2. Загальні положення
 - 3.3. Формати файлів вихідних даних

- 3.4. Параметри етапів
- 3.4.1. Параметри завантаження часового ряду
- 3.4.2. Параметри перетворення часового ряду
- 3.4.3. Параметри розкладання часового ряду
- 3.4.4. Параметри відновлення часового ряду
- 4. Робота із графікою

Завдання на практику (Аналіз часового ряду з використанням програми "Caterpillar")

1. Створити базу даних основних макроекономічних показників України за 1996-2016 роки: ВВП, експорт та імпорт товарів та послуг, грошова маса, інфляція, курс долара США до гривні, обсяг вкладів населення в банках тощо. Для роботи вибрати 5 часових рядів (не можуть повторюватися з іншими студентами). Кожен з рядів має містити не менше 70 спостережень. Для визначення потрібної інформації скористатися сайтами:

- a. www.me.gov.ua,
- b. www.bank.gov.ua

2. Імпортувати Вашу базу даних до "Caterpillar".

3. Провести графічний аналіз рядів даних. Визначити наявність трендового та сезонного компонентів.

Для обраних рядів даних визначити основні числові характеристики: довжину гусениці; кількість головних компонентів для аналізу.

4. Провести аналіз динаміки часових рядів на основі методу «Гусениця».

Практичне заняття 4. «Приклад реалізації методу "Гусениця" для аналізу одномірних часових рядів»

Питання для дискусії:

- 1. Аналіз динаміки виробництва м'яса всіх видів (у живій вазі) в Україні
- 2. Оцінка внеску складових досліджуваного часового ряду
- 3. Прогнозування обсягів виробництва м'яса всіх видів (у живій вазі) в Україні

Завдання на практику (Аналіз часового ряду з використанням програми "Caterpillar")

1. Створити базу даних основних показників для аналізу динаміки виробництва м'яса всіх видів (у живій вазі) в Україні за 1996-2016 роки. Для роботи вибрати 3-5 часових рядів. Кожен з рядів має містити не менше 70 спостережень. Для визначення потрібної інформації скористатися сайтами:

- a. www.me.gov.ua,
- b. www.bank.gov.ua

2. Імпортувати Вашу базу даних до "Caterpillar".

3. Провести графічний аналіз рядів даних. Визначити наявність трендового та

сезонного компонентів.

Для обраних рядів даних визначити основні числові характеристики: довжину гусениці; кількість головних компонентів для аналізу.

4. Провести аналіз динаміки часових рядів на основі методу «Гусениця».

Практичне заняття 5. «Аналіз багатомірних часових рядів»

Питання для дискусії:

1. Опис базового алгоритму
2. Модельні приклади: сингулярні розкладання траекторних матриць
3. Разделимость
4. Приклади обробки рядів методами SSA

Практичне заняття 6. «Приклад реалізації методу «Гусениця» для аналізу багатомірних часових рядів»

Питання для дискусії:

1. Аналіз часових рядів виробництва яєць і середніх цін на яйця в Україні
2. Прогнозування економічного, внутрішнього виробничого потенціалу й цін продуктів тваринництва аграрного ринку

Питання для самостійної роботи студентів

Тема 1. Аналіз одномірних часових рядів

1. Предмет, метод, завдання курсу
2. Базовий алгоритм методу «Гусениця»
3. Формальний опис базового алгоритму методу "Гусениця"

Тема 2. Вибір параметрів при застосуванні методу "Гусениця"

1. Вибір довжини гусениці
2. Відбір головних компонентів

Тема 3. Опис програми "Caterpillar", що реалізує метод "Гусениця"

1. Загальна структура програми
2. Графічне представлення результатів
3. Посібник користувача
4. Робота із графікою

Тема 4. Приклад реалізації методу "Гусениця" для аналізу одномірних часових рядів

1. Аналіз динаміки виробництва м'яса всіх видів (у живій вазі) в Україні
2. Оцінка внеску складових досліджуваного часового ряду
3. Прогнозування обсягів виробництва м'яса всіх видів (у живій вазі) в Україні

Тема 5. Аналіз багатомірних часових рядів

1. Опис базового алгоритму
2. Модельні приклади: сингулярні розкладання траекторних матриць
3. Приклади обробки рядів методами SSA

Тема 6. Приклад реалізації методу «Гусениця» для аналізу багатомірних часових рядів

1. Аналіз часових рядів виробництва яєць і середніх цін на яйця в Україні
2. Прогнозування економічного, внутрішнього виробничого потенціалу й цін продуктів тваринництва аграрного ринку

Тести для самоконтролю (дайте всі вірні відповіді)

1. Який з перерахованих методів аналізу дозволяє розкласти часовий ряд на складові, які можна інтерпретувати як тренд, періодичні й сезонні компоненти й шум?

- 1) Кластерний аналіз.
- 2) Сингулярний спектральний аналіз.
- 3) Дисперсійний аналіз.
- 4) Методи авторегресії й ковзних середніх.

2. Що є відмінною рисою методу «Гусениця»? (може бути кілька правильних відповідей)

- 1) Він не вимагає попереднього завдання моделі ряду.
- 2) Можна проводити аналіз на базі коротких часових рядів.
- 3) Дозволяє розкласти часовий ряд на складові, які можна інтерпретувати як тренд, періодичні й сезонні компоненти й шум.
- 4) Необхідно заздалегідь задати параметричний вид тренда.

3. У чому полягає сутність методу «Гусениця»?

- 1) У перетворенні одновимірного ряду в багатовимірний і дослідженні отриманої багатовимірної траєкторії методом ковзної середньої;
- 2) У перетворенні одновимірного ряду в багатовимірний й наступному відновленні багатовимірного ряду;
- 3) У перетворенні одновимірного ряду в багатовимірний, дослідженні отриманої багатовимірної траєкторії методом головних компонентів і наступному відновленні багатовимірного ряду;
- 4) У дослідженні багатовимірної траєкторії методом авторегресії й наступному відновленні вихідного ряду;

4. Виберіть усі можливі складові вихідного ряду, які можуть бути виділені з використанням методу «Гусениця»

- 1) повільний тренд, сезонні й періодичні складові (якщо вони є) і випадкові варіації;
- 2) повільний тренд, сезонні й періодичні складові;
- 3) сезонні й періодичні складові, випадкові варіації;
- 4) тільки повільний тренд і випадкові варіації;

5. Що є основним керуючим параметром методу «Гусениця»?

- 1) Число власних чисел кореляційної матриці;
- 2) Число ступенів волі функції $f(t)$;
- 3) Довжина гусениці;
- 4) Кількість головних компонентів;

6. Які набори інтерпретуємих об'єктів використовують найчастіше в процесі відбору головних компонентів, інформативних у тому або іншому змісті?

- 1) Набори середніх і відносних коефіцієнтів варіації;
- 2) Набори власних чисел або векторів кореляційної матриці, набір головних компонентів;
- 3) Набори логарифмів власних чисел кореляційної матриці;
- 4) Тільки набори головних компонентів;

7. Яка комп'ютерна програма дозволяє проводити сингулярний спектральний аналіз часових рядів?

- 1) «SPSS».
- 2) «Stata».
- 3) «Statistica».
- 4) «Caterpillar».

8. Робота програми «Caterpillar» розділена на три послідовні етапи (не вважаючи читання даних з файлу), кожний з яких робить обчислювальні дії, а саме:

- 1) перетворення ряду, розкладання, відновлення;
- 2) відновлення ряду, розкладання, перетворення;
- 3) розкладання ряду, відновлення, перетворення;
- 4) денормування ряду, перетворення, відновлення;

9. Які види аналізу можна проводити з використанням програми «Caterpillar»? *(може бути кілька правильних відповідей)*

- 1) Аналіз динаміки тимчасового ряду.
- 2) Оцінка внеску складових досліджуваного тимчасового ряду.
- 3) Прогнозування значень часових рядів.
- 4) Аналіз взаємозв'язків двох і більш тимчасові рядів.

10. Які статистичні оцінки можна використовувати для визначення заходу точності отриманого прогнозу? *(може бути кілька правильних відповідей)*

- 1) MAPE.
- 2) Коефіцієнт кореляції.
- 3) Коефіцієнт Спірмена.
- 4) Коефіцієнт Генрі Тейла.

МЕТОДИЧНІ РЕКОМЕНДАЦІЇ ДО ВИКОНАННЯ ПИСЬМОВИХ САМОСТІЙНИХ РОБІТ

ІНДИВІДУАЛЬНО-РОЗРАХУНКОВА РОБОТА

Індивідуальні завдання є обов'язковою частиною самостійної роботи студента.

Розрахункове завдання виконується згідно з «Методичними вказівками до виконання індивідуального розрахункового завдання з навчальної дисципліни».

Написання індивідуальної роботи має сприяти глибшому засвоєнню студентами дисципліни, спонукає ґрунтовно вивчати спеціальні наукові видання вітчизняних і зарубіжних авторів та інтернет-ресурси.

Індивідуальне завдання передбачає самостійне виконання студентом розрахунково-графічної роботи з обраної теми дослідження. Студенти обирають проблемну ситуацію із запропонованих у переліку або за власним бажанням, збирають необхідні дані, здійснюють розрахунки та роблять відповідні висновки.

Індивідуальне завдання оцінюється за критеріями:

- самостійності виконання;
- логічності та послідовності викладення матеріалу;
- деталізації плану;
- повноти та глибини розкриття теми, проблемної ситуації, аналітичної частини;
- наявності ілюстрацій (таблиці, рисунки, схеми і т. д.);
- кількості використаних джерел;
- використання статистичної інформації, додаткових літературних джерел та ресурсів мережі Internet;
- відображення практичного досвіду;
- обґрунтованості висновків;
- якості оформлення, презентації та захисту індивідуального розрахункового завдання.

**4. ПИТАННЯ, ЗАДАЧІ, ЗАВДАННЯ АБО КЕЙСИ ДЛЯ
ПОТОЧНОГО КОНТРОЛЮ ЗНАНЬ І ВМІНЬ
ЗДОБУВАЧІВ ВИЩОЇ ОСВІТИ, ДЛЯ КОНТРОЛЬНИХ
РОБІТ, ПЕРЕДБАЧЕНИХ НАВЧАЛЬНИМ ПЛАНОМ,
ПІСЛЯТЕСТАЦІЙНОГО МОНІТОРИНГУ НАБУТИХ
ЗНАНЬ І ВМІНЬ З НАВЧАЛЬНОЇ ДИСЦИПЛІНИ**

Питання для поточного контролю знань і вмінь здобувачів вищої освіти за темами

Тема 1. Аналіз одновимірних часових рядів

Контрольні питання

1. Предмет, метод, завдання курсу
2. Базовий алгоритм методу «Гусениця»
3. Формальний опис базового алгоритму методу "Гусениця"

Тема 2. Вибір параметрів при застосуванні методу "Гусениця"

Контрольні питання

1. Вибір довжини гусениці
2. Відбір головних компонентів

Тема 3. Опис програми "Caterpillar", що реалізує метод "Гусениця"

Контрольні питання

1. Загальна структура програми
2. Графічне представлення результатів
3. Посібник користувача
4. Робота із графікою

Тема 4. Приклад реалізації методу "Гусениця" для аналізу одновимірних часових рядів

Контрольні питання

1. Аналіз динаміки виробництва м'яса всіх видів (у живій вазі) в Україні
2. Оцінка внеску складових досліджуваного часового ряду
3. Прогнозування обсягів виробництва м'яса всіх видів (у живій вазі) в Україні

Тема 5. Аналіз багатовимірних часових рядів

Контрольні питання

1. Опис базового алгоритму
2. Модельні приклади: сингулярні розкладання траекторних матриць
3. Приклади обробки рядів методами SSA

Тема 6. Приклад реалізації методу «Гусениця» для аналізу багатовимірних часових рядів

Контрольні питання

1. Аналіз часових рядів виробництва яєць і середніх цін на яйця в Україні
2. Прогнозування економічного, внутрішнього виробничого потенціалу й цін продуктів тваринництва аграрного ринку

**Тестові завдання поточного контролю знань і вмінь студентів
з навчальної дисципліни**

Варіант 1 (приклад)

1. Який з перерахованих методів аналізу дозволяє розкласти часовий ряд на складові, які можна інтерпретувати як тренд, періодичні й сезонні компоненти й шум?

- 1) Кластерний аналіз.
- 2) Сингулярний спектральний аналіз.
- 3) Дисперсійний аналіз.
- 4) Методи авторегресії й ковзних середніх.

2. Що є відмінною рисою методу «Гусениця»? *(може бути кілька правильних відповідей)*

- 1) Він не вимагає попереднього завдання моделі ряду.
- 2) Можна проводити аналіз на базі коротких часових рядів.
- 3) Дозволяє розкласти часовий ряд на складові, які можна інтерпретувати як тренд, періодичні й сезонні компоненти й шум.
- 4) Необхідно заздалегідь задати параметричний вид тренда.

3. У чому полягає сутність методу «Гусениця»?

- 1) У перетворенні одновимірного ряду в багатовимірний і дослідженні отриманої багатовимірної траєкторії методом ковзної середньої;
- 2) У перетворенні одновимірного ряду в багатовимірний й наступному відновленні багатовимірного ряду;
- 3) У перетворенні одновимірного ряду в багатовимірний, дослідженні отриманої багатовимірної траєкторії методом головних компонентів і наступному відновленні багатовимірного ряду;
- 4) У дослідженні багатовимірної траєкторії методом авторегресії й наступному відновленні вихідного ряду;

4. Виберіть усі можливі складові вихідного ряду, які можуть бути виділені з використанням методу «Гусениця»

- 1) повільний тренд, сезонні й періодичні складові (якщо вони є) і випадкові варіації;
- 2) повільний тренд, сезонні й періодичні складові;
- 3) сезонні й періодичні складові, випадкові варіації;
- 4) тільки повільний тренд і випадкові варіації;

5. Що є основним керуючим параметром методу «Гусениця»?

- 1) Число власних чисел кореляційної матриці;
- 2) Число ступенів волі функції $f(t)$;
- 3) Довжина гусениці;
- 4) Кількість головних компонентів;

6. Які набори інтерпретуємих об'єктів використовують найчастіше в процесі відбору головних компонентів, інформативних у тому або іншому змісті?

- 1) Набори середніх і відносних коефіцієнтів варіації;
- 2) Набори власних чисел або векторів кореляційної матриці, набір головних компонентів;
- 3) Набори логарифмів власних чисел кореляційної матриці;
- 4) Тільки набори головних компонентів;

7. Яка комп'ютерна програма дозволяє проводити сингулярний спектральний аналіз часових рядів?

- 1) «SPSS».
- 2) «Stata».
- 3) «Statistica».
- 4) «Caterpillar».

8. Робота програми «Caterpillar» розділена на три послідовні етапи (не вважаючи читання даних з файлу), кожний з яких робить обчислювальні дії, а саме:

- 1) перетворення ряду, розкладання, відновлення;
- 2) відновлення ряду, розкладання, перетворення;
- 3) розкладання ряду, відновлення, перетворення;
- 4) денормування ряду, перетворення, відновлення;

9. Які види аналізу можна проводити з використанням програми «Caterpillar»? *(може бути кілька правильних відповідей)*

- 1) Аналіз динаміки тимчасового ряду.
- 2) Оцінка внеску складових досліджуваного тимчасового ряду.
- 3) Прогнозування значень часових рядів.
- 4) Аналіз взаємозв'язків двох і більш тимчасові рядів.

10. Які статистичні оцінки можна використовувати для визначення заходу точності отриманого прогнозу? *(може бути кілька правильних відповідей)*

- 1) MAPE.
- 2) Коефіцієнт кореляції.
- 3) Коефіцієнт Спірмена.
- 4) Коефіцієнт Генрі Тейла.

**5. ЗАВДАННЯ ПІДСУМКОВОГО КОНТРОЛЮ –
СЕМЕСТРОВИЙ ЗАЛІК
(дворівнева шкала оцінювання)**

Основні питання для підготовки до заліку

1. Предмет, метод і завдання курсу.
2. Роль і значення сингулярного спектрального аналізу в дослідженні динаміки соціально-економічних процесів і явищ.
3. Основні переваги сингулярного спектрального аналізу в дослідженні динаміки соціально-економічних процесів і явищ.
4. Основні етапи базового алгоритму методу «Гусениця» при проведенні сингулярного спектрального аналізу.
5. Формальний опис базового алгоритму методу «Гусениця».
6. Вибір довжини гусениці.
7. Відбір головних компонент.
8. Загальна структура програми «Caterpillar».
9. Основні принципи роботи програми «Caterpillar», що реалізує метод «Гусениця» при проведенні сингулярного спектрального аналізу в економіко-статистичних дослідженнях.
10. Основні етапи практичного застосування методу «Гусениця» для аналізу одномірних тимчасових рядів.
11. Графічне представлення результатів у програмі «Caterpillar».
12. Особливості прогнозування часових рядів з використанням методу «Гусениця» при проведенні сингулярного спектрального аналізу.

ЗАЛІКОВІ ЗАВДАННЯ

Залікове завдання № 1 (приклад)

1. Предмет, метод і завдання курсу.

2. Відбір головних компонент.

3. Тести

1. Який з перерахованих методів аналізу дозволяє розкласти часовий ряд на складові, які можна інтерпретувати як тренд, періодичні й сезонні компоненти й шум?

- 1) Кластерний аналіз.
- 2) Сингулярний спектральний аналіз.
- 3) Дисперсійний аналіз.
- 4) Методи авторегресії й ковзних середніх.

2. Що є відмінною рисою методу «Гусениця»? *(може бути кілька правильних відповідей)*

- 1) Він не вимагає попереднього завдання моделі ряду.
- 2) Можна проводити аналіз на базі коротких часових рядів.
- 3) Дозволяє розкласти часовий ряд на складові, які можна інтерпретувати як тренд, періодичні й сезонні компоненти й шум.
- 4) Необхідно заздалегідь задати параметричний вид тренда.

3. У чому полягає сутність методу «Гусениця»?

- 1) У перетворенні одновимірного ряду в багатовимірний і дослідженні отриманої багатовимірної траєкторії методом ковзної середньої;
- 2) У перетворенні одновимірного ряду в багатовимірний й наступному відновленні багатовимірного ряду;
- 3) У перетворенні одновимірного ряду в багатовимірний, дослідженні отриманої багатовимірної траєкторії методом головних компонентів і наступному відновленні багатовимірного ряду;
- 4) У дослідженні багатовимірної траєкторії методом авторегресії й наступному відновленні вихідного ряду;

4. Виберіть усі можливі складові вихідного ряду, які можуть бути виділені з використанням методу «Гусениця»

- 1) повільний тренд, сезонні й періодичні складові (якщо вони є) і випадкові варіації;
- 2) повільний тренд, сезонні й періодичні складові;
- 3) сезонні й періодичні складові, випадкові варіації;
- 4) тільки повільний тренд і випадкові варіації;

5. Що є основним керуючим параметром методу «Гусениця»?

- 1) Число власних чисел кореляційної матриці;
- 2) Число ступенів волі функції $f(t)$;
- 3) Довжина гусениці;
- 4) Кількість головних компонентів;

6. Які набори інтерпретуємих об'єктів використовують найчастіше в процесі відбору головних компонентів, інформативних у тому або іншому змісті?

- 1) Набори середніх і відносних коефіцієнтів варіації;
- 2) Набори власних чисел або векторів кореляційної матриці, набір головних компонентів;
- 3) Набори логарифмів власних чисел кореляційної матриці;
- 4) Тільки набори головних компонентів;

7. Яка комп'ютерна програма дозволяє проводити сингулярний спектральний аналіз часових рядів?

- 1) «SPSS».
- 2) «Stata».
- 3) «Statistica».
- 4) «Caterpillar».

8. Робота програми «Caterpillar» розділена на три послідовні етапи (не вважаючи читання даних з файлу), кожний з яких робить обчислювальні дії, а саме:

- 1) перетворення ряду, розкладання, відновлення;
- 2) відновлення ряду, розкладання, перетворення;
- 3) розкладання ряду, відновлення, перетворення;
- 4) денормування ряду, перетворення, відновлення;

9. Які види аналізу можна проводити з використанням програми «Caterpillar»? *(може бути кілька правильних відповідей)*

- 1) Аналіз динаміки тимчасового ряду.
- 2) Оцінка внеску складових досліджуваного тимчасового ряду.
- 3) Прогнозування значень часових рядів.
- 4) Аналіз взаємозв'язків двох і більш тимчасові рядів.

10. Які статистичні оцінки можна використовувати для визначення заходу точності отриманого прогнозу? *(може бути кілька правильних відповідей)*

- 1) MAPE.
- 2) Коефіцієнт кореляції.
- 3) Коефіцієнт Спірмена.
- 4) Коефіцієнт Генрі Тейла.

6. КРИТЕРІЇ ОЦІНЮВАННЯ ЗНАНЬ СТУДЕНТІВ ТА РОЗПОДІЛ БАЛІВ

СХЕМА НАРАХУВАННЯ БАЛІВ

Структура засобів контролю та розподіл балів із дисципліни наведена в таблицях.

Узагальнена схема нарахування балів (денна форма навчання)

Поточний контроль, самостійна робота, індивідуальні завдання										Залікова робота	Сума
Поточне оцінювання						Проміжне тестування	Контрольна робота, передбачена навчальним планом	Індивідуальне завдання	Разом		
Розділ 1			Розділ 2								
T1	T2	T3	T4	T5	T6						
2	2	4	4	4	4	20	-	20	60	40	100

T1, T2 ... – теми розділів.

Узагальнена схема нарахування балів (заочна форма навчання)

Поточний контроль, самостійна робота, індивідуальні завдання										Залікова робота	Сума
Поточне оцінювання						Проміжне тестування	Контрольна робота, передбачена навчальним планом	Індивідуальне завдання	Разом		
Розділ 1			Розділ 2								
T1	T2	T3	T4	T5	T6						
2	2	4	4	4	4	20	-	20	60	40	100

T1, T2 ... – теми розділів.

Оцінювання знань, умінь та навичок студентів включає ті види занять, які згідно з програмою навчальної дисципліни передбачають лекційні, практичні заняття, самостійну роботу.

Перевірка та оцінювання знань студентів проводиться в наступних формах:

- поточне оцінювання роботи і знань студентів під час практичних занять;
- складання проміжного контролю знань за змістовими розділами (тестування);
- складання заліку.

Поточне оцінювання знань студентів здійснюється під час проведення практичних, і має на меті перевірку рівня підготовленості студента до виконання конкретної роботи. Об'єктами поточного контролю є:

- активність та результативність роботи студента протягом семестру над вивченням програмного матеріалу дисципліни;
- відвідування занять;
- виконання індивідуального науково-дослідного завдання;
- складання проміжного контролю.

Контроль систематичного виконання самостійної роботи та активності на практичних заняттях проводиться за такими критеріями:

- О розуміння, ступінь засвоєння теорії та методології проблем, що розглядаються;
- О ступінь засвоєння фактичного матеріалу навчальної дисципліни;
- О ознайомлення з рекомендованою літературою, а також із сучасною літературою з питань, що розглядаються;
- О уміння поєднувати теорію із практикою при розгляді практичних ситуацій, розв'язанні задач, проведенні розрахунків при виконанні індивідуальних завдань, та завдань, винесених на розгляд в аудиторії;
- О оволодіння методами економіко-статистичної обробки даних із використанням комп'ютерних технологій;
- О логіка, структура, стиль викладу матеріалу в письмових роботах і при виступах в аудиторії, уміння обґрунтовувати свою позицію, здійснювати узагальнення інформації та робити висновки.

Оцінювання знань студента під час виконання завдань для самостійної роботи проводиться за 4- бальною шкалою.

Оцінка «відмінно» ставиться за умови відповідності виконаного завдання студента або його усної відповіді до всіх зазначених критеріїв. Відсутність тієї чи іншої складової знижує оцінку.

При оцінюванні практичних занять увага приділяється також їх якості та самостійності, своєчасності здачі виконаних завдань викладачу (згідно з графіком навчального процесу). Якщо якась із вимог не буде виконана, то оцінка буде знижена.

Проміжний контроль рівня знань передбачає виявлення опанування студентом лекційного матеріалу та вміння застосування його для вирішення практичної ситуації й проводиться у вигляді тестування. При цьому тестове завдання може містити як запитання, що стосуються суто теоретичного матеріалу, так і запитання, спрямовані на вирішення невеличкого практичного завдання.

Проміжний тестовий контроль проводиться один раз на семестр. Загальна тривалість тестів – 1,5 години.

Тестування складається з 10 тестів. Одна правильна відповідь на кожен із тестів дорівнює 2 балам. Тестове завдання містить запитання одиничного і множинного вибору різного рівня складності. Тести для проміжного контролю обираються із загального переліку тестів за відповідними темами.

Підсумковий контроль за курсом – у формі заліку.

До складання заліку допускають студентів, що мають задовільну кількість балів зі складених тестів з основних навчальних елементів розділів, написання та захисту індивідуального науково-дослідного завдання та інших завдань передбачених програмою дисципліни.

Залік здійснюється за заліковими завданнями, які містять три питання: два теоретичних питання і тести. Вони дають можливість здійснити оцінювання знань студента за дисципліною. Залікове завдання оцінюється за дворівневою шкалою.

КРИТЕРІЇ ОЦІНЮВАННЯ (ЗАСОБИ ДІАГНОСТИКИ)

Критерії оцінювання результативності роботи студентів при виконанні самостійної роботи

Якісними критеріями оцінювання виконання індивідуальних завдань студентами є:

1. Повнота виконання завдання:

- Елементарна;
- Фрагментарна;
- Повна;
- Неповна.

2. Рівень самостійності студента

- під керівництвом викладача;
- консультація викладача;
- самостійно.

3. *Сформованість навчально-інформаційних умінь* (роботи з підручником, володіння різними способами читання, складання плану, рецензій, конспекту, вміння користуватися бібліотекою, спостереження, експеримент тощо)

4. *Сформованість навчально-інтелектуальних умінь* (визначення понять, аналіз, синтез, порівняння, класифікація, систематизація, узагальнення, абстрагування, вміння відповідати на запитання, виконувати творчі завдання тощо);

5. *Рівень сформованості фахових методичних вмінь* (вміння застосовувати на практиці набуті знання):

- низький – володіння умінням здійснювати первинну обробку навчальної інформації без подальшого її аналізу;
- середній – уміння вибирати відомі способи дій для виконання фахових завдань;
- достатній – застосовує набуті знання у стандартних практичних ситуаціях;
- високий – володіння умінням творчо-пошукової діяльності.

Критерії оцінювання здобувачів вищої освіти за відповіді на питання поточного контролю

Завданням поточного контролю є перевірка розуміння та засвоєння певного матеріалу, вироблених навичок проведення розрахункових робіт, умінь самостійно опрацьовувати тексти, здатності осмислити зміст теми чи розділу, умінь публічно чи письмово представити певний матеріал (презентація).

Якісними критеріями оцінювання виконання завдань поточного контролю є:

1. *Повнота відповіді або виконання завдання:*
 - елементарна;
 - фрагментарна;
 - повна;
 - неповна.
2. *Рівень сформованості логічних умінь:*
 - елементарні дії;
 - операція, правило, алгоритм;
 - правила визначення понять;
 - формулювання законів і закономірностей;
 - структурування суджень, доводів, описів.

Критерії оцінювання здобувачів вищої освіти за виконання індивідуальної розрахункової роботи

Індивідуально-розрахункова робота студента складається з двох частин: перша – опрацювання теоретичного питання, друга – виконання практичного завдання. Максимальна оцінка за виконання завдань індивідуальної розрахункової роботи – 20 балів.

Виконання завдань оцінюється за такими критеріями:

- 1) теоретичне питання:
 - повнота й ґрунтовність викладу;
 - аргументованість тверджень;
 - суб'єктне усвідомлення змісту;
 - термінологічна коректність;
- 2) практичне завдання:
 - технологічна грамотність;
 - методична грамотність;
 - обґрунтованість висновків;
 - правильність оформлення.

Шкала оцінювання індивідуальної розрахункової роботи

Кількість балів	Теоретичне питання	Практичне завдання
15-20	Повне засвоєння та суб'єктне усвідомлення матеріалу. Твердження чітко аргументовані. Продemonстровано термінологічну грамотність	Продemonстровано методичну й технологічну грамотність. Методичні рішення обґрунтовано. Оформлення відповідає вимогам.
10-14	Повне засвоєння матеріалу, але недостатнє суб'єктне його усвідомлення. Нечітка аргументація тверджень. Часткова термінологічна некоректність.	Наявність незначних методичних і технологічних помилок, а також помилок в оформленні роботи. Методичні рішення обґрунтовано.
5-9	Часткове засвоєння матеріалу, суб'єктне його не усвідомлення. Аргументація відсутня. Термінологічна неграмотність.	Наявність значної кількості методичних і технологічних помилок, а також в оформленні роботи.

0-4	Теоретичний матеріал не засвоєно. Аргументація відсутня. Термінологічна неграмотність.	Методична й технологічна неграмотність. Неправильне оформлення роботи.
-----	--	--

Критерії оцінювання на заліку

Оцінювання знань студента проводиться за дворівневою шкалою (*зараховано* - відмінно, добре, задовільно; *не зараховано* - незадовільно). За залік студент може отримати максимум 40 балів:

1. Для отримання оцінки «відмінно» (35-40 балів) студент повинен:
 - укластися у встановлений строк підготовки відповіді;
 - викласти теоретичний матеріал чітко, коротко, зв'язно й обґрунтовано;
 - навести вірне рішення задачі та тестів.
2. Для одержання оцінки «добре» (25-34 бала) студент повинен:
 - укластися у встановлений строк підготовки відповіді;
 - викласти теоретичний матеріал зв'язно й обґрунтовано;
 - навести вірне рішення задачі;
 - можливі помилки у відповідях на тести.
3. Для отримання оцінки «задовільно» (15-24 бала) студент повинен:
 - викладати теоретичний матеріал у доступній для розуміння формі;
 - можливі помилки при розв'язанні задачі та в тестах.
4. Оцінку «незадовільно» (1-14 балів) отримують студенти, відповіді яких можуть бути оцінені нижче вимог, сформульованих у попередніх пунктах.

Кожне завдання екзамену оцінюється окремо. Загальна оцінка дорівнює сумі оцінок за усі завдання (засоби контролю). Якщо одна з оцінок «незадовільно», то загальна оцінка не може бути вищою за «задовільно».

Шкала оцінювання залікової роботи:

- за кожну правильну відповідь на теоретичні питання завдань № 1 і 2 студент одержує по 7 балів;
- за правильно виконане завдання № 3 (розрахункова задача) – 16 балів;
- за правильно виконане завдання № 4 (тести) студент одержує 10 балів.

Підсумкова оцінка з навчальної дисципліни визначається як загальна оцінка, яка враховує оцінки з кожного виду контролю (оцінки поточного та проміжного контролю за роботу протягом семестру, індивідуальне завдання та оцінка за результатами підсумкового заліку).

***Зведена шкала оцінювання роботи студентів з дисципліни
«Сучасні методи прогнозування»***

Види робіт	Максимум балів
Поточний контроль на заняттях	10
Проміжне тестування	20
Індивідуальна робота	30
РАЗОМ	60
<i>Залік</i>	40
ВСЬОГО	100

У відповідності до набраних студентом балів за розділами та за залік оцінка знання матеріалу проводиться за дворівневою шкалою оцінювання згідно з Методикою переведення показників успішності знань студентів.

Шкала оцінювання

Сума балів за всі види навчальної діяльності протягом семестру	Оцінка за національною шкалою
	для дворівневої шкали оцінювання
90-100	зараховано
80-89	
70-79	
60-69	
50-59	
1-49	не зараховано