НАЦІОНАЛЬНИЙ ТЕХНІЧНИЙ УНІВЕРСИТЕТ УКРАЇНИ

«КИЇВСЬКИЙ ПОЛІТЕХНІЧНИЙ ІНСТИТУТ
імені ІГОРЯ СІКОРСЬКОГО»
Інститут прикладного системного аналізу
Кафедра математичних методів системного аналізу

До захисту допущено
Завідувач кафедри
_____ Оксана ТИМОЩУК
«07» 06 2021 р.

Дипломна робота

на здобуття ступеня бакалавра
за освітньо-професійною програмою «Системний аналіз і управління»
спеціальності 124 "Системний аналіз"
на тему: «Моделі та прогнозування нелінійних нестаціонарних процесів»

Виконав:
Студент IV курсу, групи КА-73
Гаврюшин Павло Андрійович _____

Керівник:
професор кафедри ММСА
к.т.н., професор Петро Іванович Бідюк _____

Консультант з економічного розділу:
доцент кафедри ТПЕ
к.е.н., доцент Надія Василівна Рощіна _____

Консультант з нормоконтролю:
доцент кафедри ММСА
к.т.н., доцент Анатолій Єпіфанович Коваленко _____

Рецензент:
Декан ФІОТ, Сергій Федорович Теленик
д.т.н., професор _____

Засвідчую, що у цій дипломній роботі
немає запозичень з праць інших авторів
без відповідних посилань.
Студент _____

Київ-2021

Національний технічний університет України

«Київський політехнічний інститут імені Ігоря Сікорського»

Інститут прикладного системного аналізу

Кафедра математичних методів системного аналізу

Рівень вищої освіти – перший (бакалаврський)

Спеціальність – 124 "Системний аналіз"

Освітньо-професійна програма «Системний аналіз і управління»

ЗАТВЕРДЖУЮ

Завідувач кафедри

_____Оксана ТИМОЩУК

«___»_____20__ р.

ЗАВДАННЯ

на дипломну роботу студенту

Гаврюшину Павлу Андрійовичу

1. Тема роботи «Моделі і прогнозування нелінійних нестаціонарних процесів», керівник роботи к.т.н., професор Бідюк Петро Іванович, затверджені наказом по університету від «___»_____ 2021 р. №_____

2. Термін подання студентом роботи 07.06.2021 р.

3. Вихідні дані до роботи: Вибірки даних часових рядів.

4. Зміст роботи: Моделювання та прогнозування нелінійних нестаціонарних процесів

5. Перелік ілюстративного матеріалу: (із зазначенням плакатів, презентацій тощо)

6. Консультанти розділів роботи:

Розділ	Прізвище, ініціали та посада консультанта	Підпис, дата	
		завдання видав	завдання прийняв

Економічний	Рощина Н.В.		

7. Дата видачі завдання: 02 квітня 2021

Календарний план

№ з/п	Назва етапів виконання дипломної роботи	Термін виконання етапів роботи	Примітка
1.	Вивчення літератури за темою роботи.	08.04.2020	Виконано
2.	Підготовка першого розділу.	10.05.2020	Виконано
3.	Підготовка другого розділу.	18.05.2020	Виконано
4.	Практична частина.	22.05.2020	Виконано
5.	Підготовка третього розділу	25.05.2020	Виконано
6.	Підготовка економічної частини	28.05.2020	Виконано
7.	Оформлення розділів відповідно до нормоконтролю.	02.06.2020	Виконано
8.	Підготовка презентації доповіді.	02.06.2020	Виконано
9.	Оформлення дипломної роботи.	05.06.2000	Виконано

Студент Павло ГАВРЮШИН

Керівник Петро БІДЮК

Реферат

Дипломна робота: 109 с., 8 табл., 39 рис., 1 додаток, 19 джерел.

Так як в загальному випадку усі достатньо складні досліджувані процеси(економічні, соціальні, екологічні тощо) підвержені впливу випадкових збурень, спровокованих дією великої кількості факторів різної природи, то ці процеси супроводжуються нестаціонарністю, що, в свою чергу, часто призводить до нелінійності, яка вимагає складніших прийомів для побудови та аналізу відповідних моделей. Очевидно, що існує необхідність якнайточніше моделювати оточуючі процеси як у сфері економіки, так і у будь-якій іншій. Тож актуальність розвитку у цьому напрямку важко переоцінити, адже це допоможе більш точно обробляти дані та представляти кращі прогнози.

Об’єкт дослідження - нелінійні нестаціонарні процеси різної природи, побудова часових рядів

Предмет дослідження – математичні моделі та прогнози нелінійних нестаціонарних економічних процесів, методи їх побудови, часові ряди

Мета проекту – Побудова адекватних моделей для аналізу та прогнозування нелінійних нестаціонарних процесів

ABSTRACT

Thesis: 109 p., 8 tabl., 39 fig., 1 appendices, 19 sources

In the general case all rather complex researched processes (economic, social, ecological, etc.) are exposed to random disturbances provoked by action of a large number of factors of various nature, these processes are followed by nonstationarity that, in turn, often leads to nonlinearity which demands more complex techniques for constructing and analyzing appropriate models. Obviously, there is a need to model the surrounding processes as accurately as possible, both in the field of economics and in any other. Therefore, the relevance of development in this direction is difficult to overestimate, because it will help to process data more accurately and present better forecasts.

The object of research is nonlinear non-stationary economic processes, construction of time series

Subject of research - mathematical models and forecasts of nonlinear nonstationary economic processes, methods of their construction, time series

The purpose of the project is to develop a system for analysis and forecasting of nonlinear nonstationary processes

Зміст

РОЗДІЛ 1
ЗАГАЛЬНА ПРОБЛЕМАТИКА ТА АКТУАЛЬНІСТЬ МОДЕЛЮВАННЯ НЕЛІНІЙНИХ НЕСТАЦІОНАРНИХ ПРОЦЕСІВ

1.1 Актуальність проблеми моделювання

Так як в загальному випадку усі достатньо складні досліджувані процеси(економічні, соціальні, екологічні тощо) підвержені впливу випадкових збурень, спровокованих дією великої кількості факторів різної природи, то ці процеси супроводжуються нестаціонарністю, що, в свою чергу, часто призводить до нелінійності, яка вимагає складніших прийомів для побудови та аналізу відповідних моделей. Очевидно, що існує необхідність якнайточніше моделювати оточуючі процеси як у сфері економіки, так і у будь-якій іншій. Тож актуальність розвитку у цьому напрямку важко переоцінити, адже це допоможе більш точно обробляти дані та представляти кращі прогнози.

1.2 Огляд нелінійних нестаціонарних процесів

Стаціонарний процес-це стохастичний(випадковий) процесс, розподіл вірогідності якого не змінюється із плином часу. Отже, наприклад, його середнє значення та дисперсія є постійними величинами.

Лінійний процес-процес, який можна із достатньою степінню точності задати лінійною формулою. Наприклад, лінійна регрісійна модель:

$$y = f(x, b) + e \qquad (1.1)$$

де e – випадкова похибка

f(x, b) – лінійна функція, у якій x є регресорами, а b – коефіцієнтами при них

Відповідно до вищесказаного, нелінійний нестаціонарний процесс є такий, у якого функція f(x, b) є нелінійною і розподіл вірогідності якого змінюється із часом.

На рисунку 1.1 предлставлено класифікацію процесів.

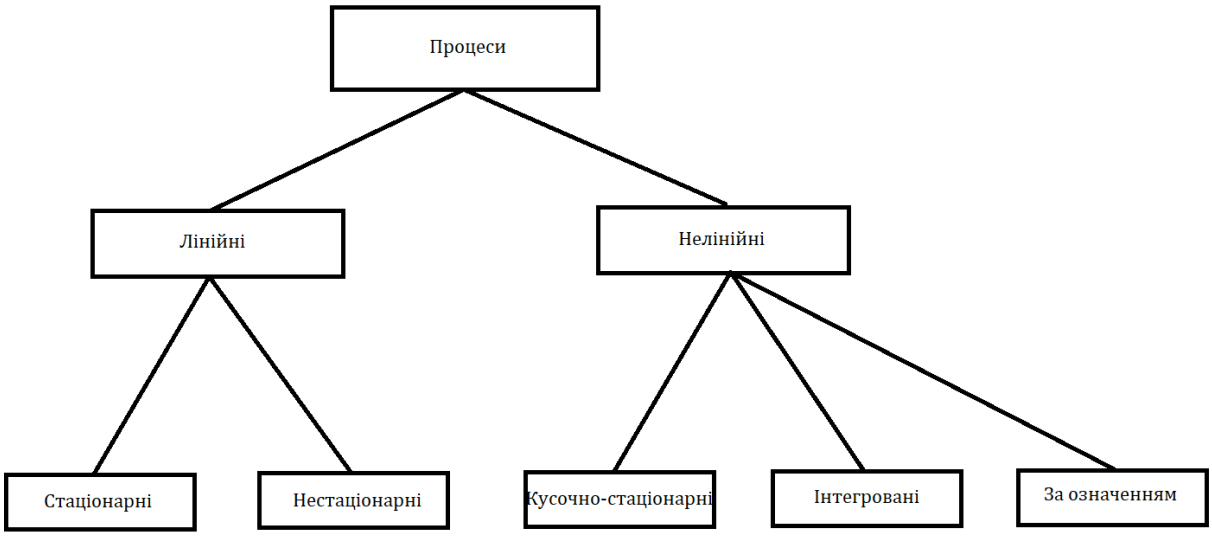


Рисунок 1.1 Класифікація процесів

Кусочно-стаціонарний процес-процес, який є стаціонарним на певних проміжках, проте не є стаціонарним у загальному

Інтегрований процес, в свою чергу, можна поділити на процес(лінійний або ні) із детермінованим трендом та процес із стохастичним трендом.

Детермінований тренд не має випадкової компоненти і його коефіцієнти не змінюються з часом.

Одним із найпростіших прикладів детермінованого тренду є лінійний тренд без шуму, тобто

$$X(t) = a + bt \tag{1.2}$$

де a – початковий рівень тренда,

b – швидкість змінення(константа) тренду, тобто середнє його змінення за одиницю часу.

X(t)-прогнозована величина в момент часу t.

Як видно із назви та функції тренду, лінійний тренд використовують для прогнозування часових рядів при наявності постійного росту чи спадання значень. Такий тренд часто використовують, наприклад, для прогнозування продажів, або за ситуації, коли наявні кілька факторів і коефіцієнт b виступає в якості усередненого впливу цих факторів.

Також є популярним детермінований, але вже не лінійний тренд

$$X(t) = a + bt + \mathrm{c}t\text{^}2, \tag{1.3}$$

де a – початковий рівень тренду

b – швидкість змінення тренду

c – прискорення змінення тренду

X(t)-прогнозована величина в момент часу t.

Такий тренд називається параболічним та є частковим випадком поліноміального тренду

Тренд такого типу використовується при наявності росту або спадання із прискоренням(заторможенням). Наприклад, зміна кількості осіб у популяції. Або, ближче до економіки, при зміні ставки за кредитом чи зняття обмежень у розподілі ресурсів.

Також існують інші типи трендів: експоненційний є характерним для процесів, що не мають обмежень для росту; логарифмічний використовують для опису процесів із зменшенням динаміки росту(падіня) тощо

Стохастичний тренд – тренд, який має випадкову компоненту та коефіцієнти якого змінюються з часом. Тобто тренд може задаватися так само як і детермінований(лінійний, параболічний тощо), проте із додатковою випадковою величиною.

Нелінійним процессом за означенням є будь який процесс, який неможливо достатньо точно представити лінійною функцією. Розглянемо поняття гетероскедастичності.

Гетероскедастичність – ознака процесу, яка являє собою неоднорідність спостережень, тобто зміну дисперсії процесу із плином часу. Через це такі процеси зазвичай важко описати лінійною функцією. Прикладами саме таких процесів є GARCH, ARCH та інші. Відхилення часто використовують у якості міри ризику, тому гетероскедастичні процеси використовують для моделювання та прогнозування економічних процесів.

1.3 Тести на нестаціонарність, гетероскедастичність, нелінійність

Виділяють два типи стаціонарності процесів, а саме:

---Сильна стаціонарність

---Слабка стаціонарність(також стаціонарність у широкому сенсі)

Слабка стаціонарність вимагає:

---Існування матсподівання та дисперсії процесу

$$E[X(t)] < \infty \ \forall t \epsilon T$$

$$D[X(t)] < \infty \ \forall t \epsilon T$$

де T-множина всіх моментів часу

---Математичне сподівання не залежить від часу

$$E[X(t)] = E[X(t-k)] = a = const$$

де $k \epsilon T$

---Дисперсія не залежить від часу

$$D[X(t)] = D[X(t-k)] = \sigma = const$$

---Автоковаріаційна функція залежить тільки від різниці моментів часу

$$Cov[X(t),X(t-k)]=Cov[X(t-s),X(t-k-s)]=const$$

де $t \epsilon T$

Сильна стаціонарність не вимагає існування матсподівання та дисперсії.

Отже, слабка стаціонарність є більш сильною умовою ніж сильна

Тести на гетероскедастичність

Виділяють наступні тести для перевірки на гетероскедастичність:

--- тест Уайта

--- тест Голдфелда-Квандта

--- тест Бройша-Пагана

--- тест Глейзера

Розглянемо алгоритм кожного тесту

Тест Уайта

Суть тесту Уайта полягає у формуванні основної і альтернативної гіпотези щодо гетероскедастичності процесу. Основною гіпотезою приймається гомоскедастичність. Далі будується допоміжна модель регресії для квадратів залишків, які були отримані в результаті оцінки параметрів вихідної регресії.

Допоміжна регресія містить константу, та ненадлишкові із наступних регресорів: самі регресори, їх квадрати та їх взаємні добутки.

Далі вираховується тестова статистика за формулою nR^2, де R^2-коефіцієнт детермінаціїї допоміжної регресії. При істинності нульової гіпотези тестова статистика має хі-квадрат розподіл із (m-1) ступенями свободи, де m-кількість коефіцієнтів останньої регресії.

Для прийняття нульової гіпотези(гомоскедастичність) nR^2 має бути незначною, а саме менша за Xi^2(m-1).

Наприклад, маємо наступну регресію

$$X(t)=a_0 + a_1x_1(t) + a_2x_2(t) + \varepsilon(t) \quad (1.4)$$

тобто вектор вимірів незалежних змінних має вигляд: $[1\ x_1\ x_2]^T$. В даному випадку всього є дев'ять можливих змінних, але 1 в квадраті залишається одиницею, а взаємні добутки регресорів на 1 призводять до надлишків. Тому множина всіх ненадлишкових змінних, яка складається із регресорів, їхніх квадратів та взаємних добутків, має вигляд: $[1\ x_1\ x_2\ x_1^2\ x_2^2\ x_1x_2]$.

Отже, відповідний Хі-квадрат розподіл має 5 ступенів свободи. Після оцінки параметрів регресії(наприклад, методом МНК), знаходимо nR^2 і порівнюємо із Хі^2(5) і обираємо відповідну гіпотезу. Якщо nR^2 < Xi^2 α (5) буде прийнята нульова гіпотеза із рівнем значущості α.

Тест Голдфелда-Квандта

Цей тест застосовують припускаючи залежність гетероскедастичності від якоїсь однієї змінної-регрессора X_i. Алгоритм методу наступний:

Припустимо позитивну кореляцію між X_i та відхиленням.

Як і у тесті Уайта, формується основна і альтернативна гіпотези щодо гетероскедастичності процесу. Нульвою гіпотезою приймається гомоскедастичність.

Спочатку відсортувати дані за спаданням по змінній X_i, після чого відкинути якийсь відсоток середніх значень(як правило, відкидають близько чверті даних).Нехай було відкинуто m даних. Далі будуються моделі авторегресіїї окремо для двох отриманих груп даних. Наступним кроком є обчислення RSS кожної регресії. Припустимо, що RSS_1 та RSS_2 є помилками регресій побудованих за менешими значеннями X_i відповідно. Тоді значення

$$P = \frac{RSS(1)}{RSS(2)} \quad (1.5)$$

Буде мати(в припущенні істинності гетероскедастичності) розподіл Фішера із $(\frac{n-m}{2} - k, \frac{n-m}{2} - k)$ ступенями свободи, де n-обсяг початкової вибірки, k-кількість регресорів.

Таким чином, якщо $P < F_{\alpha}(\frac{n-m}{2} - k, \frac{n-m}{2} - k)$, то гіпотеза про існування гетероскедастичності буде відхилена із рівнем довіри α. Якість тесту залежить від кількості відкинутих даних. Якщо відкинути забагато даних, то залишиться мало інформації для адекватного оцінювання. Якщо ж відкинути замало даних, то різниця між RSS_1 та RSS_2 буде невеликою.

Також даний тест можна реалізовувати і за різного обсягу вибірок для двох авторегресій. В такому випадку

$$P = \frac{RSS(1)(m1-k)}{RSS(2)(m2-k)} \quad (1.6)$$

де m1 та m2-обсяг вибірок.

Тест Бройша-Пагана

Тест Бройша-Пагана застосовується при припусканні, що дисперсія залишків залежить від декількох змінних.

Як і у попередніх тестах, розглядається нульова гіпотеза щодо гомоскедастичності процесу та альтернативна гіпотеза щодо його гетероскедастичності.

Розглянемо наступну регресію

$$X(t)=x^t(t)\,\beta + \varepsilon(t) \qquad (1.7)$$

Де $x^т$-вектор регрессорів

Припускається, що гетероскедастичність має наступну форму

$$E[\varepsilon(t)] = 0$$

$$D[\varepsilon(t)] = E[\varepsilon^2(t)] = \sigma^2 = h(\alpha z^t(t)) \qquad (1.8)$$

де $\alpha = [\alpha_1\ \alpha_2\ \dots\ \alpha_p]$ - вектор невідомих коефіцієнтів

$z^t(t) = [1\ z_1\ z_2\ \dots\ z_p]$ – вектор змінних(тих, від яких, за припущенням, залежить дисперсія залишків

h – деяка невід'ємна функція(наприклад, квадрат)

Нульова гіпотеза H_0: $\alpha_1 = \alpha_2 = \dots = \alpha_p = 0$

Отже,

$\sigma^2_\varepsilon = h(\alpha_1) = \text{const}$

Таким чином, алгоритм тесту зводиться до наступного:

1. Оцінити параметри регресії за допомогою методу МНК, знайти залишки

$$e(t) = x(t)-x^т(t)\beta \qquad (1.9)$$

Та дисперсію

$$\sigma^2_e = \sigma^2_\varepsilon = \frac{RSS}{n} \quad (1.10)$$

2. Знайти оцінки регресії $e(t)/\sigma^2_e$ на z(t) за допомогою ЗМНК та обчислити значення похибки ESS за формулою:

$$ESS = \beta^T X^T X\beta - \mu_y \quad (1.11)$$

де $\mu_y -$ середнє значення послідовності X(t).

3. Статистика тесту вираховується як $\frac{RSS}{2}$ допоміжної регресії і має асимптотично розподіл хі-квадрат із p-1 степенями свободи. Отже, якщо $\frac{RSS}{2} > Xi^2_\alpha(p-1)$, то гіпотеза щодо гомоскедастичності буде відхилена із рівнем значущості α.

4. Також можна використати іншу статистику, а саме nR^2 для тієї ж допоміжної регресії. Дана статистика при прийнятній нульовій гіпотезі матиме розподіл хі-квадрат із p-1 ступенями свободи. Обидва підходи є асимптотично еквівалентними.

Тест Глейзера

Нехай маємо регресію вигляду

$$X(t)=x^t(t)\,\beta + \varepsilon(t) \quad (1.12)$$

1. Оцінюємо параметри регресії методом МНК

2. Знаходяться похибки

3. Будуються наступні допоміжні регресії

$$|e(t)| = a + bx^{T}(t) + \varepsilon(t)$$

$$|e(t)| = a + b*sqrt(x^{T}(t)) + \varepsilon(t)$$

$$|e(t)| = a + b/sqrt(x^{T}(t)) + \varepsilon(t)$$

4. Перевіряється значущість показника Р для кожної із регресій(значущість показника Р перевыряеться за допомогою t або F статистики). Якщо показник є значущим хоча б для однієї із регресій, то приймається гіпотеза щодо гетероскедастичності процесу.

Тести на наявність тренду

Одним із найпопулярніших тестів для перевірки на наявність тренду є тест Дікі-Фуллера, який перевіряє наявність одиничного кореня

Одиничний корінь - кажуть, що часовий ряд має одиничний корінь, якщо його перші різниці утворюють стаціонарний часовий ряд.

$$\Delta X(t) = X(t)-X(t-1) \qquad (1.13)$$

Дана умова позначається як X(t)~I(1), тобто процес має перший порядок інтегрованості.

За допомогою тесту Дікі-Фуллера перевіряється значення коефіцієнта a у авторегресії першого порядку

$$X(t) = a*X(t-1) + \varepsilon(t) \qquad (1.14)$$

Якщо значення |a| = 1, то процес має одиничний корінь, тобто ряд є нестаціонаним. При |a| < 1 процес не має одиничного корень, відповідно ряд є стаціонарним. Випадок |a| > 1 відповідає вибуховому процесу(його траєкторія все швидше віддаляется від початкового стану) і виникнення таких процесів маловірогідне через неможливість дуже швидкого зростання показників за невеликі проміжки часу у економічних системах та середовищах.

Суть тесту Дікі-Фуллера заключається у формувані двух гіпотез:

Нульова гіпотеза - одиничній корінь існує, ряд нестаціонарний

Альтернативна гіпотеза - одиничного кореня не існує, ряд нестаціонарний

Рівняння 1.13 можна переписати у вигляді

$$\triangle X(t) = bX(t-1) + \varepsilon(t) \qquad (1.15)$$

де $\triangle X(t)$ віпдовідає виразу 1.13, а b=a-1. Таким чином, нульовій гіпотезі відповідає рівність нулю коефіцієнта b, а альтернативній - b < 1.

Тест пропонує використовувати три допоміжні регресії

1. Модель, що описується рівнянням (1.16) являє собою випадкове блукання, відсутній зсув на константу та тренд.

$$\triangle X(t) = bX(t-1) + \varepsilon(t) \qquad (1.16)$$

2. Модель, що описується рівнянням (1.17), має зсув на константу і не має тренду

$$\triangle X(t) = b_0 + bX(t-1) + \varepsilon(t) \quad (1.17)$$

3. Модель, що описується рівнянням (1.18), має як зсув на константу, так і лінійний тренд

$$\triangle X(t) = b_0 + b_1(t) + bX(t-1) + \varepsilon(t) \quad (1.18)$$

За тестом Дікі-Фуллера передбачається оцінювання параметра b та стандартної похибки оцінки якогось(можливо, не одного) рівняння із вище представлених(наприклад, методом МНК). Но основі отриманої оцінки та похибки будується відповідна t-статистика, значення якої порівнюється зі значеннями, наведеними у таблиці Дікі-Фуллера(для відповідного рівня значущості). Якщо значення статистики менше за значення із таблиці, то процес не має одиничного кореня, гіпотеза щодо нестаціонарності відхиляється і процесс визнається стаціонарним. У протилежному випадку отримується протилежний висновок.

Також існує розширений тест Дікі-Фуллера, який застосовують у випадку наявності автокореляційних залишків, тобто коли процес є авторегресією не першого, а більш високого порядку. Тест заключається у додаванні до тестових регресій лагів відповідних різностей. Розглянемо даний алгоритм на прикладі AR(2).

$$X(t) = b_1X(t-1) + b_2X(t-2) + \varepsilon(t) \quad (1.19)$$

Дану модель можна представити у наступному вигляді

$$\triangle X(t) = (b_1 + b_2 - 1)X(t-1) - b_2\triangle X(t-1) + \varepsilon(t) \quad (1.20)$$

Якщо часовий ряд має один одиничний корінь, то перші різниці є стаціонарними. Нульовою гіпотезою є нестаціонарність $\triangle X(t-1)$, отже, якщо коефіцієнт при ньому дорівнює нулю, то отримуємо протиріччя. Таким чином, із припущенні щодо інтегрованості першого порядку часового ряду витікає, що $b_1 + b_2 - 1 = 0$. Отже, для перевірки на наявність одиничного кореня(і, відповідно, тренда), слід провести звичайний тест Дікі-Фулера для коефіцієнту $b_1 + b_2 - 1$, додавши у тестову регресію лаг першої різниці залежної змінної(тобто X).

Тести на нелінійність

Тест Рамсея

Тест Рамсея використовують з метою перевірки специфікації(тобто правильності) моделі. Він дозволяє визнчити, чи допомогає нелінійна комбінація оціненого значення залежної змінної краще пояснити зміни самої змінної. Якщо якість оцінки покращується, то модель визнається специфічно некорректною і потребує перегляду. Якщо ж якість оцінки не покращується то модель визнається прийнятною.

Розглянемо алгоритм тесту. Нехай маємо наступну регресію:

$$Y(t) = b_0 + b_1X_1(t) + b_2X_2(t) + \varepsilon(t) \quad (1.21)$$

Будуємо допоміжну регресію

$$Y(t) = b_0 + b_1X_1(t) + b_2X_2(t) + a_1Y(t)^2 + a_2Y(t)^3 + \ldots + a_pY(t)^{p+1} + \varepsilon(t) \quad (1.22)$$

H_0: $a_1 = a_2 = \ldots = a_p = 0$

H_1: хоча б один із коефіцієнтів не рівний нулю

Для перевірки гіпотези використовується статистика Фішера

$$F = \frac{RSS_R - \frac{RSS_{UR}}{l}}{\frac{RSS_{UR}}{n-k}} \quad (1.23)$$

де RSS_R – сумма квадратів залишків для вихідної моделі

RSS_{UL} – сумма квадратів залишків для допоміжної моделі

l – кількість регресорів, що є степенями прогнозованого значення залежної змінної

n – кількість спостережень(обсяг вибірки)

k – кількість регресорів у допоміжній моделі, у розглянутому випадку k = p + + 2

Якщо значення отриманої статистики більше за критичне значення F(p, n-p) для обраного рівня значущості, то гіпотеза H_0 щодо корректності специфікації моделі відхиляється, є пропущені важливі регресори.

Також даний тест дозволяє визначити чи достатньо якісно залежна змінна описується поточним набором регресерів, чи може деякі важливі та впливаючі на змінну регресори, даних про які немає не були включені до моделі. Ідея методу полягає у тому, що не включені важливі регресори будуть впливати на значення Y(t) шляхом впливання на включені до моделі регресори і таким чином значення степенів залежної змінної буде містити деяку інформацію щодо значущості цього впливу. І навпаки, за допомогою

даного тесту можно перевірити наявність непотрібних або надлишкових регресорів.

тест Бокса-Кокса

Даний тест використовується для вибору форми моделі між лінійною та логарифмічною. Алгоритм тесту наступний:

1) Перетворюємо залежну змінну насупним чином(метод Зарембкі):

$$Y^{*}(t) = \frac{Y(t)}{\left(Y(1)*Y(2)*\ldots*Y(n-1)*Y(n)\right)^{\frac{1}{n}}} = \frac{Y(t)}{e^{((\ln(Y(1))+\ln(Y(2))+\cdots+\ln(Y(n-1))+\ln(Y(n)))/n)}} \qquad (1.24)$$

2) Розраховуємо нові змінні(перетворення Бокса-Кокса) при значеннях λ від 0 до 1

$$Y(t)_{BC} = \frac{Y(t)^{*\lambda} - 1}{\lambda}$$

$$X(t)_{BC} = \frac{X(t)^{\lambda} - 1}{\lambda} \qquad (1.25)$$

3) Розраховується регресії для всіх значень(із певним кроком) λ від 0 до 1:

$$Y(t)_{BC} = b_1 + b_2 X(t)_{BC} + \varepsilon(t)$$

4) Знаходиться мінімум суми квадратів залишків та обирається та із крайніх регресій (λ = 0 або λ = 1, тобто лінійну чи логарифмічну відповідно), до якої ближче точка мінімуму.

1.4 Огляд популярних систем для аналізу фінансово-економічних процесів

Прогнозування і методи для нього дуже поширені у наш час, адже хочеться мати точний і вірогідний прогноз. Для цього існують відповідні комп’ютерні програми. Бажано мати потужний технічний інструментарій для проведення різноманітних тестів, аналізів, графічних представлень даних та різноманітних зв’язків між компонентами, але в той же час пріорітетним є зручність інтерфейсу та простота роботи із програмою.

Розглянемо популярні комп’ютерні системи предназначені для роботи у сфері економічних процесів – Econometric Views(Eviews) і Statistica, та порівняємо їх.

Statistica має модульну структуру, тобто складається із різних частин(модулів), кожна з яких використовується для вирішення певного конкретного типу завдань, а саме: аналіз часових рядів і прогнозування, нелінійне оцінювання, непараметрична статистика, моделювання структурними рівняннями, множинна регресія, факторний аналіз, дисперсійний аналіз, функціональний аналіз. Певну кількість модулів, а саме планування експерименту, аналіз процесів, контроль якості, об'єднано в єдину групу, що має назву "Промислова статистика".

При завантаженні пакету програм Statistica і при створенні нового файлу з'являється таблиця, в якій стовпці є змінними, а рядки – спостереженнями. Введеня даних у програмі Statica є доволі зручним, адже процес схожий на введеня даних в Excel, із яким знайомі багато людей.

Дана комп’ютерна система дозволяє імпортувати дані з інших Windows, MacOS, Linux і навіть DOS програм, наприклад: MS Excel, MS Access, Foxpro, Paradox, dbase, CSV, SPSS, а також з файлів *.txt.

Додаток Eviews, на відміну від Statistica, є не модульною системою, а містить так зване вікно робочого файлу. Робоче вікно має об’єктну структуру, що дозволяє працювати із різними типами інформації одночасно. Управління об'єктами здійснюється за допомогою процедур, які у свою чергу можуть самі створювати нові об'єкти.

Кожен із об'єктів має власний конкретний вигляд інформації, а саме: графіки і діаграми, моделі, ряд даних, коефіцієнти, таблиці, результати обчислень тощо.

Крім того, система Eviews має командну строку, куди вводяться деякі команди, за допомогою яких можна проводити статистичний аналіз даних. Існує можливість зберігати команди в окремому файлі, що дає користувачу можливість переглянути історію виконуваних дій.

В Eviews , на відміну від програми Statistica, необхідно задати формат даних перед їх введенням, після чого створити об'єкт типа ряд, визначити кількість змінних і спостережень. Пакет Eviews дозволяє працювати з вісьмома типами даних (річні, піврічні, квартальні, місячні, тижневі (5 днів), тижневі (7 днів), щоденні і недатовані спостереження). Програма Eviews також надає можливість імпортувати дані із інших програм, наприклад: MS Access, Gauss, ODBS, SAS, SPSS, MS Excel, Stata, ACSII, HTML. З огляду на те, що процедура введення і опису даних в програмі Eviews складніша ніж в Statistica, краще за можливістю користуватися саме імпортом даних із інших джерел.

1.4.1 Графіки та основні статистики

Важливою складовою при аналізі та моделюванні є можливість графічного представлення отриманих результатів. Для спрощення процесу

візуалізації вихідних параметрів моделі і її кінцевих результатів як у Eviews, так і у Statistica надана можливість побудови різних графіків, діаграм, спектрограм, корелограм і тому подібного. Перегляд вихідних даних у графічному представлення у обох системах здійснюється за допомогою відповідних програм командного рядка.

Перегляд результатів у вигляді графіків в програмі Statistica відбувається за допомогою відповідних кнопок з робочого модуля, до того ж додаток надає можливість задати автоматичну побудову графіка після кожної проведеної процедури, а також є можливість вибору масштабу. Побудова корелограм АКФ і ЧАКФ в пакеті Statistica виконується лише в різних вікнах і вказані межі білого шуму. Щодо Eviews, у ньому можна переглянути відповідні корелограми і в одному вікні, проте межі білого шуму там не вказуються. Результати моделювання в Eviews в графічному режимі можна подивитися скориставшись командами основного меню. Для перегляду числових характеристик досліджуваних даних (середнє значення стандартне відхилення, ексцес, вірогідність ,тощо) в Statistica треба зайти в окремий модуль ‘’Основні статистики/Таблиці”(Basic Statistics/tables), але вагомим плюсом є наявність критичних значень для різних розподілів.

У додатку Eviews ці дії здійснюється за допомогою безпосередньо команд меню. Обидва пакети мають ідентичні набори описових статистик.

1.4.2 Лінійна регресія

Інструментарій для оцінки коефіцієнтів однофакторної та багатофакторної лінійної регресії у програмі Statistica знаходиться в окремому модулі “Множинна регресія” (Multiple regression). Результати представлені в діалоговому вікні, там представлені: коефіцієнти, оцінені за допомогою МНК, статистики Стьюдента оцінки значущості коефіцієнтів, коефіцієнт детермінації, статистика Фішера оцінки значущості регресії,

коефіцієнт кореляції (матриця кореляцій), статистика Дарбіна-Уотсона. Серьозними недоліками додатка Statistica є: Оцінка коефіцієнтів простої регресії лише за допомогою МНК; Оцінка на гетероскедастичність проводиться в окремому модулі(за допомогою тесту Спірмена в модулі “Непараметричні статистики”).

Інструментарій Eviews предоставляє ширші можливості для оцінки регресій, - він дозволяє проводити оцінку регресії не тільки за допомогою МНК, але ще й за допомогою ЗМНК і нелінійного МНК, також методом максимальної правдоподібності. Для вибору потрібного методу досить набрати назву методу в командному рядку при оцінці коефіцієнтів моделі. Eviews предоставляє можливість робити поправку на гетероскедастичність з урахуванням характеру залежності помилок від незалежної змінної. За замовчанням гетероскедастичність визначається тестом Уайта, проте за допомогою відповідних команд можна обрати інші тести(Глейзера, Парку тощо).

Програма Statistica дозволяє вирішити проблему мультиколлінеарності регресорів двома способами: за допомогою гребеневої регресії та за допомогою методу головних компонент(модуль “Факторний аналіз”(Factor Analysis).

1.4.3 Нелінійна регресія

У програмі Statistica методи для оцінки нелінійної регресії можна знайти в окремому модулі “Нелінійне оцінювання”. Можна скористатися як запропонованими видами залежності(регресія, кусочно-лінійна регресія, регресія експоненційного зростання, логіт/пробіт), так і задати вигляд залежності самостійно. Коефіцієнти нелінійної регресії довільного вигляду оцінюються ітеративними методи, наприклад Хука-Джівса, квазі-

ньютонівський, симплексний та ін. В якості результатів оцінки виводиться статистика Фішера та індекс детерміації. Підібрати гладку функцію можна лиша на основі візуальних даних. У системі Eviews підібрати вигляд найкращої нелінійної функції можна на основі тесту Бокса-Кокса. Оцінка коефіцієнтів проводиться методами нелінійного МНК і зваженого МНК.

1.4.4 Моделі із дискретною залежною змінною

Оцінка моделей бінарного вибору(логіт/пробіт) легко здійснюжться в програмі Statistica в модулі “Нелінійне оцінювання”. Вихідними даними є R2-статистика, оцінені методом максимальної правдоподібності, параметри моделі, обмежена логарифмічна функція правдоподібності, логарифмічна функція правдо-подібності. Щодо Eviews, то він має інструментарій для побудови моделі не тільки бінарного, а і множинного вибору як з неврегульованими, так і з порядковими альтернативами. Для вибору потрібної моделі слід в полі вибору методу оцінювання обрати метод, відповідний шуканій моделі. Вихідними параметрами приймаються логарифмічні функції правдоподібності, псевдо-коефіцієнт детермінації та R2-статистика.

1.4.5 Моделювання стаціонарних процесів

Серьозною перевагою програми Eviews перед Statistica є наявність тестів на стаціонарність(Eviews дозволяє застосувати як розширений так і звичайний тест Дікі-Фулера). Система Statistica не має модуля для перевірки стаціонарності ряду і судити щодо його стаціонарності можна лише на основі аналізу графіка ряду. Інструментарій для побудови моделей АРКС в програмі Statistica знаходиться у модулі “Аналіз часових рядів/Прогнозування”.

1.4.6 Ідентифікація моделі

Програма Eviews дозволяє провести ідентифікацію моделі АРКС ще і за допомогою q-статистики (тест Люнга-Бокса). Також є можливість порівняти дві значимі моделі АРКС, тобто виробити їх селекцію, за допомогою критеріїв Акайке і Шварца. Ці критерії Eviews вираховує на основі мінімальності дисперсії помилки. Інструментарій для побудови моделей стаціонарних рядів із урахуванням зміни дисперсії(GARCH та ARCH) присутній лише у Eviews, де до того ж знайдену модель АРКС за допомогою ARCH методу можна протестувати на гетероскадестичність.

За допомогою системи Eviews можна провести специфікацію моделі виправлення помилки і векторної авторегресійної моделі (дослідження коінтеграції між декількома змінними). Для перевірки залучається процедура Йохансена, що визначає кількість векторів коінтеграції в групі тимчасових рядів і забезпечує оцінки максимальної правдоподібності векторів коінтеграції і векторів швидкості приведення.

1.4.7 Моделювання нестаціонарних рядів

Звичайне моделювання нестаціонарних рядів проводиться на основі моделі АРІКС, де порядок інтеграції є порядком взяття різниць. У системі Statistica за допомогою процесу перетворення ряду обчислюються різниці аж поки ряд не стане стаціонарним, щоправда, як уже було сказано, стаціонарність можна первірити лише за допомогою аналізу графіків, після чого ідентифікують і будують модель АРКС. У системі Eviews для аналогічних цілей використовується розширений тест Діки-Фуллера, до того ж перевірка на стаціонарність проводиться автоматично після взяття різниці певного порядку (потрібний порядок треба вказати в діалоговому вікні).

1.5 Висновки до розділу

У сучасному світі всі процеси підлягають певному математичному опису, причому всі із оточуючих нас процесів за своєю природою є нелінійними і нестаціонарними і підвласні впливу багатьох чинників. Не дивлячись на те, що деякі процеси можна описати за допомогою лінійних моделей, більшість із них потребує більш складного формулювання та моделювання. Виключенням не є і фінансово-економічна складова навколишнього світу, при чому вона потребує все більш ретельного то якісного підходу для точніших прогнозів. У цьому розділі були приведені необхідні визначення, наприклад взагалі визначення та природу нелінійності та нестаціонарності, також розглянуто велику кількість різноманітних тестів для визначення властивостей процесу(нелінійнсть, гетероскедастичність, стаціонарність), наведено плюси і мінуси кожного тесту у порівняним один з одним, адже якість оцінки побудованої моделі багато в чому залежить від доречності вибору того чи іншого тесту. Якість же загалом побудовнаої моделі багато в чому залежить від правильності вибору її функціонального представлення та оцінки її параметрів. Для оцінки параметрів найпоширенішими є метод найменших квадратів та його модифікації(зокрема, зміщений МНК), іноді замість квадратичної функції обирають іншу невід’ємну функцію, метод максимальної правдоподібності тощо. Наприкінці було розглянуто дві найпопулярніші на сьогодні комп’ютерні системи - Statistica та Eviews.

РОЗДІЛ 2

ВИБРАНЫ МОДЕЛЫ ОПИСУ ННП

2.1 Метод групового урахування аргументів

Метод групового урахування аргументів – метод для моделювання багатопараметричних даних, який заключається у породженні і виборі моделей оптимальної складності(під оптимальністю мається на увазі кількість параметрів). Метод базується на послідовному селективному відборі моделей, на основі яких будуються складніші. За рахунок послідовного ускладнення моделей збільшується і їх точність.

Для роботи алгоритму слід визначити правила ускладнення моделі, систему опорних функцій, критерій селекції та методу регуляризації згідно зовнішнім критеріям.

--- Правила ускладнення моделі – деяке правило за яким змінюється модель на кожному кроці

--- Система опорних функцій – загальний вигляд досліджуваних моделей

--- Критерій селекції – певний параметр, на онснові якого вібдувається відсів непідходящих моделей(тобто модель визнається достатньо підходящою).

Загальний алгоритм методу можна представити у наступному вигляді:

1. Обробка даних відповідно до системи обраних опорних функцій
2. генерація множини моделей-претендентів;
3. обчислення критеріїв селекції, пошук моделі оптимальної складності.

Розглянемо поліноміальні алгоритми методу групового урахування аргументу. Початкові дані розділяються на перевірочну та навчальну вибірки. Нехай M – кількість вузлів інтерполяції; m — кількість членів повного

поліному регресії. Якщо m>M, то розв'язок можливо отримати тільки за допомогою МГВА.

Припустимо, що повний опис об'єкту задається деякою функціональною залежністю $y=f(x_1, x_2, \ldots, x_n)$. Замінимо такий опис кількома рядами часткових описів, які є першим рядом селекції.

$$y_1 = f(x_1, x_2),\ y_2 = f(x_1, x_3),\ \ldots,\ y_{n-1} = f(x_{n-1}, x_n),\ y_n = f(x_2, x_3),\ y_s = f(x_{n-1}, x_n)$$

де $s = C_n^2$.

Другий ряд селекції має наступний вигляд:

$$z_1 = f(y_1, y_2),\ z_2 = f(y_1, y_3),\ \ldots\ z_p = f(y_{s-1}, y_s)$$

де $p = C_n^2$.

Алгоритми МГВА відрізняються виглядом функції часткового опису f.

Розглянемо наступні види функції часткового опису:

лінійний частковий опис –

$$f(x_i, x_j) = A_0 + A_1x_i + A_2x_j \qquad (2.1)$$

неповний квадратичний частковий опис -

$$f(x_i, x_j) = A_0 + A_1x_i + A_2x_j + A_3x_ix_j \qquad (2.2)$$

повний квадратичний частковий опис -

$$f(x_i, x_j) = A_0 + A_1x_i + A_2x_j + A_3x_ix_j + A_4x_i^2 + A_5x_j^2 \qquad (2.3)$$

де A_i - нечітке число.

На кожний наступний рівень передається лише деяка кількість найкращих моделей.

Правило зупинки селекції: ряди селекції нарощуються поки критерій незміщеності розв’язків зменшується. При досягненні мінімуму селекцція припиняється.

Зовнішні критерії оптимальності

Під час селекції моделей використовуються критерії регулярності та незміщеності. Ці критерії визначаються на основі перевірочної вибірки.

Вибірка N поділяється не навчальну вибірку N_A, за допомогою якої проводиться оцінка параметрів моделі, та перевірочну вибірку N_B, на підставі якої обирається оптимальна модель.

Середньоквадратичне відхилення на перевірочній вибірці визначається як коефіцієнт регуляторності, тобто:

$$Kr = \sum_{t \epsilon N_B} (y_t^m - y_t)^2 / N_в \qquad (2.4)$$

де Kr – коефіцієнт регуляції.

Зважаючи на те, що на невеликих проміжках точність апроксимації буде зберігатися, критерій є підходящим варінтом для короткострокового прогнозу, адже відхилення вихідних даних буде неістотним. Незважаючи на те, що при процесі селекції можуть виявитися відкинутими важливі змінні, їх вплив все одно буде частково врахованим через інші змінні.

Критерій незміщеності моделі(також критерій несуперечливості). Даний критерій базується на тому, що для моделі поведінки одного й того самого об’єкту по різним вибіркам даних за умови рівних умов їх отримання мають виявитися близькими одна до одної.

Цей критерій визначається наступним чином:

$$Un = \frac{1}{R_1 + R_2} * \sum_{r=1}^{R_1 + R_2} (z_r^* - z_r^{**})\text{^}2 \qquad (2.5)$$

де R_1, R_2 – обсяг першої та другої підвибірок даних;

z_r^*, z_r^{**} — значення прогнозу першої та другої моделі.

Для відбіру моделей на етапі селекції було вирішено скористатися критерієм, що представляє собою лінійну комбінацію двох розглянутих критеріїв у наступному вигляді:

$$L = \alpha * Kr + (1-\alpha)*Un \quad (2.6)$$

де α – ваговий коефіцієнт, тобто α є [0;1]

Побудова часткової моделі нечіткого МГВА(НМГВА)

Для побудови даної моделі скористаємося моделлю лінійної інтервальної регресії, яка має насупний вигляд:

$$Y = A_1z_1 + A_2z_2 + \ldots + A_nz_n \quad (2.7)$$

Де A_i – нечіткі числа, z_i - деякі змінні.

Змінні z_i відповідають змінним x_i та x_j для квадратичної часткової моделі НМГВА наступним чином:

$$z_1 = 1, z_2 = x_i, z_3 = x_j, z_4 = x_i x_j, z_5 = x_i^2, z_6 = x_j^2.$$

Метод оцінювання даної моделі є наступним. Нехай відомо M спостережень 1 залежної змінної(y) від n незалежних(z). $z_i=(z_{i1},\ldots,z_{iM})$ – вихідний вектор спостережень незалежних змінних, $y=(y_1,\ldots,y_M)$ — вихідний вектор спостережень залежної змінної.

Перед нами стоїть задача оцінити коефіцієнти рівняння, тобто числа A_i.

Постановка задачі залежить від вигляду чисел A. Розглянемо нечіткі числа з наступними функціями належності:

- трикутною,
- гаусівською,
- дзвоноподібною,

та побудуємо для них відповідні моделі.

Першими розглянемо нечіткі числа з трикутною функцією належності. Нехай B_i — нечітке число з трикутною функцією належності. Тоді його функція належності має вигляд:

$$\mu(B_i) = \begin{cases} \frac{x-a}{b-a}, & x \in [a; b] \\ \frac{c-x}{c-b}, & x \in [b; c] \end{cases}$$

число B_i можна записати у вигляді центру a_i та ширини c_i: $B_i = (a_i; c_i)$.

Отже, Y розраховується за наступною формулою:

$$Y = (\sum a_i z_i; \sum c_i z_i) \qquad (2.8)$$

Відношення вкладення двох інтервалів B_i та B_j ($B_i \subset B_j$) задається наступною системою нерівностей:

$$\begin{cases} a_j - c_j \geq a_i - c_i \\ a_j + c_j \leq a_i + c_i \end{cases}$$

Тоді необхідну нам оціночну лінійну інтервальну модель можна побудувати наступним чином:

1. Задані значення y_i включаються в оціночний інтервал y_i^*.
2. Ширина оціночного інтервалу має бути найменшою.

Ці вимоги зводяться до наступної задачі лінійного програмування (для n-ї точки спостереження):

$$\min(c_0 + c_1*|x_i| + c_2*|x_j| + c_3*|x_i x_j| + c_4*|x_i^2| + c_5*|x_j^2|)$$

$$a_0 + a_1*x_i + a_2*x_j + a_3*x_i x_j + a_4*x_i^2 + a_5*x_j^2 -$$

$$- (c_0 + c_1*|x_i| + c_2*|x_j| + c_3*|x_i x_j| + c_4*|x_i^2| + c_5*|x_j^2|) \leq y_k$$

$$a_0 + a_1*x_i + a_2*x_j + a_3*x_i x_j + a_4*x_i^2 + a_5*x_j^2 +$$

$$+ (c_0 + c_1*|x_i| + c_2*|x_j| + c_3*|x_i x_j| + c_4*|x_i^2| + c_5*|x_j^2|) \geq y_k$$

$$c_p \geq 0,\ p = \overline{0,5}$$

З цього випливає, що маючи дані по M спостереженням щодо існуючих величин, завдання пошуку коефіцієнтів моделі формулюється наступним чином(для всіх точок спостереження).

$\min(c_0*M + c_1*\sum_{n=1}^{M} |x_i| + c_2*\sum_{n=1}^{M} |x_j| + c_3*\sum_{n=1}^{M} |x_i * x_j| +$

$+ c_4*\sum_{n=1}^{M} |x_i^2| + c_5*\sum_{n=1}^{M} |x_j^2|$

$a_0 + a_1*x_i + a_2*x_j + a_3*x_ix_j + a_4*x_i^2 + a_5*x_j^2 -$

$- (c_0 + c_1*|x_i| + c_2*|x_j| + c_3*|x_ix_j| + c_4*|x_i^2| + c_5*|x_j^2|) \leq y_k$

$a_0 + a_1*x_i + a_2*x_j + a_3*x_ix_j + a_4*x_i^2 + a_5*x_j^2 +$

$+ (c_0 + c_1*|x_i| + c_2*|x_j| + c_3*|x_ix_j| + c_4*|x_i^2| + c_5*|x_j^2|) \geq y_k$

де n – номер спостереження, тобто $n = \overline{0, M}$

$c_p \geq 0$

$p = \overline{0,5}$

Перед нами стоїть задача мінімізувати область зміни вихідних значень у шляхом знаходження таких a_i та $c_i(i = \overline{0,5})$, тобто центру та ширини інтервалу відповідно, за яких розсіювання величини Y є мінімальним та за умови потрапляння усіх значень Y у цей інтервал. Така задача є задачею лінійного програмування, щоб її розв’язати перейдемо до двоїстої задачі, сформульованої наступним чином:

$\max(\sum_{n=1}^{M} y_n * \delta_{n+M} - \sum_{n=1}^{M} y_n * \delta_n)$

$\sum_{n=1}^{M} \delta_n - \sum_{n=1}^{M} \delta_{n+M} = 0$

$\sum_{n=1}^{M} X_{ni} * \delta_n - \sum_{n=1}^{M} X_{ni} * \delta_{n+M} = 0$

…

$\sum_{n=1}^{M} X_{kj}^2 * \delta_n - \sum_{n=1}^{M} X_{kj}^2 * \delta_{n+M} = 0$

$\sum_{n=1}^{M} \delta_n + \sum_{n=1}^{M} \delta_{n+M} \leq M$

$\sum_{n=1}^{M} |X_{ni}| * \delta_n - \sum_{n=1}^{M} |X_{ni}| * \delta_{n+M} \leq \sum_{n=1}^{M} |X_{ni}|$

…

$\sum_{n=1}^{M} |X_{kj}^2| * \delta_n - \sum_{n=1}^{M} |X_{kj}^2| * \delta_{n+M} = \sum_{n=1}^{M} |X_{kj}^2|$

$\delta_n \geq 0, \text{n} = \overline{1,2,\dots,M}$

Щоб знайти оптимальні значення змінних d_i та a_i і визначити модель залежності треба розв'язати поставлену задачу сиплекс-методом та знайти оптимальні значення двоїстих змінних. так як при $\delta_k = 0, k = \overline{1,2 \cdot M}$ всі обмеження виконуються, то така задача має розв'язок

Наступними розглянемо нечіткі числа із гауссівською функцією належності.

Нехай B-нечітке число із гауссівською функцією належності. Тоді його фунцкія належності має наступний вигляд:

$$\mu_B(x) = e^{-\frac{1}{2} \cdot \frac{(x-a)^2}{c^2}}$$

Задамо число B парою чисел a та c, тобто B(a, c), де a – центр, c –ширина функції. Тоді сумою двох нечітких чисел $B(a_1, c_1)$ та $B(a_2, c_2)$, визначеною принципомом узагальнення Заде, буде деяке нечітке число $B_3(a_1 + a_2, c_1 + c_2)$.

Тоді Y розрахується за наступною формулою:

$$\text{Y} = (\sum a_i * z_i, \ \sum c_i * z_i) \qquad (2.10)$$

Нехай оціночна лінійна інтервальна модель для часткової моделі НМГВА має вигляд (2.7). Тоді необхідно знайти такі нечіткі числа A_i, тобто такі a_i і c_i, щоб:

1. Спостереження y_k належало оціночній множині Y_k як мінімум зі ступінню α такою, що. α є [0;1].
2. Ширина оціночного інтервалу рівня α була найменшою.

Розглянемо спочатку другу вимогу.

Ширина оціночного інтервалу рівня α:

$$d_\alpha = y_r - y_l = 2 \cdot (y_r - a)$$

$(y_r - a)$ знайдемо із умови:

$$\alpha = \exp\left\{ -\frac{1}{2} \cdot \frac{(y_r - a)^2}{c^2} \right\}$$

Тоді $\ln(\alpha) = -\frac{1}{2} * \frac{(y_r - a)^2}{c^2}$

$$y_r - a = c \cdot \sqrt{-2 \cdot \ln \alpha}$$

Або

$$d_\alpha = 2 * c * \sqrt{-2 * \ln(a)}$$

У такому разі цільову функцію можна представити у наступному вигляді:

$\min\sum_{n=1}^{M} c_{\alpha}^{n} = \min\sum_{k=1}^{M} 2c_k * \sqrt{-2 * \ln(a)} = 2\sqrt{-2\ln(a)} \min\sum_{k=1}^{M} c_k = 2\sqrt{-2\ln(a)} \min\sum_{k=1}^{M}\sum_{i=1}^{n} c_i * |z_{ik}|$

$2\sqrt{-2\ln(a)}$ – додатнє число, яке не впливає на значення d_i, що мінімізує функцію. Тому маємо право скоротити на цю константу, що і зробимо

Остаточно отримуємо наступне:

$\min(c_0*M + c_1*\sum_{k=1}^{M}|x_i| + c_2*\sum_{k=1}^{M}|x_j| + c_3*\sum_{k=1}^{M}|x_i * x_j| + c_4*\sum_{k=1}^{M}|x_i^2| + c_5*\sum_{k=1}^{M}|x_j^2|)$

Тепер розглянемо першу вимогу: $\mu(y_k) \geq \alpha$.

Вона є еквівалентною наступному:

$$\exp\left\{-\frac{1}{2}\cdot\frac{(y_k-a_k)^2}{c_k^{\,2}}\right\} \geq \alpha$$

$$\frac{1}{2}\cdot\frac{(y_k-a_k)^2}{c_k^{\,2}} \geq \ln\alpha$$

$$\frac{(y_k-a_k)^2}{c_k^{\,2}} \leq -2\ln\alpha$$

$$\left|\frac{(y_k-a_k)}{c_k}\right| \leq \sqrt{-2\ln\alpha}$$

$$\begin{cases} y_k - a_k \leq c_k\sqrt{-2\ln\alpha} \\ y_k - a_k \geq -c_k\sqrt{-2\ln\alpha} \end{cases}$$

$$\begin{cases} a_k + c_k\sqrt{-2\ln\alpha} \geq y_k \\ a_k - c_k\sqrt{-2\ln\alpha} \leq y_k \end{cases}.$$

Так як

$$a_k = \sum_{i=1}^{n} a_i \cdot z_{ki}\,,\quad c_k = \sum_{i=1}^{n} c_i \cdot |z_{ki}|\,.$$

То поточна задача зводиться до наступної задачі лінійного програмування:

Для розв'язання даної задачі перейдемо до двоїстої задачі. Вона запишеться наступним чином:

$$\min\left(c_0 \cdot M + c_1 \cdot \sum_{k=1}^{M} |X_i| + c_2 \cdot \sum_{k=1}^{M} |X_j| + c_3 \cdot \sum_{k=1}^{M} |X_i \cdot X_j| + c_4 \cdot \sum_{k=1}^{M} X_i^2 + c_5 \cdot \sum_{k=1}^{M} X_j^2 \right)$$

$$\alpha_0 + a_1 \cdot X_i + \mathrm{K} + a_5 \cdot X_j^2 + (c_0 + c_1 \cdot |X_i| + \mathrm{K} + c_5 \cdot |X_j^2|)\sqrt{-2\ln\alpha} \geq y_k$$

$$\alpha_0 + a_1 \cdot X_i + \mathrm{K} + a_5 \cdot X_j^2 - (c_0 + c_1 \cdot |X_i| + \mathrm{K} + c_5 \cdot |X_j^2|)\sqrt{-2\ln\alpha} \leq y_k$$

$$k = \overline{1,M}\,,\ c_i > 0, i = \overline{0,5}$$

Перейдемо до наступної двоїстої задачі:

$$\max\left(\sum_{k=1}^{M} Y_k \cdot \delta_{k+M} - \sum_{k=1}^{M} Y_k \cdot \delta_k\right)$$

$$\sum_{k=1}^{M} \delta_k - \sum_{k=1}^{M} \delta_{k+M} = 0$$

$$\sum_{k=1}^{M} X_{ki}\delta_k - \sum_{k=1}^{M} X_{ki} \cdot \delta_{k+M} = 0$$

....

$$\sum_{k=1}^{M} X^2{}_{kj}\delta_k - \sum_{k=1}^{M} X^2{}_{kj} \cdot \delta_{k+M} = 0$$

$$\sum_{k=1}^{M} \delta_k + \sum_{k=1}^{M} \delta_{k+M} \leq \frac{M}{\sqrt{-2\ln\alpha}}$$

$$\sum_{k=1}^{M}\left|X_{ki}\right|\cdot\delta_k+\sum_{k=1}^{M}\left|X_{ki}\right|\cdot\delta_{k+M}\le\frac{\sum_{k=1}^{M}\left|X_{ki}\right|}{\sqrt{-2\ln\alpha}}$$

……

$$\sum_{k=1}^{M}\left|X^{2}{}_{kj}\right|\cdot\delta_k+\sum_{k=1}^{M}\left|X^{2}{}_{kj}\right|\cdot\delta_{k+M}\le\frac{\sum_{k=1}^{M}\left|X^{2}{}_{kj}\right|}{\sqrt{-2\ln\alpha}}\ .$$

$$\delta_k\ge 0, k=\overline{1,2\cdot M}$$

Щоб знайти оптимальні значення змінних d_i та a_i і визначити модель залежності треба розв'язати поставлену задачу сиплекс-методом та знайти оптимальні значення двоїстих змінних. так як при $\delta_k=0, k=\overline{1,2\cdot M}$ всі обмеження виконуються, то така задача має розв'язокНечіткі числа з дзвоноподібною функцією належності

Нехай B-нечітке число із дзвоноподібною функцією належності. Тоді його фунцкія належності має наступний вигляд:

$$\mu_B(x)=\frac{1}{1+\left(\frac{x-a}{c}\right)^2}$$

Задамо число B парою чисел a та c, тобто B(a, c), де a – центр, c –ширина функції. Тоді сумою двох нечітких чисел $B(a_1, c_1)$ та $B(a_2, c_2)$, визначеною принципомом узагальнення Заде, буде деяке нечітке число $B_3(a_1 + a_2, c_1 + c_2)$.

Маємо наступну задачу лінійного прграмування:

$$\min_{a_i,c_i}\left(c_1\cdot\sum_{k=1}^{M}\left|X_{k1}\right|+\ldots+c_n\cdot\sum_{k=1}^{M}\left|X_{kn}\right|\right)$$

$$\alpha_0 + a_1 \cdot X_i + \text{K} + a_5 \cdot X_j^{\,2} + (c_0 + c_1 \cdot |X_i| + \text{K} + c_5 \cdot |X_j^{\,2}|)\sqrt{\frac{1-\alpha}{\alpha}} \geq y_k$$

$$\alpha_0 + a_1 \cdot X_i + \text{K} + a_5 \cdot X_j^{\,2} - (c_0 + c_1 \cdot |X_i| + \text{K} + c_5 \cdot |X_j^{\,2}|)\sqrt{\frac{1-\alpha}{\alpha}} \leq y_k$$

$$k = \overline{1, M}$$

$$c_i > 0, i = \overline{0,5}$$

Отримуємо двоїсту задачу наступного вигляду:

$$\max(\sum_{k=1}^{M} Y_k \cdot \delta_{k+M} - \sum_{k=1}^{M} Y_k \cdot \delta_k)$$

$$\sum_{k=1}^{M} \delta_k - \sum_{k=1}^{M} \delta_{k+M} = 0$$

$$\sum_{k=1}^{M} X_{ki}\delta_k - \sum_{k=1}^{M} X_{ki} \cdot \delta_{k+M} = 0$$

....

$$\sum_{k=1}^{M} X^{\,2}{}_{kj}\delta_k - \sum_{k=1}^{M} X^{\,2}{}_{kj} \cdot \delta_{k+M} = 0$$

$$\sum_{k=1}^{M} \delta_k + \sum_{k=1}^{M} \delta_{k+M} \leq M \cdot \sqrt{\frac{\alpha}{1-\alpha}}$$

$$\sum_{k=1}^{M} |X_{ki}| \cdot \delta_k + \sum_{k=1}^{M} |X_{ki}| \cdot \delta_{k+M} \leq \sum_{k=1}^{M} |X_{ki}| \cdot \sqrt{\frac{\alpha}{1-\alpha}}$$

.....

$$\sum_{k=1}^{M} |X^{\,2}{}_{kj}| \cdot \delta_k + \sum_{k=1}^{M} |X^{\,2}{}_{kj}| \cdot \delta_{k+M} \leq \sum_{k=1}^{M} |X^{\,2}{}_{kj}| \cdot \sqrt{\frac{\alpha}{1-\alpha}}$$

$$\delta_k \geq 0, k = \overline{1,2 \cdot M}$$

Опис ряду селекції

Моделі, розглянуті вище є самооранізованими системами, тобто вони можуть змінювати свою структуру шляхом збільшення ступеня входження вихіних змінних та підстроювання коефіцієнтів моделі. В якості вигляду шуканої моделі виступав поліном, що залежав від двох змінних, тому на

першії ітерації було отримано C_n^2 нових моделей, серед яких, за допомогою розглянутих вище критеріїв обирається певна кількість найкращих та на їх основі проводять прогнозування(даний алгоритм називається однорядним методом МГВА). Проте можна збільшити кількість ітерацій, тобто на основі обраних на 1 ітерації моделей побудувати наступний ряд і отримати кращі прогнози(даний алгоритм називається багаторядним методом МГВА). Так як він є точніщим, розглянемо його докладніше.

Для роботи даного алгоритму необхідно ввести опорну функцію, за допомогою якої генеруватимуться нові моделі, а в якості аргументів вона має приймати функції, отримані на попередньому рівні . В розглянутому вище прикладі опорна функція задавалась як поліном 2-го степеня від двох змінних $Y_{ij}^{k+1} = f(Y_i^k, Y_j^k)$, де k-номер ряду.

На рисунку 2.1 представлено алгоритм роботи багаторядовогу алгоритму

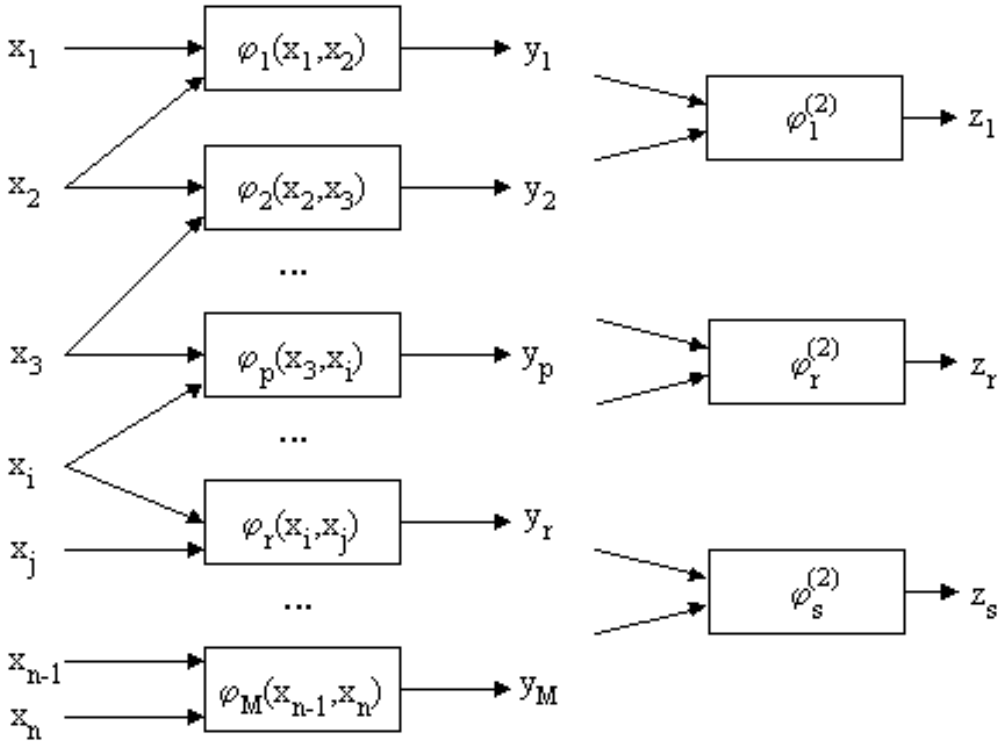


Рисунок 2.1 - Механізм роботи багаторядного алгоритму

Після кожної ітерації на поточному ряді за певним комплексним критерієм відбиралось деяка кількість кращих моделей, на основі яких відбуватиметься генерація моделей на наступному рівні. Критерієм зупинки процесу генерації є зростання обраного критерію на наступному ряді селекції.

2.2 Регресійні моделі процесів із трендами різних видів

Для початку нагадаємо визначення та сенс тренду.

Тренд – основна тенденція зміни значення часового ряду

Тренд може бути як стохастичним, так і детермінованим. Тобто функція тренду відповідно має чи не має у собі випадкової складової.

Розглянемо деякі популярні функції задання тренду

Логарифмічний тренд

Логарифмічний тренд задається наступною функцією:

$$Y_i = a*\ln(t_i)+b \qquad (2.11)$$

де x-регресори, b-початкове значення, a – значення, на яке змінюється наступне значення часового ряду(при $a > 0$ або $a < 0$ значення ряду збільшуються або зменшуються відповідно). В загальному випадку тренд може містити певну випадкову змінну, що має певну функцію розподілу.

Властивості логарифмічного тренду:

1. Якщо $b > 0$, то рівні зростають, але з уповільненням, а при $b <0$ рівні тренду зменшуються, також з уповільненням.

2. Абсолютні зміни рівнів по модулю завжди зменшуються з часом.

3. Прискорення абсолютних змін мають знак, протилежний самим абсолютним змінам, а по модулю поступово зменшуються.

4. Темпи зміни (ланцюгові) поступово наближаються до 100% при t -> ∞

Такий вигляд тренду використовується при прогнозуванні процесів, зміна значень яких з часом зменшується. Наприклад, при появі на ринку нового товару, швидкість зміни попиту на нього спочатку є великою, але з плином часу вона зменшується.

Гіперболічний тренд

Гіперболічний тренд має наступну функцію:

$$Y_i = a + \frac{b}{t} \qquad (2.12)$$

При b > 0 даний тренд відображає зростаючий процес із поступовим зменшенням росту, а при t < 0 – спадаючий процес із поступовим зменшенням росту. Також можна побачити, що при t -> ∞ складова $\frac{b}{t} \to 0$, а отже процес має асимптоту a. Саме в цьому криється основна відмінність такого тренду від логарифмічного, адже значення часового ряду при логарифмічному тренді, хоча і схожі за поведінкою на значення при гіперболічному тренді, теоретично можуть зростати або спадати необмежено.

Властивості гіперболічного тренду:

1. Абсолютний приріст, скорочення рівнів, пришвидшення абсолютних змін, темп змін - всі ці показники не є постійними. При b > 0 рівні уповільнено зменшуються, негативні абсолютні зміни, а також позитивні прискорення теж зменшуються, ланцюгові темпи змін ростуть і прагнуть до 100%

2. При $b < 0$ рівні уповільнено зростають, позитивні абсолютні зміни, а також негативні прискорення і ланцюгові темпи зростання уповільнено зменшуються, прагнучи до 100%.

Реальні умови процесів рідко можуть дозволити необмежене зростання значень, проте при прогнозуванні нас як правило цікавить поведінка процесу в межах якогось скінченого проміжку часу, тож використання логарифмічного тренду є цілком виправданим.

Степінний тренд

Такий тренд виглядає наступним чином:

$$Y_i = a_0 + a_1 * t^{a2} \qquad (2.13)$$

Степінний тренд дозволяє моделювати кілька типів процесів в залежності від значення коефіцієнтів. У випадку з позитивними значеннями a1 можна виділити наступні три ситуації:

1. $0 < a_2 < 1$ - зростання з уповільненням. Такий тип процесів частіше зустрічається в практичному прогнозуванні. Зростання в такому випадку є необмеженим, але з кожним наступним наглядом зміни показника стають все менше і менше.

2. $a_2 > 1$ - зростання з прискоренням. По суті, в такому випадку ми отримуємо зростаючу гілку дещо модернізованої параболи.

3. $a_2 < 0$ - зниження з уповільненням. В такому випадку одержуємо фактично гіперболу, що має асимптоту в точці a_0. При $a_2 \neq -1$ отримується дещо модернізована гіпербола, проте основні її властивості зберігаються.

Коефіцієнт a_2 являє собою коефіцієнт еластичності, який показує, на скільки відсотків зміниться Y зі зміною t на 1%.

Показниковий тренд

Функція даного тренду є наступною:

$$Y_i = a_0 * a_1^t \quad (2.14)$$

Тренд такого виду часто замінюють на експоненційний тренд

За допомогою такого тренду описуть або процеси із вибуховим ростом, або процеси, що спаданяють із замедленням.

Розглянувши функцію можна побчити, що в залежності від значення a_1 отримаємо 4 випадки:

1. При $a_1 > 1$ моделюється процес зростання з вибуховим прискоренням.

2. $0 < a1 < 1$ – вказує на процес уповільнення з наближенням до нуля.

3. $-1 < a1 < 0$ – маємо процес уповільнення з наближенням до нуля. До того ж, значення часового ряду чергуються за знаком.

4. $a1 < -1$ – моделюється знакозмінний необмежений ряд.

3 та 4 випадок через власну знакозмінність на практиці використовуються дуже рідко(наприклад, їх іноді використовують для моделювання сезонних процесів). До того ж процес із 1 випадку через свою вибуховість може зустрітись лише на достатньо короткому короткому часовому проміжку, адже природа реальних процесів рідко дозволяє необмежений вибуховий ріст.

Показниковий тренд на практиці часто зводять до **експоненційного.** Зробивши у (2.2.4) наступну заміну $b_1 = \ln(a)$, отримаємо:

$$Y_i = a_0 * \exp(b_1 * t) \quad (2.15)$$

Тут коефіцієнт t вказує на скільки відсотків Y змінюється з часом(з кожним наступним спостереженням у збільшується на наступну величину:

$$(\exp(b_1) - 1) * 100\% \quad (2.16)$$

Параметри кожної моделі можна оцінити за допомогою МНК

Поліноміальний тренд

Загальный вигляд поліноміального тренду є настпуним:

$$Y(t) = \alpha_0 + \sum_{i=1}^{k} \alpha_i t^i + \varepsilon(t) \quad (2.17)$$

Тренди такого типу використовються для опису процесів із великою кількістю проміжків зростання та спадання. Він підходить для аналізу великої вибірки даних нестабільної величини(обсяг продажів, біржові ціни тощо). Щодна із оглянутих вище функцій тренду не підійде для аналізу такого процесу, адже вони використовуються для більш-менш однонаправлених процесів. Чим більшою є степінь полінома, тим більше він має точок екстремуму, а отже може охопити більшу кількість проміжків зростання та спадання. Тож за наявності досить сильної нестабільності даних слід обирати поліном високої степені.

Найпростішим виглядом поліноміального тренду є **лінійний тренд**.

$$Y(t) = a_0 + a_1 * t \quad (2.18)$$

Такий вид тренду є найпростішим та інтуітивно зрозумілим. Він використовується для аналізу рядів, значення яких змінюються із приблизно постійною швидкістю. Наприклад, його часто використовують для аналізу та прогнозування обсягу продаж на невеликих проміжках часу(наприклад, сезон), причому товар не має бути новим, адже за появи нового товару на ринку обсяг попиту і, відповідно, продажу скоріш за все не буде лінійною функцією. Цю модель часто використовують у випадках, коли складніші моделі виявляються лише незначно точнішими, через її простоту. Також вона може бути основою для деяких складніших процесів.

Наступною моделлю є **параболічний тренд**, який задається наступною функцією:

$$Y(t) = a_0 + a_1 t + a_2 t^2 \qquad (2.19)$$

Сенс коефіцієнтів є наступним:

a_0 показує середній рівень тренду, який був на початку відліку

a_1 відповідає за середній приріст, який уже не є константою, а змінюється рівномірно із середнім прискоренням 2c, яке є константою і головним параметром моделі параболічного тренду.

Отже, тренд у вигляді поліному 2 порядку відображає тенденцію таких процесів, яким властиво приблизно постійне прискорення абсолютних змін рівнів. Таким, наприклад, можна вважати процес зміни популяції деякого міста за короткий проміжок часу.

Загалом, вибір степені тренду залежить від значень часового ряду. За допомогою параболи 3 порядку можна аналізувати процес, що має 2 точки

екстремуму, 4 порядку-3 точки екстремуму, n-ого порядку - (n-1) точок екстремуму. Але важливо розуміти, що степінь поліному не слід обирати близькою до кількості спостережень, адже незважаючи на те, що ми отримаємо дуже хорошу апроксимуючу модель, через свою складність вона із високою степінню вірогідності не підходитиме для прогнозування. Слід обирати степінь поліному вагомо меншу за кількість спострежень, адже тоді вона, хоча і не буде такою точною на тестових даних, зможе зберегти та відтворити загальну динаміку процесу.

Коефіцієнти поліноміальної моделі можна оцінювати за допомогою МНК

Система МНК-рівнянь для поліноміальної функції порядку k виглядає наступним чином.

$$\begin{matrix} n & \sum_n x_t & \dots & \sum_n x_t^k \\ \sum_n x_t & \sum_n x_t^2 & \dots & \sum_n x_t^{k+1} \\ \dots & \dots & \dots & \dots \\ \sum_n x_t^k & \sum_n x_t^{k+1} & \dots & \sum_n x_t^{2k} \end{matrix} * \begin{matrix} a_0 \\ a_1 \\ \dots \\ a_k \end{matrix} = \begin{matrix} \sum_n y_t \\ \sum_n x_t * y_t \\ \dots \\ \sum_n x_t^k * y_t \end{matrix}$$

, де n – кількість спостережень

2.3 Моделі гетероскедастичних процесів

Як уже було сказано раніше, гетероскедастичністю називається властивість процесу, яка заключається у непостійності дисперсії його випадкової помилки.

У цьому розділі буде розглянуто деякі існуючі моделі таких процесів.

ARCH(АРУГ)-модель

Ця модель була запропонована американським економістом Робертом Енглом у 1982 році. Її використовують в економетриці для прогнозування майбутньої волатильності економічних величин. За даною моделлю аналізу

часових умовна дисперсія ряду залежить від його минулих значень та випадкової складової.

Наступну модель називають Авторегресійною моделлю умовної гетероскедастичності порядку(тобто кількості попередніх величин, які застосовуються для прогнозування) q:

$$Y_t = \overline{X} * \overline{\beta} + \varepsilon_t$$

$$r_t = \sqrt{\alpha_0 + \sum_{i=1}^{q} \alpha_i * \varepsilon_{t-i}^2} \qquad (2.20)$$

де $\varepsilon_t \sim N(0;d)$, r_t – дисперсія випадкової помилки ε_t

Тобто, як бачимо, за цією моделлю припускається, що значення умовної дисперсії дійсно залежить лише від попередніх значень ряду(ε_t)

Квадратична функція вказує на те, що невеликі значення (менші за 1) будуть “зглажуватися”, а значення більше за 1 навпаки, будуть вносити ще більший вклад внаслідок піднесення їх до квадрату.

Для зручності запису позначимо величину $\sqrt{\alpha_0 + \sum_{i=1}^{q} \alpha_i * \varepsilon_{t-i}^2}$ як деяке h, тобто $\sqrt{\alpha_0 + \sum_{i=1}^{q} \alpha_i * \varepsilon_{t-i}^2} = h$

Розглянемо умовне матсподівання та умовну дисперсію ряду ε_t відносно минулих значень ряду:

$E(r_t | r_{t-1}, r_{t-2}, \ldots) = \sqrt{h_t} * E(\varepsilon_t | r_{t-1}, r_{t-2}, \ldots) = 0$

$D(r_t | r_{t-1}, r_{t-2}, \ldots) = h_t * D(\varepsilon_t | r_{t-1}, r_{t-2}, \ldots) = h_t * d$

Тобто умовне матсподівання величини ε_t залежно від минулих значень дорінює нулю, а умовна дисперсія дорівнює $h_t * d$.

Безумовне матсподівання також рівне нулю. Безумовна дисперсія дорівнює

$$D = \frac{D_\varepsilon}{1-\sum_{i=1}^{q} \alpha_i}$$

Отже, необхідною умовою стаціонарності є :

$$\sum_{i=1}^{q} \alpha_i < 1 \qquad (2.21)$$

Мінус даної моделі заключається в наступному: Дисперсія є величиною невід'ємною, а отже є необхідною наступна умова:

$$\alpha_n \geq 0 \ \forall \ n \in [0;q]$$

При частковому випадку при $d = 1$, тобто $\varepsilon_t \sim N(0;1)$, отримуємо:

$$E(r_t \mid r_{t-1}, r_{t-2}, \ldots) = 0$$

$$D(r_t \mid r_{t-1}, r_{t-2}, \ldots) = h_t$$

На практиці точно описати достатньо складний процес за допомогою даної моделі важко, тому було розроблено певні її модернізації.

GARCH-General ARCH(УАРУГ)

На відміну від АРУГ моделі при прогнозуванні на основі УАРУГ приймається залежність умовної дисперсії не лише від минулих значень ряду, а ще і від минулих значень самої дисперсії.

Модель наступного вигляду називають моделлю УАРУГ(p, q), де p-порядок значень дисперсії, q-порядок значень ряду.

$$Y_t = \overline{X} * \overline{\beta} + \varepsilon_t$$

$$r_t^2 = \alpha_0 + \sum_{i=1}^{q} \alpha_i * \varepsilon_{t-i}^2 + \sum_{j=1}^{p} \beta_j * r_{t-j}^2 \qquad (2.22)$$

Як бачимо, формула прогнозу має ще один доданок у порівнянні з АРУГ, а саме $\sum_{j=1}^{p} \beta_j * r_{t-j}^2$. Цей доданок якраз відповідає за внесок попередніх значень умовної дисперсії до її прогнозу.'

Маємо наступну необхідну умову стаціонарності для даного процесу:

$$\sum_{i=1}^{p} \beta_i + \sum_{n=1}^{q} \alpha_i < 1 \qquad (2.23)$$

Матсподівання процесу дорівнює нулю.

Безумовна дисперсія даного процесу є постійною і дорівнює

$$D = \frac{D_\varepsilon}{1 - \sum_{i=1}^{p} \beta_i - \sum_{i=1}^{q} \alpha_i}$$

Отже, якщо $\sum_{i=1}^{p} \beta_i + \sum_{n=1}^{q} \alpha_i = 1$, то процес є інтегрованим(IGARCH)

Антисимметричні GARCH-моделі

Даний клас моделей має на меті врахувати існуючу у реальних економічних процесах наступну асиметрію: погані новини мають сильніший вплив на волатильність ніж хороші. Розглянуті вище моделі не беруть це до уваги, адже використовують невід'ємнозначні функції(безпосередньо у розглянутому випадку-квадратичну), тож знак вхідних даних не грає ролі.

Антисиметрію ще називають ефектом важіля.

ПУАРУГ(Порогова УАРУГ) модель

Дана модель є прикладом антисиметричної моделі, специфікація моделі має наступний вигляд:

$$r_t^2 = \sum_{j=1}^{p} \beta_j * r_{t-j}^2 + \sum_{i=1}^{q} \alpha_i * \varepsilon_{t-i}^2 + \sum_{k=1}^{q} \gamma_i * \varepsilon_{t-k}^2 * \ I_{t-k}^- \qquad (2.24)$$

Як бачимо, тут є ще один доданок, а саме $\sum_{k=1}^{q} \gamma_i * \varepsilon_{t-i}^2 * \ I_{t-k}^-$, де I_{t-k}^- є індикатором значення Y, тобто:

$$\begin{cases} I_{t-k}^- = 1, \text{якщо } \varepsilon_t < 1 \\ I_{t-k}^- = 0, \text{якщо } \varepsilon_t > 1 \end{cases}$$

Саме цей доданок допомагає внести різницю у впливі негативного та позитивного шоку на процес. Як бачимо, при $\varepsilon_t < 1,$ тобто за умови негативного шоку значення ε_t впливає на значення $(\alpha_i + \gamma_i)$, а у випадку позитивного шоку лише на значення α_i. Отже, при $\gamma_i > 0$ негативний шок(погані новини) збільшуються волатильність процесу, а при $\gamma_i < 0$ - навпаки зменшують.

Експоненційна АУАРУГ(ЕУАРУГ)

Має наступну специфікацію

$$\mathrm{Ln}(r_t^2) = \alpha_0 + \sum_{j=1}^{p} \beta_j \ln(r_{t-j}^2) + \sum_{i=1}^{q} \alpha_i \left|\frac{\varepsilon_{t-i}^2}{r_{t-i}^2}\right| + \sum_{k=1}^{u} \gamma_k \left(\left|\frac{\varepsilon_{t-k}^2}{r_{t-k}^2}\right|\right) \quad (2.25)$$

Так як зліва знаходиться логарифм, то це гарантує невід'ємні значення умовної дисперсії. Присутність ефекту важеля перевіряється тестуванням гіпотези щодо $\gamma_k < 0$. Асиметрія присутня, якщо $\gamma_k \neq 0$.

Взагалі кажучи, значення випадкової помилки не обов'язково має гауссову функцію розподілу

2.4 Висновки

У 2 розділі дипломної роботи було розглянуто певні існуючі моделі нелінійних нестаціонарних процесів, методи їх побудови, наведено способи знаходження та оцінки їх коефіцієнтів. Також було розглянуто ситуації коли ці моделі доцільно застосовувати для отримання точнішої моделі та прогнозу. Наведені моделі будуть використовуватися для практичної частини дипломної роботи

РОЗДІЛ 3

ПОБУДОВА, АНЛІЗ ТА ПРОГНОЗУВАННЯ ПРОЦЕСІВ

3.1 Вступ

В цьому розділі мова буде йти про безпосередню побудову моделей демографічних процесів та побудова прогнозів на основі обраних моделей. На початковому етапі було проведено збір необхідних даних, що описують демографічні процеси в Україні, тепер необхідно провести їх попередній аналіз та приступати до процесу побудови моделей.

У цьому розділі буде розглянуто безпосередньо моделювання певних процесів та прогнозування на основі отримкних моделей. Загалом, алгоритм дій є наступним:

1. Попередній погляд на дані та їх обробка, поділ на навчальну та тестову вибірки.
2. Виділення основних характеристик процесу з метою побудови адекватнішої моделі. Проведення тестів на стаціонарність, аналіз корелограми та прийняття висновку щодо наявності авторегресії.
3. Проведення тестів на гетероскедастичність та прийняття рішення щодо вигляду моделі.
4. Вибір методу оцінювання коефіцієнтів моделі.
5. Побудова декількох конкуруючих моделей.
6. Оцінка якості моделі за певними критеріями.
7. Порівняння результатів моделей та вибір найкращої.
7. Прогнозування майбутніх значень та їх порівняння із фактичними.
8. Порівняння результатів прогнозу моделей (якщо більше ніж одна) та вибір найкращої.

3.2 Пошук даних та вибір середовища для роботи

Дані для роботи було знайдено на різноманітних інтернет ресурсах, наприклад, kaggle.

Для виконання роботи було вирішено скористатися комп'ютерною програмою Eviews, бо вона надає широкі можливості для для роботи із часовими рядами та має потужний інструментарій для їх аналізу. Вона дозволяє писати користувачу власні внутрішні програми за допомогою текстового вікна. Цією програмою часто користуються для роботи із часовими рядами.

3.3 Моделювання та прогнозування процесів

3.3.1 Перша вибірка даних

Ця вибірка показує значення певних погодніх параметрів(температуру, швидкість вітру, вологість, тиск) у період із 01/01/2017 до 04/24/2017

Нашою метою буде побудова адекватної моделі, за допомогою якої можна описати та спрогнозувати наступні значення температури залежно від інших наявних параметрів.

Дані було вирішено поділити наступним чином:

З 01/01/2017 по 04/01/2017– навчальна вибірка

З 04/02/2017 по 04/24/2017 – тестова вибірка

На рисунку 3.1 представлено графік процесу, із якого можна побачити значення температури на усьому проміжку спостережень.

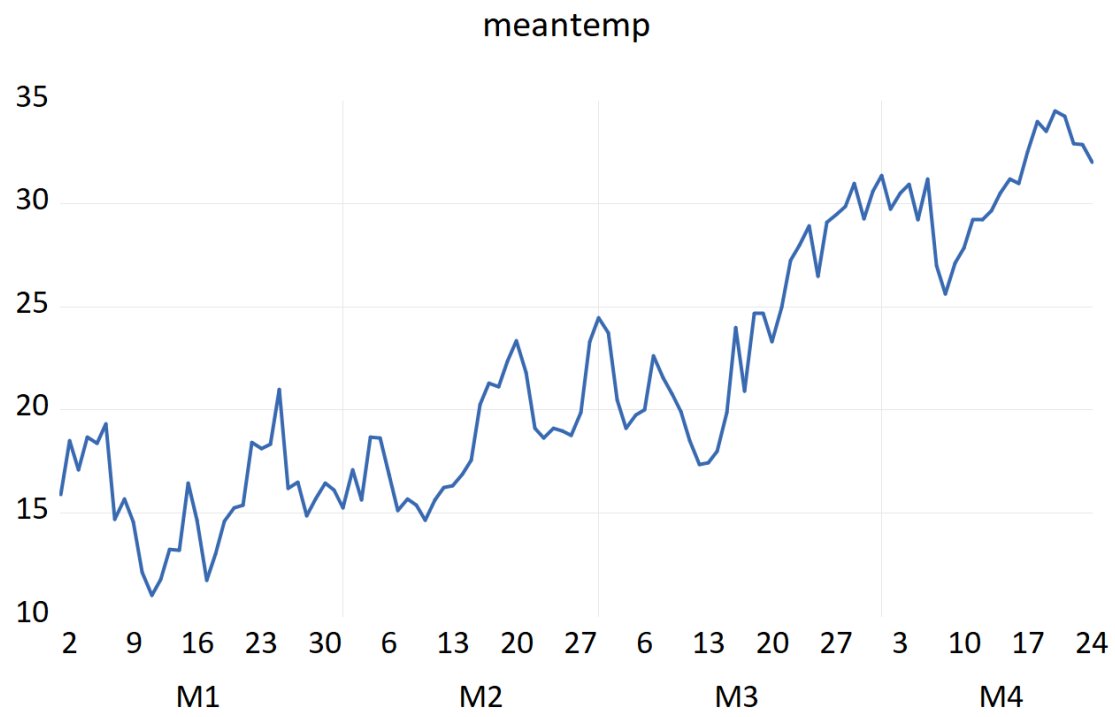


Рис 3.1 – графік процесу(вся вибірка)

На рисунку 3.2 можна побачити значення температури на навчальній вибірці спостережень.

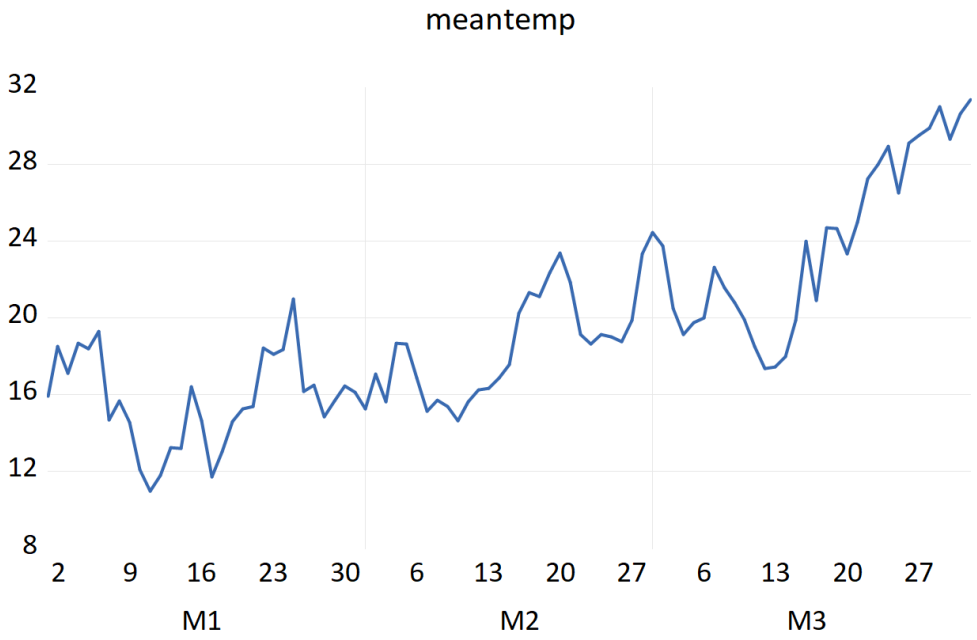


Рисунок 3.2 - графік процесу(навчальна вибірка)

Зверху(Рис 3.3) представлені результати перевірки на стаціонарність за допомогою розширеного тесту Дікі-Фулера, із яких робимо висновок щодо наявності нестаціонарності процесу. Тест проводився із припущенням щодо наявності деякого тренду процесу(виходячи із його графіку), про правильність такого припущення показали і результати тесту, а саме статистична значимість обраних критеріїв (рис 3.4). Також даний тест можна проводити і без припущення щодо наявності тренду, проте у такому випадку статистична значимість критеріїв перевірки буде невисокою, що каже про некоректність обраного варінту.

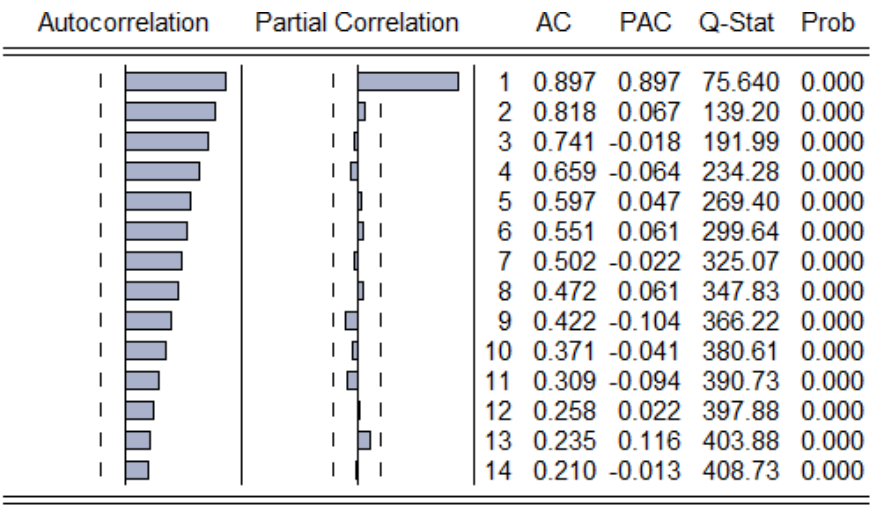
		t-Statistic	Prob.*
Augmented Dickey-Fuller test statistic		-2.798786	0.2017
Test critical values:	1% level	-4.063233	
	5% level	-3.460516	
	10% level	-3.156439	

Рисунок 3.3 – тест Дікі-Фулера

Variable	Coefficient	Std. Error	t-Statistic	Prob.
MEANTEMP(-1)	-0.177957	0.063584	-2.798786	0.0063
C	2.161264	0.889762	2.429037	0.0172
@TREND("1/01/2017")	0.031770	0.011410	2.784385	0.0066

Рисунок 3.4 - значимість критеріїв тесту Дікі-Фулера

Як бачимо із Аналізу АКФ та ЧАКФ (рис 3.5), процес можна представити у вигляді авторегресії



Autocorrelation	Partial Correlation		AC	PAC	Q-Stat	Prob
		1	0.897	0.897	75.640	0.000
		2	0.818	0.067	139.20	0.000
		3	0.741	-0.018	191.99	0.000
		4	0.659	-0.064	234.28	0.000
		5	0.597	0.047	269.40	0.000
		6	0.551	0.061	299.64	0.000
		7	0.502	-0.022	325.07	0.000
		8	0.472	0.061	347.83	0.000
		9	0.422	-0.104	366.22	0.000
		10	0.371	-0.041	380.61	0.000
		11	0.309	-0.094	390.73	0.000
		12	0.258	0.022	397.88	0.000
		13	0.235	0.116	403.88	0.000
		14	0.210	-0.013	408.73	0.000

Рисунок 3.5 – тест АКФ та ЧАКФ

Для початку побудуємо модель лінійної регресії враховуючи всі наявні параметри

Як видно (Рис 3.6), така модель має у собі статистично незначимі регресори(WIND_SPEED, MEANPRESSURE), а також невисокі значення її оцінок. Така модель не є цілком адекватною, спробуємо її покращити.

Variable	Coefficient	Std. Error	t-Statistic	Prob.
WIND_SPEED	-0.020425	0.106662	-0.191492	0.8486
C	38.27214	3.871471	9.885685	0.0000
MEANPRESSURE	-0.002812	0.003304	-0.850957	0.3971
HUMIDITY	-0.252261	0.023239	-10.85509	0.0000

R-squared	0.603642	Mean dependent var	19.43492
Adjusted R-squared	0.589974	S.D. dependent var	4.830603
S.E. of regression	3.093191	Akaike info criterion	5.139245
Sum squared resid	832.4015	Schwarz criterion	5.249612
Log likelihood	-229.8356	Hannan-Quinn criter.	5.183771
F-statistic	44.16617	Durbin-Watson stat	0.426716
Prob(F-statistic)	0.000000		

Рисунок 3.6 – оцінки моделі (1 модель)

З огляду на рисунок 3.5, спробуємо побудувати модель авторегресії 4 порядку. Як видно із рисунка 3.7, точність моделі значно повисилась.

Variable	Coefficient	Std. Error	t-Statistic	Prob.
MEANTEMP(-1)	0.825486	0.097747	8.445122	0.0000
MEANTEMP(-2)	0.070596	0.129037	0.547098	0.5858
MEANTEMP(-3)	-0.034457	0.128389	-0.268381	0.7891
MEANTEMP(-4)	-0.093376	0.104531	-0.893285	0.3744
HUMIDITY	-0.087280	0.017906	-4.874398	0.0000
WIND_SPEED	-0.099532	0.056846	-1.750894	0.0838
C	10.82184	2.199672	4.919752	0.0000

R-squared	0.905446	Mean dependent var	19.52131
Adjusted R-squared	0.898355	S.D. dependent var	4.918236
S.E. of regression	1.568025	Akaike info criterion	3.814549
Sum squared resid	196.6963	Schwarz criterion	4.012956
Log likelihood	-158.9329	Hannan-Quinn criter.	3.894441
F-statistic	127.6799	Durbin-Watson stat	1.877354
Prob(F-statistic)	0.000000		

Рисунок 3.7

Проте зараз у моделі наявні надлишкові регресори. Спробуємо поступово їх повидаляти. На рисунку 3.8 бачимо результат. Було видалено авторегресори 2 та 3 порядку і, як бачимо, оцінки моделі трохи покращились, а інформаційні критерії зменшились.

Variable	Coefficient	Std. Error	t-Statistic	Prob.
MEANTEMP(-1)	0.858852	0.062317	13.78197	0.0000
MEANTEMP(-4)	-0.091833	0.074944	-1.225357	0.2239
HUMIDITY	-0.087582	0.017669	-4.956861	0.0000
WIND_SPEED	-0.099232	0.056224	-1.764940	0.0813
C	10.86133	2.170290	5.004550	0.0000

R-squared	0.905091	Mean dependent var	19.52131
Adjusted R-squared	0.900461	S.D. dependent var	4.918236
S.E. of regression	1.551691	Akaike info criterion	3.772322
Sum squared resid	197.4351	Schwarz criterion	3.914040
Log likelihood	-159.0960	Hannan-Quinn criter.	3.829387
F-statistic	195.4965	Durbin-Watson stat	1.947626
Prob(F-statistic)	0.000000		

Рисунок 3.8

Рівняння моделі до рисунку 3.8

$$Y(k) = 0.858852*y(k-1) - 0.091833*y(k-4) - 0.087582*x_1 - 0.099232*x_2 + + 10.86133$$

Де x_1 - humidity

x_2 – wind_speed

За допомогою тесту Жака-Бера було визначено нормальність розподілення залишків

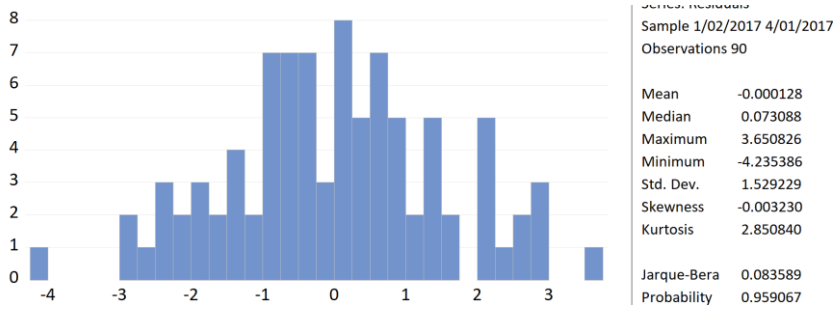


Рисунок 3.9 – статистика щодо залишків (1 модель)

Згадуючи про нестаціонарність процесу, проведемо тести на гетероскедастичність. Дивлячись лише на графік процесу (Рис 3.2), можна помітити певну неоднорідність спостережень, у точу числі і різницю у відхиленні. Результати тесту Уайта на гетероскедастичність виявились наступними:

Heteroskedasticity Test: White
Null hypothesis: Homoskedasticity

F-statistic	1.409368	Prob. F(10,79)	0.1916
Obs*R-squared	13.62532	Prob. Chi-Square(10)	0.1908
Scaled explained SS	11.51324	Prob. Chi-Square(10)	0.3190

Рисунок 3.10 – тест Уайта

Тест Уайта не виявив гетероскедастичність процесу, адже значення F – статистики більше навіть за 10%. Проте, з огляду на хоч і вище, але все ж невелике значення отриманої статистики і з огляду на початкові висновки на основі графіку, було вирішено все ж розглянути даний процесс у припущенні його гетероскедастичності, тобто розглянемо віпдовідні АРУГ-моделі та їх

можливі модернізації. На рисунку 3.10 представлені оцінки 2 моделі(а саме УАРУГ(1,1).

Variable	Coefficient	Std. Error	z-Statistic	Prob.
MEANTEMP(-1)	0.878530	0.057617	15.24769	0.0000
MEANTEMP(-4)	-0.081330	0.044763	-1.816919	0.0692
HUMIDITY	-0.093226	0.011831	-7.879893	0.0000
WIND_SPEED	-0.122695	0.001307	-93.85386	0.0000
C	10.83986	0.987164	10.98081	0.0000
Variance Equation				
C	0.298247	0.184081	1.620191	0.1052
RESID(-1)^2	-0.184227	0.056422	-3.265186	0.0011
GARCH(-1)	1.053518	0.079956	13.17617	0.0000
R-squared	0.903572	Mean dependent var		19.52131
Adjusted R-squared	0.898868	S.D. dependent var		4.918236
S.E. of regression	1.564058	Akaike info criterion		3.668742
Sum squared resid	200.5948	Schwarz criterion		3.895492
Log likelihood	-151.5903	Hannan-Quinn criter.		3.760047
Durbin-Watson stat	1.937077			

Рисунок 3.11 – оцінка 2 моделі

Як видно із рисунка 3.10, така модель також має хороші показники точності. Було змодельвано ще кілька УАРУГ-моделей, проте жодна з них не виявилась кращою за модель 2, зупинимось на ній

Формула для моделі 2 є наступною

$$Y(k) = 0.878530*y(k-1) – 0.081330*x_1(k) - 0.093226*x_1(k) – 0.122695*x_2(k) + + 10.83986 + r(k)$$

де $r(k) \sim N(0;o^2(k))$

$$o^2(k) = 0.298247 – 0.184227*r^2(k-1) + 1.053518*o^2(k-1)$$

За допомогою тесту Жака-Бера було визначено нормальність розподілення залишків (Рис 3.12)

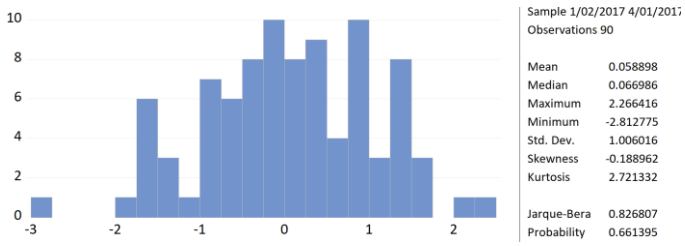


Рисунок 3.12 – статистика щодо залишків(2 модель)

Таким чином, було отримано дві моделі-кандидати. Порівняємо їх.

Як бачимо із таблиці 3.1, перша модель трохи краще пояснює зміну температури через відомі параметри, проте друга модель має трохи кращий інформаційний коефіцієнт Хеннана-Квіна.

Таблиця 3.1

	R-squared	ADJ R-squared	SUM SQ RESID	Akaike
1	0.905091	0.900461	197.4351	3.772322
2	0.903572	0.898888	200.5948	3.668742

Як бачимо, 1 модель має трохи кращу точність при меншому коефіцієнту Акаіке. Загалом, обидві моделі є хорошими, тож спробуємо змоделювати прогноз на основі як 1 так і 2.

Прогнози моделей

Як і очікувалося, 1 модель (рис 3.12) надала кращий прогноз ніж друга(рисунок 3.13), хоча і не набагато. До того ж слід виділити, що модель є точною при невеликих змінах температур(близько 2 градусів на день), а от за умови стрімкої зміни рівня температури модель не може швидко зорієнтуватися і видати точний прогноз, як бачимо із малюнка, модель запізнюється приблизно на день. Це відбувається через наявність у рівнянні

авторегресії першого порядку, проте, нажаль, цей доданок є надто важливим щоб його видаляти.

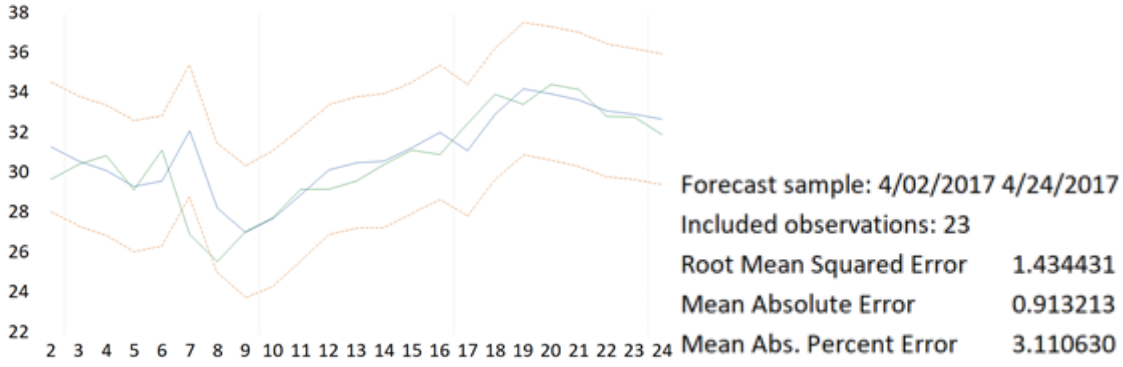


Рисунок 3.13 – прогноз 1 моделі

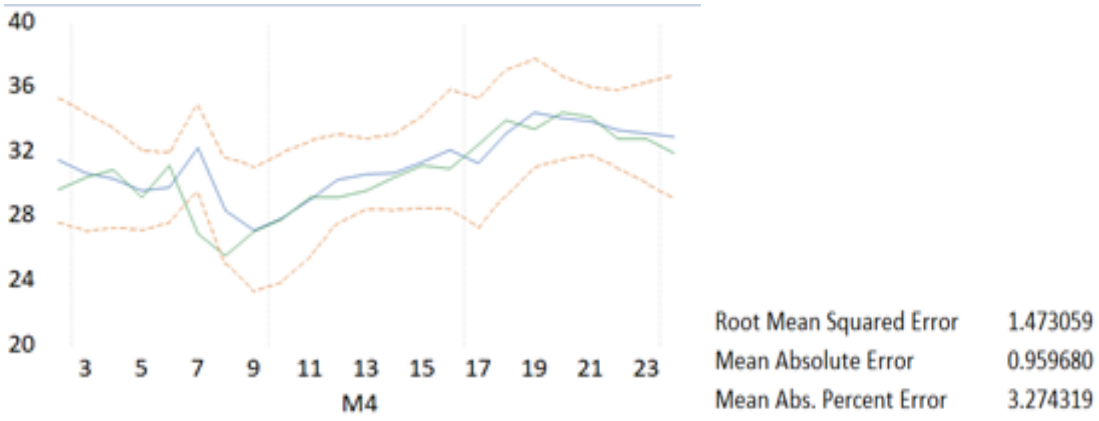


Рисунок 3.14 – прогноз 2 моделі

Загалом, модель є достатньо адекватної, особливо за умови більш-менш поступової зміни температури.

3.3.2 Друга вибірка даних

2 вибірка даних являє собою щоденну кількість відвідувань сайту протягом 1000 днів(починаючи з 14 вересня 2016 року).

Навчальна вибірка – 800 спостережень

Тестова вибірка – 200 спостережень

Із вигляду графіка(рисунок 3.13) можна зробити попередні заключення щодо даного процесу. По-перше, цей процес глобально має синусовидні коливання з періодом приблизно рік. Також від початку року і до кінця квітня кількість відвідувань сайту поступово збільшується, після чого у період травень-серпень кількість відвідувань навпаки зменшується, а у період із вересня по січень знову виростає. Також приблизно раз на 3 місяці спостерігається доволі різке та короткочасне зниження кількості відвідувань.

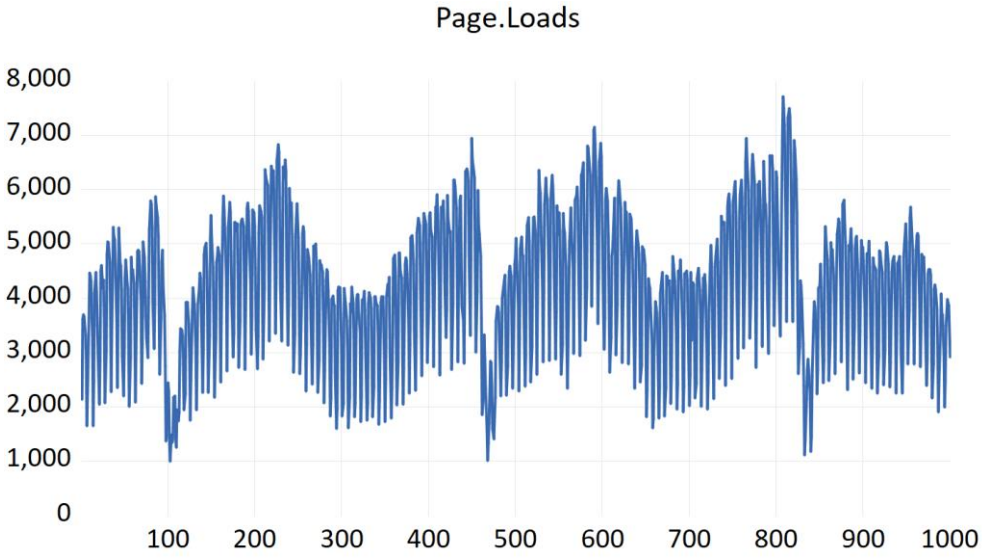


Рисунок 3.15 – графік процесу

На рисунку 3.14 можна побачити, що через кожні 7 днів значення повторюються(тобто є близькими один до одного), тож маємо семидневний період.

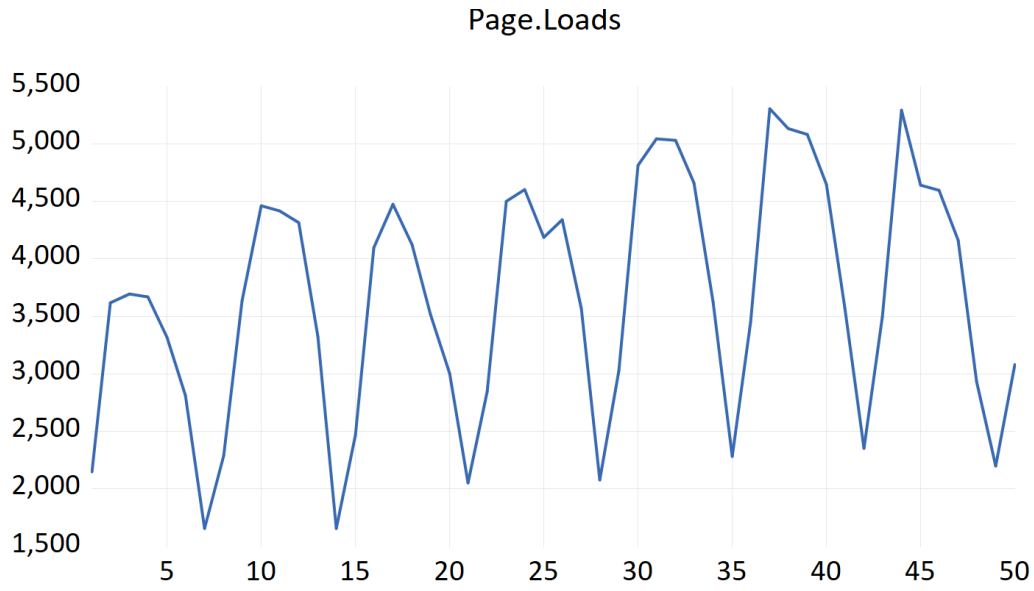


Рисунок 3.16 – перші 50 спостережень

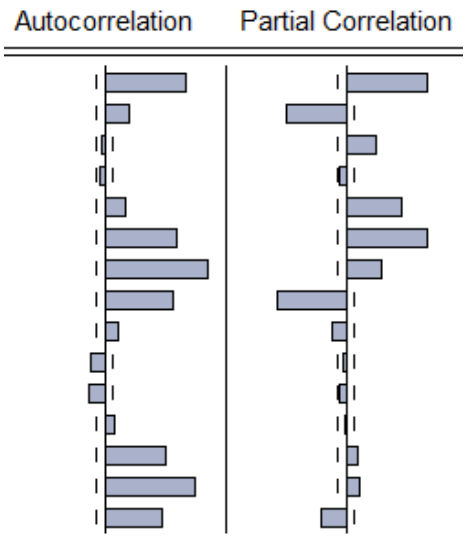


Рисунок 3.17 – АКФ та ЧАКФ

Із наведених корелограм видно семидневну періодичність, що повністю узгоджується із графіком даного процесу.

Також було проведено тест Дікі-Фулера на стаціонарність, внаслідок чого було встановлено нестаціонарність процесу.

З огляду на графіки процесу(рис 3.13, 3.14) та корелограми(рис 3.15), спробуємо побудувати авторегресію даного процесу, а саме взявши 1, 2, 6 та 7 лаги, а також день тижня.

Результати моделі є непоганими(рисунок 3.16), проте ми маємо незначний регресор(page_loads(-6)). Видалимо його.

Variable	Coefficient	Std. Error	t-Statistic	Prob.
PAGE_LOADS(-1)	0.632827	0.027353	23.13517	0.0000
PAGE_LOADS(-2)	-0.118987	0.019135	-6.218208	0.0000
PAGE_LOADS(-6)	-0.008884	0.022351	-0.397464	0.6911
PAGE_LOADS(-7)	0.435462	0.026014	16.73966	0.0000
C	1161.240	97.35979	11.92730	0.0000
DAY_OF_WEEK	-225.5442	14.66893	-15.37564	0.0000

R-squared	0.897175	Mean dependent var	4164.052
Adjusted R-squared	0.896522	S.D. dependent var	1315.060
S.E. of regression	423.0289	Akaike info criterion	14.94030
Sum squared resid	1.41E+08	Schwarz criterion	14.97567
Log likelihood	-5917.827	Hannan-Quinn criter.	14.95389
F-statistic	1373.357	Durbin-Watson stat	1.790892
Prob(F-statistic)	0.000000		

Рисунок 3.18 – результат

Результати не погіршилися(рисунок 3.17), тож цей регресор дійсно був надлишковим.

Variable	Coefficient	Std. Error	t-Statistic	Prob.
PAGE_LOADS(-1)	0.634365	0.027064	23.43962	0.0000
PAGE_LOADS(-2)	-0.121491	0.018059	-6.727434	0.0000
PAGE_LOADS(-7)	0.430225	0.022419	19.19026	0.0000
C	1136.772	75.38600	15.07935	0.0000
DAY_OF_WEEK	-222.2083	12.02421	-18.48007	0.0000

R-squared	0.897154	Mean dependent var	4164.052
Adjusted R-squared	0.896632	S.D. dependent var	1315.060
S.E. of regression	422.8028	Akaike info criterion	14.93797
Sum squared resid	1.41E+08	Schwarz criterion	14.96746
Log likelihood	-5917.907	Hannan-Quinn criter.	14.94930
F-statistic	1718.493	Durbin-Watson stat	1.791274
Prob(F-statistic)	0.000000		

Рисунок 3.19 – модель після видалення регресора(1 модель)

Як бачимо із рисунку 3.18, існує певна залежність між залишками моделі. Тож є сенс розглянути АРУГ та УАРУГ моделі.

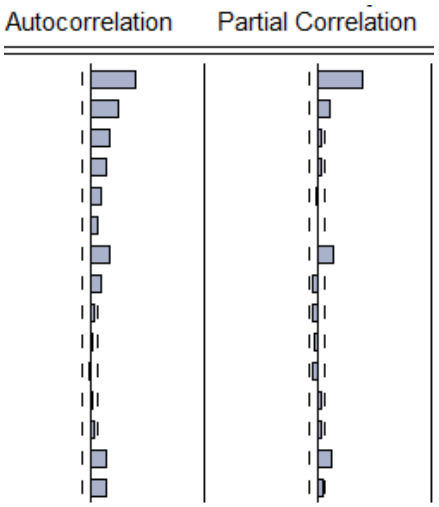


Рисунок 3.20 – АКФ та ЧАКФ залишків

З огляду на корелограми залишків, побудуємо модель УАРУГ(7, 7).

Модель УАРУГ(7, 7) виявилась не точнішою за попередню, і до того ж має багато надлишкових параметрів. Спробуємо зменшити розмірність моделі. Побудуємо модель УАРУГ(2, 2). Вона виявилася трохи гіршою за минулі.

Variable	Coefficient	Std. Error	z-Statistic	Prob.
PAGE_LOADS(-1)	0.633206	0.031931	19.83053	0.0000
PAGE_LOADS(-2)	-0.123662	0.019155	-6.455859	0.0000
PAGE_LOADS(-7)	0.433176	0.026305	16.46768	0.0000
C	1142.842	78.74843	14.51257	0.0000
DAY_OF_WEEK	-223.5243	12.69249	-17.61076	0.0000
Variance Equation				
C	96017.16	26465.63	3.627994	0.0003
RESID(-1)^2	0.138385	0.047411	2.918837	0.0035
RESID(-2)^2	0.013620	0.056899	0.239373	0.8108
RESID(-3)^2	-0.051190	0.030570	-1.674519	0.0940
RESID(-4)^2	0.017466	0.037042	0.471517	0.6373
RESID(-5)^2	0.001588	0.034962	0.045421	0.9638
RESID(-6)^2	-0.021424	0.036992	-0.579148	0.5625
RESID(-7)^2	0.231374	0.072975	3.170586	0.0015
GARCH(-1)	0.207038	0.238180	0.869253	0.3847
GARCH(-2)	-0.011483	0.199862	-0.057454	0.9542
GARCH(-3)	-0.016707	0.193439	-0.086367	0.9312
GARCH(-4)	-0.031783	0.166974	-0.190347	0.8490
GARCH(-5)	-0.005187	0.148080	-0.035031	0.9721
GARCH(-6)	0.005064	0.165540	0.030589	0.9756
GARCH(-7)	0.003426	0.097917	0.034991	0.9721

R-squared	0.897128	Mean dependent var	4164.052
Adjusted R-squared	0.896606	S.D. dependent var	1315.060
S.E. of regression	422.8561	Akaike info criterion	14.81346
Sum squared resid	1.41E+08	Schwarz criterion	14.93138
Log likelihood	-5853.535	Hannan-Quinn criter.	14.85878
Durbin-Watson stat	1.785675		

Рисунок 3.21 – Результат УАРУГ(7,7)

GARCH = C(6) + C(7)*RESID(-1)^2 + C(8)*RESID(-2)^2 + C(9)
*GARCH(-1) + C(10)*GARCH(-2)

Variable	Coefficient	Std. Error	z-Statistic	Prob.
PAGE_LOADS(-1)	0.427943	0.028662	14.93064	0.0000
PAGE_LOADS(-2)	-0.095817	0.014906	-6.428055	0.0000
PAGE_LOADS(-7)	0.620233	0.021950	28.25623	0.0000
C	804.5058	69.11757	11.63967	0.0000
DAY_OF_WEEK	-142.4968	12.11827	-11.75884	0.0000
Variance Equation				
C	11919.89	16734.03	0.712315	0.4763
RESID(-1)^2	0.377295	0.064992	5.805219	0.0000
RESID(-2)^2	-0.314404	0.099190	-3.169714	0.0015
GARCH(-1)	0.916856	0.278210	3.295549	0.0010
GARCH(-2)	-0.049625	0.122876	-0.403859	0.6863

R-squared	0.886766	Mean dependent var	4164.052
Adjusted R-squared	0.886192	S.D. dependent var	1315.060
S.E. of regression	443.6421	Akaike info criterion	14.77633
Sum squared resid	1.55E+08	Schwarz criterion	14.83530
Log likelihood	-5848.817	Hannan-Quinn criter.	14.79900
Durbin-Watson stat	1.070784		

Рисунок 3.22 – Результат УАРУГ(2,2)

На рисунку 3.20 представлены оцінки моделі Уаруг(2, 1)

C	113059.1	31167.60	3.627455	0.0003
RESID(-1)^2	0.390529	0.065191	5.990532	0.0000
RESID(-2)^2	0.145219	0.121128	1.198888	0.2306
GARCH(-1)	-0.182285	0.299659	-0.608308	0.5430

R-squared	0.885466	Mean dependent var	4164.052
Adjusted R-squared	0.884885	S.D. dependent var	1315.060
S.E. of regression	446.1814	Akaike info criterion	14.77581
Sum squared resid	1.57E+08	Schwarz criterion	14.82888
Log likelihood	-5849.610	Hannan-Quinn criter.	14.79621
Durbin-Watson stat	1.034143		

Рисунок 3.23 – Результат УАРУГ(2,1)

Було побудовано ще кілька АРУГ/УАРУГ моделей, проте всі вони виявились гіршими

Із графіку залишків(рисунок 3.21) видно, що близько як раз на рік спостерігається різке спадання точності моделі. Зважаючи на це, було

проведено детальніший аналіз та з’ясовано точні номери спостережень цих викидів, а саме 99-103, 464-468(тобто 21-25 грудня) та штучно зроблено та додано до моделі функцію, яка приймає нульові значення на всіх спостереженнях, окрім вищезазаначених(на них вона приймає значення -1, -0.8, -0.6, -0.6, -0.6. Це дозволяє зменшити відповідні модельовані дані та як наслідок і похибку. Звичайно, для прогнозування такий метод може не підійти, адже якщо такі викиди у майбутньому трапляться із здвигом хоча б на день, це призведе то неточності. Проте ми принаймні можемо очікувати різке зменшення кількості відвідувань приблизно у цей період.

Після додавання функції до моделі помилки моделі дійсно зменшились, це видно із рисунків 3.22 та 3.23

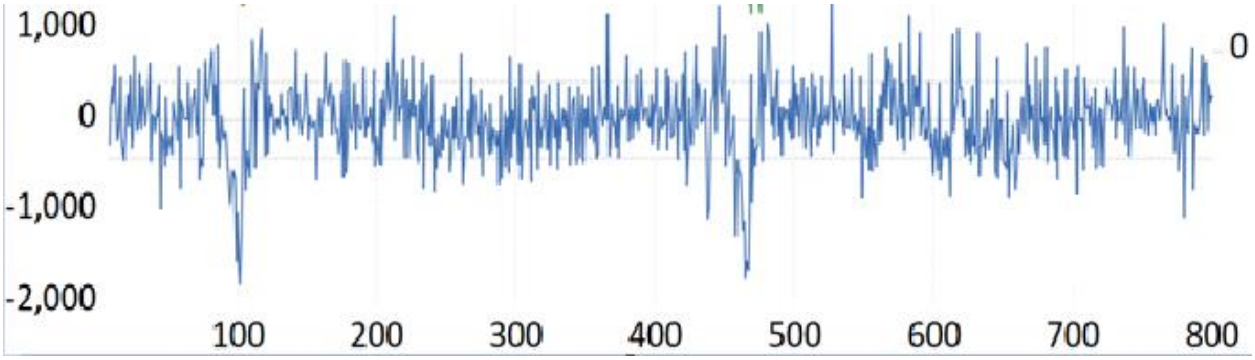


Рисунок 3.24 – графік залишків 1 моделі

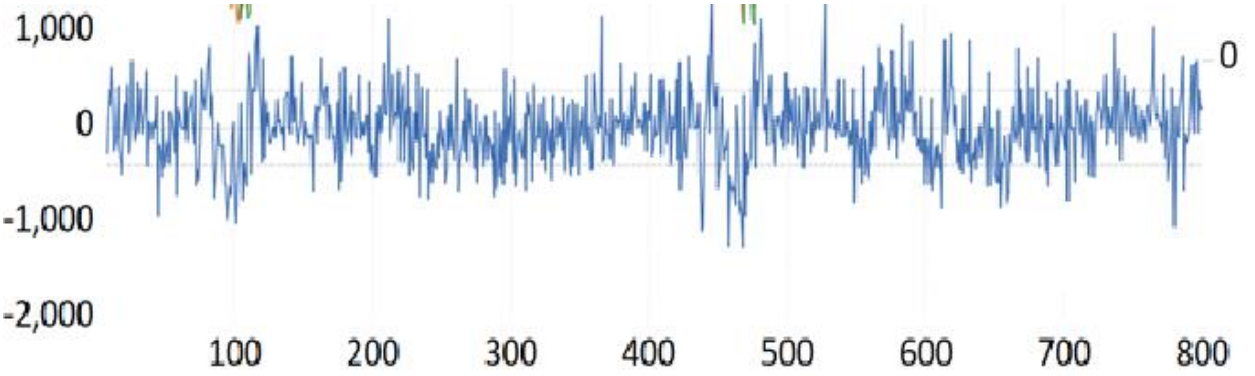


Рисунок 3.25 – Результат після додавання корегуючої функції

PAGE_LOADS(-1)	0.543631	0.026017	20.89501	0.0000
PAGE_LOADS(-2)	-0.112861	0.016628	-6.787402	0.0000
PAGE_LOADS(-7)	0.503093	0.021497	23.40259	0.0000
DAY_OF_WEEK	-195.1611	11.28804	-17.28919	0.0000
C	1087.756	69.46752	15.65849	0.0000
SERIES01	2133.440	177.6593	12.00860	0.0000

R-squared	0.913081	Mean dependent var	4164.052
Adjusted R-squared	0.912529	S.D. dependent var	1315.060
S.E. of regression	388.9358	Akaike info criterion	14.77224
Sum squared resid	1.19E+08	Schwarz criterion	14.80762
Log likelihood	-5851.194	Hannan-Quinn criter.	14.78584
F-statistic	1653.483	Durbin-Watson stat	1.802714
Prob(F-statistic)	0.000000		

Рисунок 3.26 – Оцінка моделі із функцією

До того ж оцінка моделі також дещо покращилась

Також, враховуючи періодичність та нестаціонарність досліджуваного процесу, спробуємо описати його відповідною ARIMA-моделлю. На рисунку 3.26 бачимо результати тесту Дікі-Фулера для різниць першого порядку, із чого робимо висновок щодо інтегрованості 1 порядку

Augmented Dickey-Fuller test statistic	-8.591618	0.0000

Рисунок 3.27

Отже, відповідний параметр ARIMA моделі буде рівний 1. Далі, подивившись на корелограму залишків(рис 3.27), можна зробити висновок, що різниці моделі найбільше залежать від значень різниці 7 днів тому. З огляду на попередні моделі, першу модель ARIMA було вирішено будувати лише з складовими ar(1) ar(2) та ar(7) та ar(183). Результати представлені на рисунку 3.28

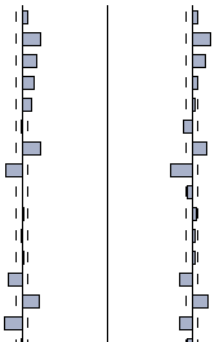


Рисунок 3.28 – корелограма залишків

C	1961.329	64.00904	30.64143	0.0000
DAY_OF_WEEK	-488.1465	15.34049	-31.82079	0.0000
SERIES01	652.2477	127.6405	5.110038	0.0000
AR(1)	-0.064822	0.016190	-4.003827	0.0001
AR(2)	-0.039114	0.010475	-3.734081	0.0002
AR(7)	0.912980	0.017506	52.15093	0.0000
AR(183)	-0.030973	0.018106	-1.710673	0.0875
MA(7)	-0.625101	0.031439	-19.88282	0.0000
SIGMASQ	104014.2	3764.788	27.62818	0.0000

R-squared	0.902122	Mean dependent var	5.244055
Adjusted R-squared	0.901131	S.D. dependent var	1031.515
S.E. of regression	324.3443	Akaike info criterion	14.42447
Sum squared resid	83107367	Schwarz criterion	14.47722
Log likelihood	-5753.575	Hannan-Quinn criter.	14.44474
F-statistic	910.1580	Durbin-Watson stat	2.205563
Prob(F-statistic)	0.000000		

Рисунок 3.29 – оцінка 1 ARIMA

Як бачимо, модель показує себе дуже добре. Після перебору кількох варіантів моделі ARIMA було виявлено найкращий, а саме ARIMA(7,7,1)(причому беруться до уваги лише залишки семидневної давності). На рисунку 3.29 наведені результати.

Variable	Coefficient	Std. Error	t-Statistic	Prob.
C	1969.840	56.98075	34.57026	0.0000
DAY_OF_WEEK	-490.3467	14.35749	-34.15268	0.0000
SERIES01	641.9275	74.63227	8.601205	0.0000
AR(1)	-0.182373	0.021778	-8.374019	0.0000
AR(2)	-0.197934	0.020845	-9.495319	0.0000
AR(3)	-0.161013	0.020815	-7.735498	0.0000
AR(4)	-0.132045	0.021801	-6.056813	0.0000
AR(5)	-0.113095	0.022015	-5.137201	0.0000
AR(6)	-0.145678	0.022433	-6.493999	0.0000
AR(7)	0.762654	0.027816	27.41778	0.0000
AR(183)	-0.065355	0.020591	-3.173916	0.0016
MA(7)	-0.580453	0.035040	-16.56552	0.0000
SIGMASQ	95866.36	3988.269	24.03708	0.0000

R-squared	0.909789	Mean dependent var	5.244055
Adjusted R-squared	0.908412	S.D. dependent var	1031.515
S.E. of regression	312.1729	Akaike info criterion	14.35647
Sum squared resid	76597219	Schwarz criterion	14.43267
Log likelihood	-5722.409	Hannan-Quinn criter.	14.38574
F-statistic	660.5764	Durbin-Watson stat	2.069522
Prob(F-statistic)	0.000000		

Рисунок 3.30 – оцінка 2 ARIMA

У таблиці 3.2 представлені оцінки конкурентних моделей. На основі таблиці можна зробити висновок, що найкращою моделлю виявилась модель із додатковою функцією для фіксування викидів.

Таблиця 3.2 – порівняння моделей

	RMSE	R-Sum	Akaike	Durbin-Watson
Функ модель	0.913	$1.19*10^8$	14.77	1.80
Модель	0.897	$1.41*10^8$	14.93	1.79
УАРУГ(7, 7)	0.897	$1.41*10^8$	14.81	1.79
УАРУГ(2, 2)	0.887	$1.55*10^8$	14.78	1.07
АРУГ(2, 1)	0.885	$1.57*10^8$	14.78	1.03
ARIMA(7,7,1)	0.91	$7.65*10^7$	14.36	2.06

Отже, найкращими моделями виявились модель із додатковою функцією для фіксування викидів та ARIMA.

$y(k) = 1087.756 + 0.543631*y(k-1) - 0.112861*y(k-2) + 0.503093*y(k-7) –$

$-195.1611*x_1(k) +2133.440*x_2(k)$

де $y(k)$ – кількість відвідувань у день k

$x_1(k)$ – день тижня у день k

$x_2(k)$ – додаткова функція

$y(k) = 1969.840 – 490*x_1(k) + 641.9275*x_2(k) – 0.182372*y(k-1) - 0.197934*y(k-2) - 0.161013*y(k-3) - 0.132045*y(k-5) - 0.113095*y(k-6) – 0.145678*y(k-7) – 0.065355*y(k -183) – 0.580453*o(t-7) + o(t)$

де $o \sim N(0;95866.36)$

Проведемо прогнозування на основі даних моделей

Прогнозування виявилося достатньо точним, на рисунках 3.30 та 3.31 відповідно можна побачити графіки та оцінки прогнозу

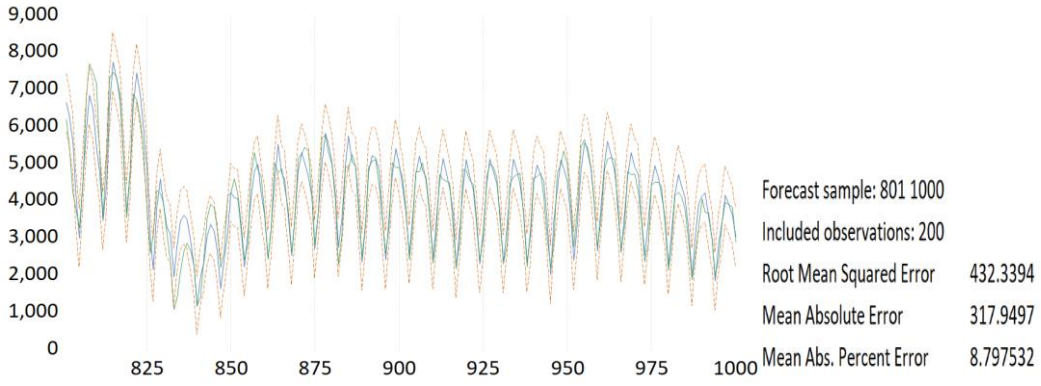


Рисунок 3.31 – Прогноз 1 моделі

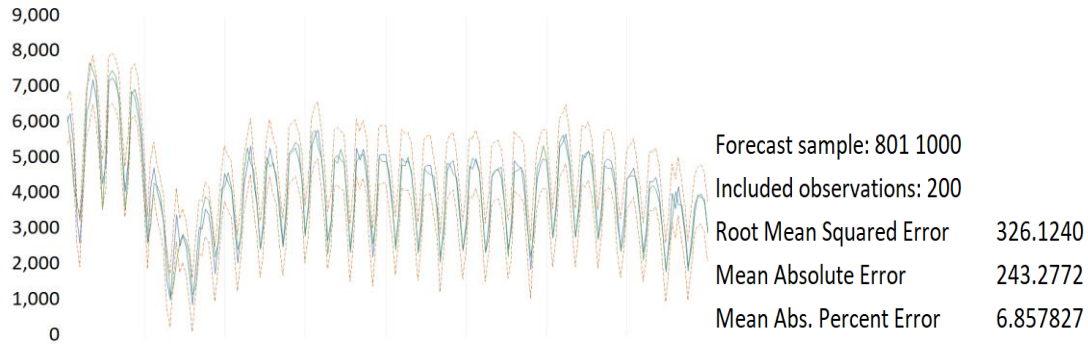


Рисунок 3.32 – Прогноз 2 моделі

Тож обидві моделі є цілком пригодними для моделювання даного процесу. У той час як перша модель має трохи кращий коефіцієнт детермінації, друга модель має помітно більшу точність. Загалом, слід визнати, що 2 модель показуж себе дещо краще за першу.

3.4 Висновки до розділу

У даному розділі бакалаврської дипломної роботи було безпосередньо розглянуто, проаналізовано, змодельовано та спрогнозовано певні реальні нелінійні нестаціонарні процеси за допомогою розглянутих у 1 та 2 розділі теоретичних та практичних підходів. Кожен із процесів був проаналізований як програмно, так і людиною, для кожного із процесів було побудовано кілька можливих моделей його опису та з них обрано найкращу. Усі моделі виявились достатньо адекватними та доволі точно прогнозували реальні дані.

Розділ 4

ФУНКЦІОНАЛЬНО-ВАРТІСНИЙ АНАЛІЗ ПРОГРАМНОГО ПРОДУКТУ

У даному розділі проводиться оцінка основних характеристик програмного продукту, розробленого для вирішення задач моделювання та прогнозування нелінійних нестаціонарних процесів.

Функціонально-вартісний аналіз (ФВА) – це технологія, яка дозволяє оцінити реальну вартість продукту або послуги незалежно від організаційної структури компанії. ФВА проводиться з метою виявлення резервів зниження витрат за рахунок ефективніших варіантів виробництва, кращого співвідношення між споживчою вартістю виробу та витратами на його виготовлення. Для проведення аналізу використовується економічна, технічна та конструкторська інформація.

Алгоритм функціонально-вартісного аналізу включає в себе визначення послідовності етапів розробки продукту, визначення повних витрат (річних) та кількості робочих часів, визначення джерел витрат та кінцевий розрахунок вартості програмного продукту.

4.1 Постановка задачі проектування

У роботі застосовується метод ФВА для проведення техніко-економічного аналізу розробки. Оскільки рішення стосовно проектування та реалізації компонентів, що розробляється, впливають на всю систему, кожна окрема підсистема має її задовольняти.

Технічні вимоги до програмного продукту є наступні:

- функціонування на персональних комп'ютерах із стандартним набором компонентів;
- зручність та зрозумілість для користувача;
- швидкість обробки даних та доступ до інформації в реальному часі;
- можливість зручного масштабування та обслуговування;
- мінімальні витрати на впровадження програмного продукту.

4.2 Обґрунтування функцій програмного продукту

Головна функція F_0 – розробка програмного продукту, який вирішує задачу прогнозування нелінійних процесів та будує його модель. Беручи за основу цю функцію, можна виділити наступні:

F_1 – вибір середовища для роботи;

F_2 – обробка вхідних даних;

F_3 – демонстрація результатів.

Кожна з цих функцій має декілька варіантів реалізації:

Функція F_1:

а) Eviews

б) SAS

Функція F_2:

а) зчитування даних з файлу

б) введення вручну

Функція F_3:

а) виведення результату на екран

б) збереження результатів у файл

Варіанти реалізації основних функцій наведені у морфологічній карті системи (рис. 4.1).

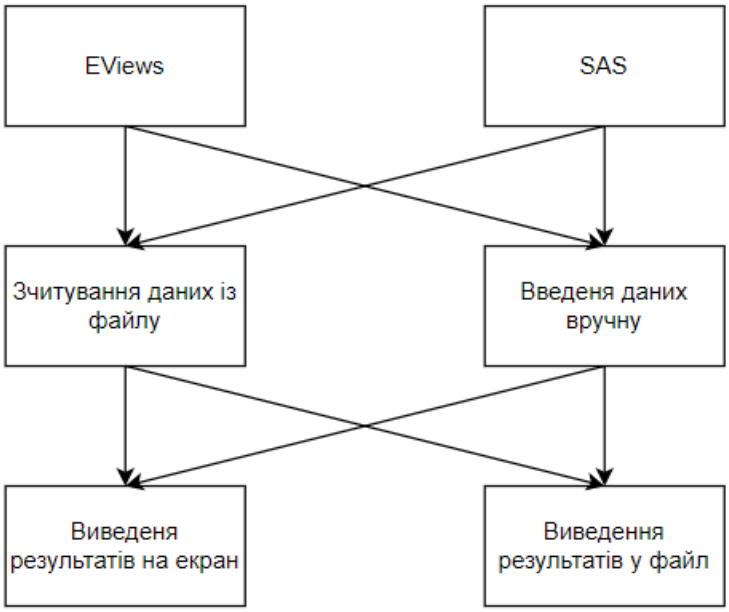


Рис 4.1 – Морфологічна карта

Морфологічна карта відображає множину всіх можливих варіанти основних функцій.

Таблиця 4.1 - Позитивно-негативна матриця

Функції	Варіанти реалізації	Переваги	Недоліки
F_1	A	Доступний для роботи, легкий інтерфейс, доступність завантаження даних популярних форматів	Неможливість роботи з багатовимірними даними.

	Б	Потужність побудови моделей для прогнозування	Велика вартість використання програми
F_2	*А*	Можна редагувати дані під час введення	Більші затрати часу, можливі помилки під час введення
	Б	Менші затрати часу	Під час вводу немає можливості відредагувати дані
F_3	*А*	Є можливість одразу побачити результати.	Відсутня історія попередніх тестів
	Б	Результати можна зберігати та використовувати для порівняння з наступними	Можливі проблеми з декодуванням символів програми

На основі цієї карти будуємо позитивно-негативну матрицю варіантів основних функцій (Таб.4.1). Робимо висновок, що при розробці програмного продукту деякі варіанти реалізації функцій варто відкинути, тому що вони не відповідають поставленим перед програмним продуктом задачам. Ці варіанти відзначені у морфологічній карті.

Функція F_1:

Дороговизна варіанту Б однозначно нас не влаштовує, тому обирається варіант А

Функція F_2:

Варіант Б не надає можливості коригування даних, тому обираємо варіант А

Функція F_3:

Розглядаємо обидва варіанти, оскільки вони забезпечують вимоги.

Таким чином, будемо розглядати такий варіанти реалізації ПП:

F_1а – F_2а – F_3а

F_1а – F_2а – F_3б

Для оцінювання якості розглянутих функцій обрана система параметрів, описана нижче.

4.3 Обґрунтування системи параметрів ПП

На основі даних, розглянутих вище, визначаються основні параметри вибору, які будуть використані для розрахунку коефіцієнта технічного рівня.

Для того, щоб охарактеризувати програмну реалізацію, будемо використовувати наступні параметри:

- *X1* – швидкодія;
- *X2* – об’єм пам’яті для обчислень та збереження даних;
- *X3* – час навчання даних;
- *X4* – точність результатів (характеризує точність розв’язків, обчислюючи вектор нев’язки).

Гірші, середні і кращі значення параметрів вибираються на основі вимог замовника й умов, що характеризують експлуатацію ПП як показано у таблиці 4.2.

Таблиця 4.2 - Основні параметри ПП

Назва Параметра	Умовні позначення	Одиниці виміру	Значення параметра		
			гірші	середні	кращі
Швидкодія	X1	оп/мс	3000	13000	21000
Об’єм пам’яті	X2	Мб	150	100	45
Час попередньої обробки даних	X3	Мс	2000	500	80
Точність результатів	X4	Число	1e-3	1e-4	1e-5

За даними таблиці 4.2 будуються графічні характеристики параметрів – рис. 4.1 – рис. 4.4.

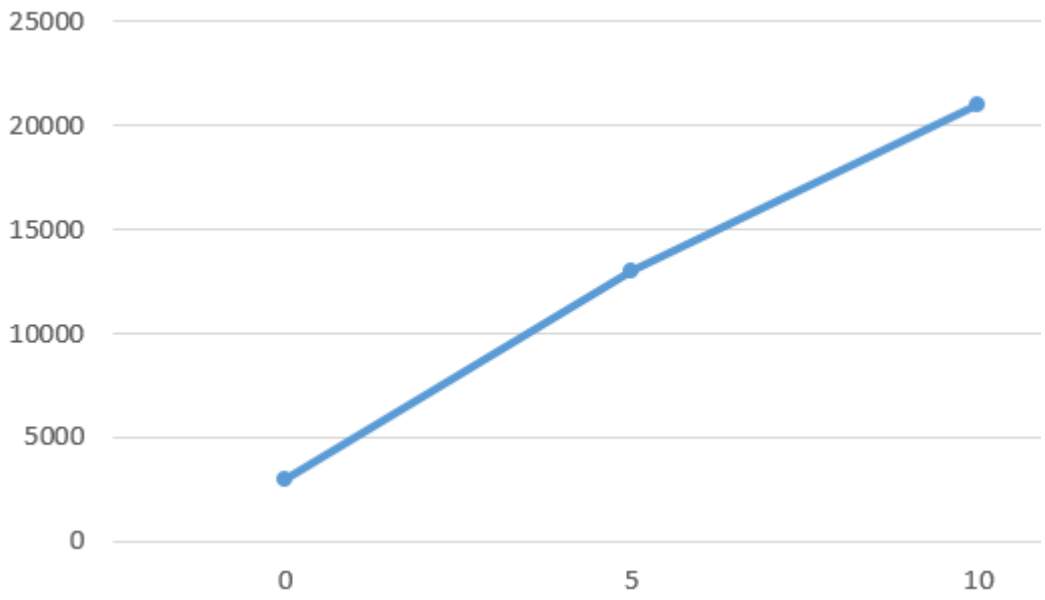


Рис 4.2 – Параметр X1

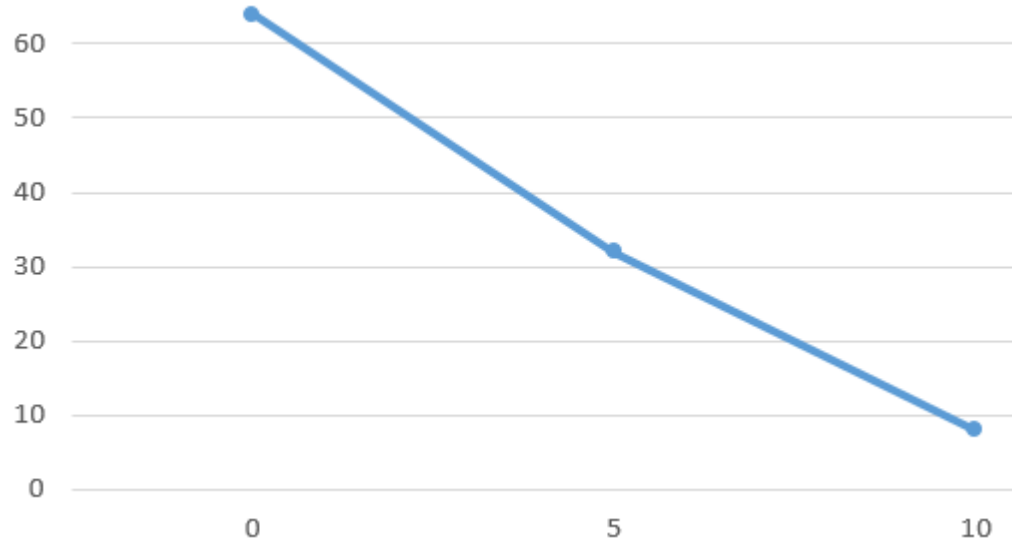


Рис 4.3 – Параметр X2

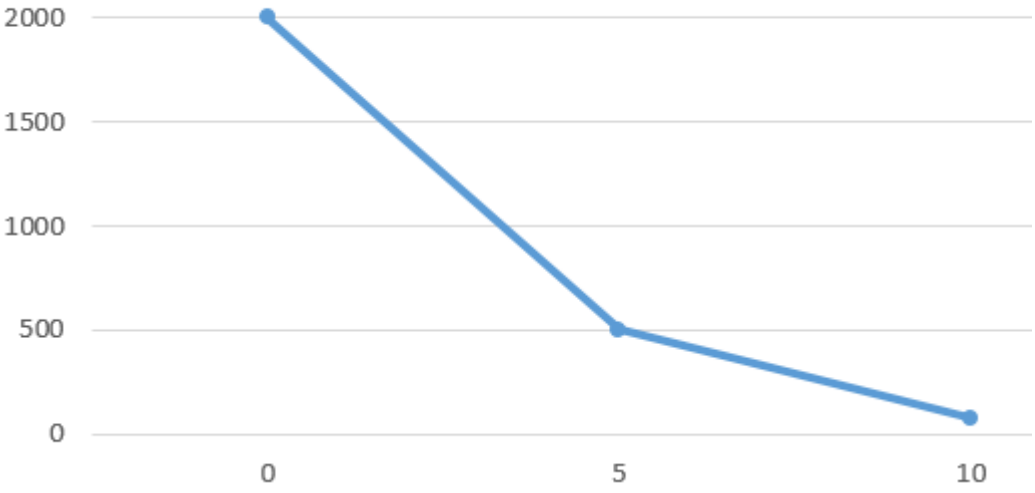


Рис 4.4 – Параметр X3

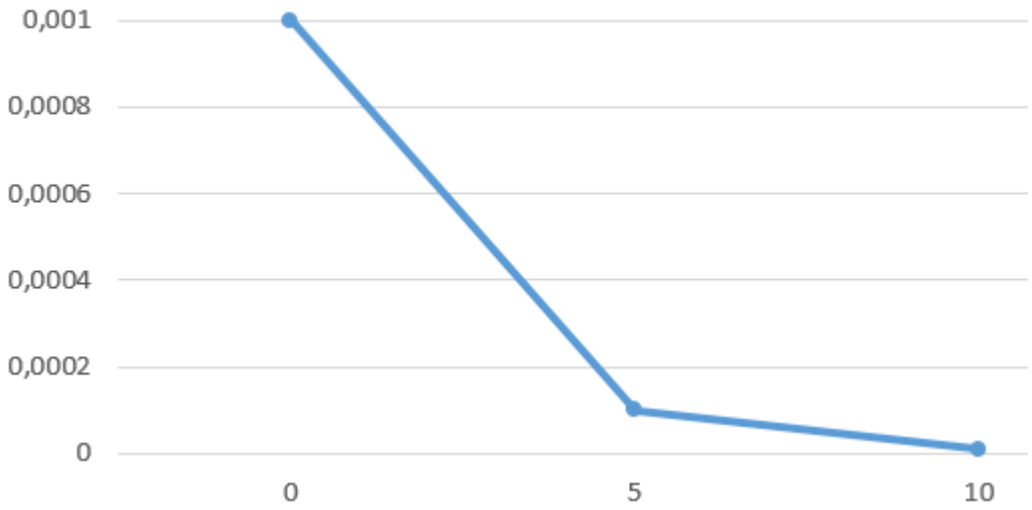


Рис 4.5 – Параметр X4

4.4 Аналіз експертного оцінювання параметрів

Після детального обговорення й аналізу кожний експерт оцінює ступінь важливості кожного параметру для конкретно поставленої цілі – розробка програмного продукту.

Значимість кожного параметра визначається методом попарного порівняння. Оцінку проводить експертна комісія із 7 людей.

Результати експертного ранжування наведені у таблиці 4.3.

Таблиця 4.3 - Результати ранжування параметрів

Позначення параметра	Назва параметра	Одиниці виміру	Ранг параметра за оцінкою експерта							Сума рангів R_i	Відхилення Δ_i	Δ_i^2
			1	2	3	4	5	6	7			
X1	Швидкодія мови	Оп/мс	3	1	1	2	3	2	1	13	-4,5	20,25

	програмування											
X2	Об’єм пам’яті	Мб	2	1	2	1	1	2	1	10	-7,5	56,25
X3	Час попередньої обробки даних	мс	2	4	2	4	3	2	3	20	2,5	6,25
X4	Точність результатів	число	3	4	5	3	3	4	5	27	9,5	90,25
	Разом		10	10	10	10	10	10	10	70	0	173

Для перевірки степені достовірності експертних оцінок, визначимо наступні параметри:

а) сума рангів кожного з параметрів і загальна сума рангів:

$$R_i = \sum_{j=1}^{N} r_{ij} R_{ij} = \frac{Nn(n+1)}{2} = 70, \qquad (4.1)$$

де N – число експертів,

n – кількість параметрів;

б) середня сума рангів:

$$T = \frac{1}{n} R_{ij} = 17{,}5 \tag{4.2}$$

в) відхилення суми рангів кожного параметра від середньої суми рангів:

$$\Delta_i = R_i - T. \tag{4.3}$$

Сума відхилень по всім параметрам повинна дорівнювати 0;

г)загальна сума квадратів відхилення:

$$S = \sum_{i=1}^{N} \Delta_i^2 = 173. \tag{4.4}$$

Порахуємо коефіцієнт узгодженості:

$$W = \frac{12S}{N^2(n^3 - n)} = \frac{12 \cdot 173}{7^2(4^3 - 4)} = 0{,}7 > W_k = 0{,}67. \tag{4.5}$$

Ранжування можна вважати достовірним, тому що знайдений коефіцієнт узгодженості перевищує нормативний, котрий дорівнює 0,67.

Скориставшись результатами ранжирування, проведемо попарне порівняння всіх параметрів і результати занесемо у таблицю 4.4.

Таблиця 4.4 - Попарне порівняння параметрів.

Параметри	Експерти							Кінцева оцінка	Числове значення
	1	2	3	4	5	6	7		
X1 і X2	<	=	<	>	>	=	=	=	1,0
X1 і X3	=	<	<	<	=	=	<	<	0,5
X1 і X4	=	<	<	<	=	<	<	<	0,5
X2 і X3	=	<	=	<	<	=	<	<	0,5
X2 і X4	<	<	<	<	<	<	<	<	0,5
X3 і X4	<	=	<	>	=	<	<	<	0,5

Таблиця 4.5 - Розрахунок вагомості параметрів

Параметриx_i	Параметриx_j				Перша ітер.		Друга ітер.		Третя ітер	
	X1	X2	X3	X4	b_i	$K_{\text{ві}}$	b_i^1	$K_{\text{ві}}^1$	b_i^2	$K_{\text{ві}}^2$
X1	1,0	1,0	0,5	0,5	3,0	0,1875	11	0,185	40,75	0,185
X2	1,0	1,0	0,5	0,5	3,0	0,1875	11	0,185	40,75	0,185
X3	1,5	1,5	1,0	0,5	4,5	0,281	16,25	0,273	59,875	0,272
X4	1,5	1,5	1,5	1,0	5,5	0,344	21,25	0,357	78,625	0,357
Всього:					16	1	59,5	1	220	1

4.5 Аналіз рівня якості варіантів реалізації функцій

Визначаємо рівень якості кожного варіанту виконання основних функцій окремо.

Таблиця 4.6 - Розрахунок показників рівня якості варіантів реалізації основних функцій ПП

Основні функції	Варіант реалізації функції	Параметри	Абсолютне значення параметра	Бальна оцінка параметра	Коефіцієнт вагомості параметра	Коефіцієнт рівня якості
F1	А	X1	21000	10	0,185	1,85
F2	А	X2	50	9	0,185	1,665
F3	А	X3	1400	2	0,273	0,546
		X4	$1*10^{-5}$	4	0,357	1,428
	Б	X3	80	10	0,273	2,73
		X4	$1*10^{-5}$	6	0,357	2,142

визначаємо рівень якості кожного з варіантів:

$$K_{K1} = 1.85 + 1.665 + 0.546 + 1.428 = 5.489$$

$$К_{К2} = 2.74 + 3.6 + 2.73 + 2.142 = 11.212$$

Із приведених розрахунків робимо висновок, що другий варіант реалізації є кращим, адже має більший коефіцієнт технічного рівня

4.6 Економічний аналіз варіантів розробки ПП

Усі варіанти включають в себе наступні завдання:

1. Розробка методів підготовки даних для їх використання в ПП:

Для першого варіанту реалізації використовуємо нормалізацію; Для другого варіанту – цифрову фільтрацію

2. Розробка проекту програмного продукту;
3. Розробка програмної оболонки;

Для першого завдання((при реалізації варіанту 1А алгоритм групи складності 3 , ступінь новизни В, вид використаної інформації — НДІ

$$T_Р = 19,\ К_П = 1.26,\ К_{СК} = 1,\ К_{СТ.М} = 1$$

$$T_1^1 = 19 * 1.26 = 20.26 \text{ людино-днів}$$

Для першого завдання((при реалізації варіанту 1Б алгоритм групи складності 3 , ступінь новизни А, вид використаної інформації — НДІ)

$$T_Р = 31,\ К_П = 1.26,\ К_{СК} = 1,\ К_{СТ.М} = 1.6.$$

$$T_1^1 = 31 * 1.26 * 1.6 = 62.5 \text{ людино-днів}$$

Для другого завдання (група складності 1 , ступінь новизни А)

$$T_Р = 98,\ К_П = 1.6,\ К_{СК} = 1,\ К_{СТ} = 0.8$$

$$T_2 = 98 * 1.6 * 0.8 = 125.4 \text{ людино-днів}$$

Для третього завдання (група складності 3 , ступінь новизни Б)

$$T_Р = 33,\ К_П = 0.7,\ 18.48 \text{ людино-днів}$$

Складаємо трудомісткість відповідних завдань для кожного з обраних варіантів реалізації програми, щоб отримати їх трудомісткість:

$$T_I = (20.26 + 125.4+18.48) \cdot 8 = 1313,12 \text{ людино-годин.}$$

$$T_{II} = (62.25+125.4 +18.48) \cdot 8 = 1649.04 \text{ людино-годин.}$$

Найбільш високу трудомісткість має варіант II.

В розробці бере участь один фінансист з окладом 40000 грн.

Зарплата за годину:

$$C=\frac{40000}{1*21*8} = 238.09 \text{ грн.}$$

Тоді, розрахуємо заробітну плату за формулою:

I. $C_{ЗП} = 238.04 * 1313,12 = 312578,1$ грн
II. $C_{ЗП=} 238.04*1649.04= 392537,5$ грн.

Відрахування на єдиний соціальний внесок становить 22%:

I. $C_{ВІД} = C_{ЗП} \cdot 0.22 = 312578,1 \cdot 0.22 = 68767.18$ грн.

II. $C_{ВІД} = C_{ЗП} \cdot 0.22 = 392537,5 \cdot 0.22 = 86358.25$ грн.

Тепер визначимо витрати на оплату однієї машино-години. ($С_М$)

Оскільки ЕОМ обслуговує один фінансист з окладом 40000 грн., з коефіцієнтом зайнятості 0.2, то для однієї машини маємо:

$$С_Г = 12 \cdot М \cdot К_З = 12 \cdot 40000 \cdot 0,2 = 96000 \text{ грн.}$$

З урахуванням додаткової заробітної плати:

$$С_{ЗП} = С_Г \cdot (1 + К_З) = 96000 \cdot (1 + 0.2) = 115200 \text{грн.}$$

Відрахування на соціальний внесок:

$$С_{ВІД} = С_{ЗП} \cdot 0.22 = 115200 \cdot 0,22 = 25344 \text{ грн.}$$

Амортизаційні відрахування розраховуємо при амортизації 25% та вартості ЕОМ – 120000 грн.

$$С_А = К_{ТМ} \cdot К_А \cdot Ц_{ПР} = 1.15 \cdot 0.25 \cdot 120000 = 34500 \text{ грн.,}$$

Витрати на ремонт та профілактику розраховуємо як:

$$С_Р = К_{ТМ} \cdot Ц_{ПР} \cdot К_Р = 1.15 \cdot 120000 \cdot 0.05 = 6900 \text{ грн.,}$$

Ефективний годинний фонд часу ПК за рік розраховуємо за формулою:

$$Т_{ЕФ} = (Д_К – Д_В – Д_С – Д_Р) \cdot t_З \cdot К_В = (365 – 110 – 9 – 16) \cdot 8 \cdot 0.9 =$$

$$= 1656 \text{годин},$$

Витрати на оплату електроенергії розраховуємо за формулою:

$$Ц_{ЕН} = 3,52$$

$$С_{ЕЛ} = Т_{ЕФ} \cdot N_С \cdot К_З \cdot Ц_{ЕН} = 1656 \cdot 0,2 \cdot 0.6 \cdot 3,52 = 699,49 \text{ грн.}$$

Накладні витрати розраховуємо за формулою:

$$С_Н = Ц_{ПР} \cdot 0.67 = 120000 \cdot 0,67 = 80400 \text{ грн.}$$

Тоді, річні експлуатаційні витрати будуть:

$$С_{ЕКС} = С_{ЗП} + С_{ВІД} + С_А + С_Р + С_{ЕЛ} + С_Н, \qquad (4.6)$$

$$С_{ЕКС} = 115200 + 25344 + 34500 + 6900 + 699 + 80400 = 263043 \text{ грн.}$$

Собівартість однієї машино-години ЕОМ дорівнюватиме:

$$С_{М\text{-}Г} = С_{ЕКС} / Т_{ЕФ} = 263043 / 1656 = 158.84 \text{ грн/год.}$$

Витрати на оплату машинного часу, в залежності від обраного варіанта реалізації, складає:

I. $С_М = 158,84 * 1313.12 = 208575.98$ грн

II. $С_М = 158,84 * 1649.04 = 261933.51$ грн

Накладні витрати складають 67% від заробітної плати:

I. $С_Н = 208575.98 * 0.67 = 139745.91$ грн.

II. $С_Н = 261933.51 * 0.67 = 175495.45$ грн.

Отже, вартість розробки ПП за варіантами становить:

I. $С_{ПП} = 312578.1 + 68767.18 + 208904.26 + 139745.91 = 729667.45$ грн.

II. $С_{ПП} = 392537.5 + 86358.25 + 262345.77 + 175495.45 = 916324.97$ грн..

4.7 Вибір кращого варіанту ПП техніко-економічного рівня

Розрахуємо коефіцієнт техніко-економічного рівня за формулою:

$К_{ТЕР1} = 5.489/ 729667.45 = 7.52 *10^{-6}$,

$К_{ТЕР2} = 11.212/ 916324.97 = 1.22 *10^{-5}$.

Як бачимо, найбільш ефективним є перший варіант реалізації програми з коефіцієнтом техніко-економічного рівня $К_{ТЕР2}$= 1.22 *10^{-5}.

4.8 Висновки до розділу

Доцільним виявилося використання другого варінту реалізації ПП, тобто використовувати програму Eviews із можливістю зчитування із файлу та збереження у файл. Було розглянуто витрати на програмний засіб моделювання та прогнозування нелінійних нестаціонарних процесів.

Розділ 5

ВИСНОВКИ ДО РОБОТИ

У бакалаврській дипломній роботі було розглянуто методики для моделювання та прогнозування нелінійних нестаціонарних процесів, та на основі цих теоретичних відомостей у комп'ютерній системі Eviews було розглянуто, проаналізовано та змодельовано кілька реальних фізичних чи економічних процесів.

Було сформовано певний алгоритм пошуку підходящої моделі для процесів. Усі побудовані моделі показали себе досить адекватно та точно як і на навчальній, так і на тестовій вибірках, що вказує на те, що даний алгоритм пошуку потрібної моделі є цілком хорошим та може застосовуватись. Було встановлено, що опис великої кількості процесів проводиться із застосуванням моделей авторегресій.

Подальша робота у цьому напрямку є актуальною, адже в оточуючий світ цілком складається із різноманітних процесів, розуміння і можливість побудови яких значно спростить життя. У подальшому можливе застосування уже існуючих теоретичних та практичних методів аналізу часових рядів для проектування та розробки власної системи та методик для моделювання та прогнозування часових рядів із застосування систем штучного інтелекту, яка(які) змогла б з часом створити конкуренцію уже існуючим і надати можливість більш точно описувати процеси, адже існуючи для цього програми за певних умов не здатні адекватно та точно проаналізувати та інтерпретувати дані що призводить до неточностей та невірних висновків.

Список використаних джерел

1. Логарифмічний тренд. Studme.

URL: https://studme.org/40992/ekonomika/logarifmicheskiy_trend

2. Лінійний тренд, параболічний тренд. Studme.

URL: https://studme.org/40990/ekonomika/modeli_trendov

3. Рысм'ятова А.: Основы эконометрики: ВМК МГУ 22.10.2014. 2-4, 6, 7-9 с.

4. Тест Уайта. Онлайн курси Coursera

URL: https://ru.coursera.org/lecture/ekonometrika/5-1-7-primier-tiesta-uaita-doska-4oyiH

5. Тест Уайта, тест Глейзера, Тест Голдфелда-Квандта, Тест Бройша-Пагана.

Studme.URL:https://studme.org/205454/ekonomika/testy_goldfelda_kvandta_gleyzera_uayta_broysha_pagana_diagnostirovaniya_geteroskedastichnosti_oshibok

6. Тест Голдфелда-Квандта. Вікіпедія, вільна енциклопедія

URL: https://ru.wikipedia.org/wiki/Тест_Голдфелда_—_Куандта

7. Канторович Г.Г.:Анализ временных рядов. Економічний журнал ВШЕ, 2002. 8, 11-13, 24-26, 31 с.

8. Демешев Б.Б. Тест Голдфелда-Квандта.

URL: https://www.youtube.com/watch?v=1ugvCUSdQxY

9. Тест Бройша-Пагана. MachineLearning, інформаційно-аналітичний ресурс

URL: http://www.machinelearning.ru/wiki/index.php?title=Критерий_Бройша-Пагана

10. Саєтова Л.Г.: Математические и инструментальные методы экономики. Технический университет имени М.Т. Калашникова, 2014. 2-4, 6-9, 13 с.

11. Тест Бройша-Пагана. Вікіпедія, вільна енциклопедія

URL: https://ru.wikipedia.org/wiki/Тест_Бройша_—_Пагана

12. Ісмагілов І.І, Кадочникова Е.І.: Многофакторная регрессия в среде Gretl. Учебное пособие Казанский федеральный университет, институт экономики и финансов, 2016. 2, 4-7, 19-21 с.

13. Метод групового урахування аргументів. URL: http://masters.donntu.org/2012/fknt/vlasov/library/art3.pdf

14. Метод групового урахування аргументів. URL: http://www.ievbras.ru/ecostat/Kiril/Library/Book1/Content393/Content393.htm

15. Авторегресійна умовна гетероскедастичність, АРУГ та УАРУГ моделі. Вікіпедія, вільна енциклопедія https://ru.wikipedia.org/wiki/Авторегрессионная_условная_гетероскедастичность#GARCH

16. Маркова А.Р.: Применение EGARCH-модели для управления структурой портфеля ценных бумаг. Международная лаборатория стохастического анализа и его приложений НИУ ВШЭ, Москва. 11, 14-16 с.

17. Федорова Е.А: Прогнозирование фондового рынка Российской федерации с помощью GARCH-моделирования. Финансовый университет при правительстве российской федерации, 2013. 2, 5, 7-9, 14 с.

18. Параболічний тренд, МНК . Studme. URL:https://studme.org/289865/matematika_himiya_fizik/parabolicheskiy_trend

19. Селютин О. Д. Аппроксимация полиномов n степени методом наименьших квадратов / А. Д. Селютин. — Текст : непосредственный // Молодой ученый. — 2018. — № 16 (202). — С. 91-96. — URL: https://moluch.ru/archive/202/49571

ДОДАТОК А. ІЛЮСТРОВАНИЙ МАТЕРІАЛ

МІНІСТЕРСТВО ОСВІТИ І НАУКИ, МОЛОДІ ТА СПОРТУ УКРАЇНИ
НАЦІОНАЛЬНИЙ ТЕХНІЧНИЙ УНІВЕРСИТЕТ УКРАЇНИ "КИЇВСЬКИЙ ПОЛІТЕХНІЧНИЙ ІНСТИТУТ"
ІНСТИТУТ ПРИКЛАДНОГО СИСТЕМНОГО АНАЛІЗУ КАФЕДРА МАТЕМАТИЧНИХ МЕТОДІВ СИСТЕМНОГО АНАЛІЗУ

Моделі і прогнозування нелінійних нестаціонарних процесів

Гаврюшин П.А.

Обєкт дослідження:

- Нелінійні нестаціонарні процеси різної природи, побудова часових рядів

Предмет дослідження:

- Математичні моделі та прогнози нелінійних нестаціонарних(гетероскедастичних) економічних процесів, методи їх побудови, часові ряди

Мета дослідження:

- Побудова адекватних моделей для аналізу та прогнозування нелінійних нестаціонарних процесів

Постановка задачі дипломної роботи

1) Аналіз особливостей функціонування процесів різної природи та збір даних для подальшого дослідження, розподілення вибірки.

2) Побудова певної кількості конкурентних моделей певних процесів на основі розглянутих теоретичних відомостей.

3) Порівняння отриманих моделей, прогнозування.

4) Аналіз отриманих результатів

Досліджувані процеси

- Нестаціонарні процеси
- Нелінійні процеси
- Гетероскедастичні процеси

Моделі, які використані для опису процесів

Модель нестаціонарного процесу:

$$y_t = \bar{a} * \bar{X}(t) + \varepsilon_t$$

Розподіл процесу змінюється часом

- АРУГ(q) : $\sigma_t^2 = a_0 + \sum_{i=1}^{q} a_i \varepsilon_{t-i}^2$
- УАРУГ(p,q): $\sigma_t^2 = a_0 + \sum_{i=1}^{q} a_i \varepsilon_{t-i}^2 + \sum_{i=1}^{p} b_i \sigma_{t-i}^2$

σ_t^2 - дисперія випадкової похибки у момент часу t

- ARIMA(p,d,q): $\Delta^d Y_t = c + \sum_{i=1}^{p} a_i \Delta^d Y_{t-i} + \sum_{j=1}^{q} b_j \Delta^d \varepsilon_{t-j} +$
- Δ^d- оператор різниці порядку d
- Також у роботі використовувались комбінації цих процесів.

Методи оцінювання параметрів

- МНК
- ЗМНК
- ММП

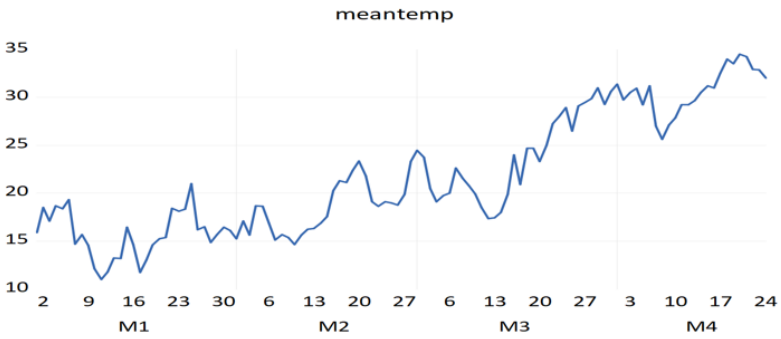
В роботі використовувався МНК та ЗМНК

Загальна методика моделювання

- Крок 1: Візуальний аналіз даних. Тести на стаціонарність, побудова та аналіз корелограм, тренд
- Крок 2: Будуємо просту модель із регресорами у вигляді інших параметрів даних
- Крок 3: Аналіз оцінки моделі, її покращення шляхом видалення надлишкових регресорів, додавання авторегресорів та випадкових збурень тощо.
- Крок 4: Оцінка отриманих моделей. Прийняття рішення щодо можливого логарифування даних, їх перегляду.
- Крок 5: Остаточно будуємо підходящі моделі, оцінюємо їх параметри.
- Крок 6: Обираємо із наявних конкурентних проблем найкращу(найкращі) на основі певних оцінок
- Крок 7: Побудова прогнозу, його оцінка, висновок

Алгоритм роботи

Досліджувані процеси

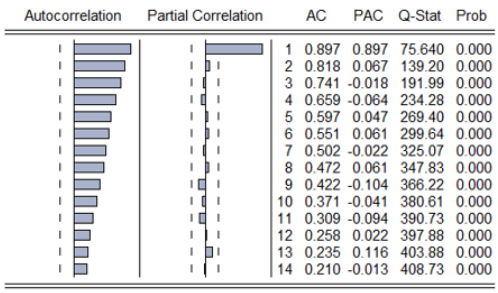


Залежність температури від тиску, вологості, вітру протягом 4 місяців

Даний процес є нелінійним та має змінну дисперсію

Попередні оцінки моделі

		t-Statistic	Prob.*
Augmented Dickey-Fuller test statistic		-2.798786	0.2017
Test critical values:	1% level	-4.063233	
	5% level	-3.460516	
	10% level	-3.156439	



Autocorrelation	Partial Correlation		AC	PAC	Q-Stat	Prob
		1	0.897	0.897	75.640	0.000
		2	0.818	0.067	139.20	0.000
		3	0.741	-0.018	191.99	0.000
		4	0.659	-0.064	234.28	0.000
		5	0.597	0.047	269.40	0.000
		6	0.551	0.061	299.64	0.000
		7	0.502	-0.022	325.07	0.000
		8	0.472	0.061	347.83	0.000
		9	0.422	-0.104	366.22	0.000
		10	0.371	-0.041	380.61	0.000
		11	0.309	-0.094	390.73	0.000
		12	0.258	0.022	397.88	0.000
		13	0.235	0.116	403.88	0.000
		14	0.210	-0.013	408.73	0.000

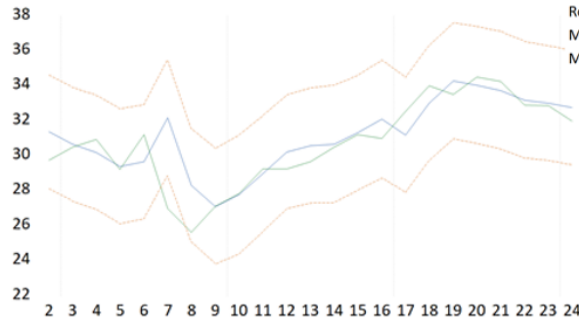
Аналіз АКФ та ЧАКФ

Variable	Coefficient	Std. Error	t-Statistic	Prob.
WIND_SPEED	-0.020425	0.106662	-0.191492	0.8486
C	38.27214	3.871471	9.885685	0.0000
MEANPRESSURE	-0.002812	0.003304	-0.850957	0.3971
HUMIDITY	-0.252261	0.023239	-10.85509	0.0000

R-squared	0.603642	Mean dependent var	19.43492
Adjusted R-squared	0.589974	S.D. dependent var	4.830603
S.E. of regression	3.093191	Akaike info criterion	5.139245
Sum squared resid	832.4015	Schwarz criterion	5.249612
Log likelihood	-229.8356	Hannan-Quinn criter.	5.183771
F-statistic	44.16617	Durbin-Watson stat	0.426716
Prob(F-statistic)	0.000000		

Регресійна модель

Результати моделювання. авторегресія 1-го порядку із залучення відомих параметрів



Forecast sample: 4/02/2017 4/24/2017
Included observations: 23

Root Mean Squared Error	1.434431
Mean Absolute Error	0.913213
Mean Abs. Percent Error	3.110630

Variable	Coefficient	Std. Error	t-Statistic	Prob.
MEANTEMP(-1)	0.858852	0.062317	13.78197	0.0000
MEANTEMP(-4)	-0.091833	0.074944	-1.225357	0.2239
HUMIDITY	-0.087582	0.017669	-4.956861	0.0000
WIND_SPEED	-0.099232	0.056224	-1.764940	0.0813
C	10.86133	2.170290	5.004550	0.0000

R-squared	0.905091	Mean dependent var	19.52131
Adjusted R-squared	0.900461	S.D. dependent var	4.918236
S.E. of regression	1.551691	Akaike info criterion	3.772322
Sum squared resid	197.4351	Schwarz criterion	3.914040
Log likelihood	-159.0960	Hannan-Quinn criter.	3.829387
F-statistic	195.4965	Durbin-Watson stat	1.947626
Prob(F-statistic)	0.000000		

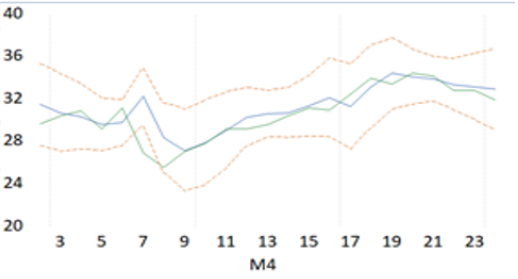
$Y(k) = 0.858852*y(k-1) – 0.091833*y(k-4) - 0.087582*x_1(k) – 0.099232*x_2(k) + +10.86133$

Результати моделювання УАРУГ(1,1)

Variable	Coefficient	Std. Error	z-Statistic	Prob.
MEANTEMP(-1)	0.878530	0.057617	15.24769	0.0000
MEANTEMP(-4)	-0.081330	0.044763	-1.816919	0.0692
HUMIDITY	-0.093226	0.011831	-7.879893	0.0000
WIND_SPEED	-0.122695	0.001307	-93.85386	0.0000
C	10.83986	0.987164	10.98081	0.0000
Variance Equation				
C	0.298247	0.184081	1.620191	0.1052
RESID(-1)^2	-0.184227	0.056422	-3.265186	0.0011
GARCH(-1)	1.053518	0.079956	13.17617	0.0000

R-squared	0.903572	Mean dependent var	19.52131
Adjusted R-squared	0.898868	S.D. dependent var	4.918236
S.E. of regression	1.564058	Akaike info criterion	3.668742
Sum squared resid	200.5948	Schwarz criterion	3.895492
Log likelihood	-151.5903	Hannan-Quinn criter.	3.760047
Durbin-Watson stat	1.937077		

Root Mean Squared Error	1.473059
Mean Absolute Error	0.959680
Mean Abs. Percent Error	3.274319



$Y(k) = 0.878530*y(k-1) – 0.081330*y(k-4) – 0.093226*x_1(k) – 0.122695*x_2(k) + 10.83986 + r(k)$

де

$r(k) \sim N(0;o^2(k))$

$o^2(k) = 0.289957 – 0.201992 * r^2(k-1) + 1.071513*o^2(k-1)$

Порівняння моделей

	R-squared	ADJ R-squared	SUM SQ RESID	Akaike
1	0.905091	0.900461	197.4351	3.772322
2	0.903572	0.898888	200.5948	3.668742

Forecast sample: 4/02/2017 4/24/2017
Included observations: 23
Root Mean Squared Error 1.434431
Mean Absolute Error 0.913213
Mean Abs. Percent Error 3.110630

- 1 модель

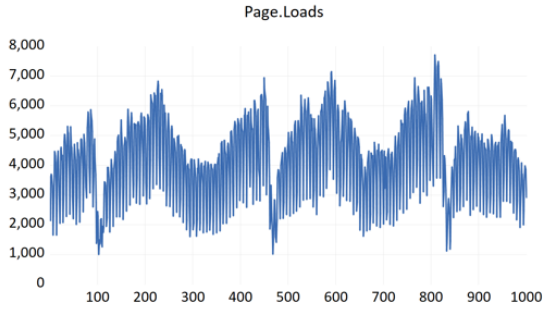
Root Mean Squared Error 1.473059
Mean Absolute Error 0.959680
Mean Abs. Percent Error 3.274319

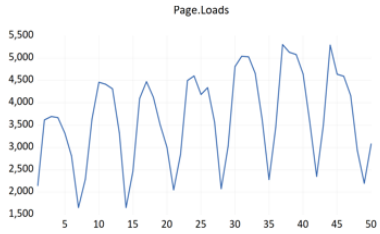
- 2 модель

- 1 модель є трохи кращою.
- Обидві моделі добре себе показують при відносно невеликій зміні температури(2-3 градуси на день)

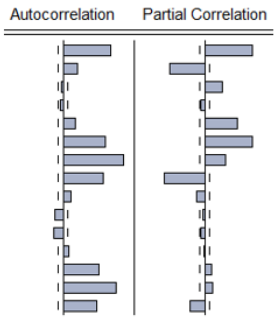
2 вибірка даних

- Вибірка показує щоденну кількість відвідувань сайту протягом 1000 днів.





Процес є періодичним, спостерігаються різкі викиди, їх неоднорідність.



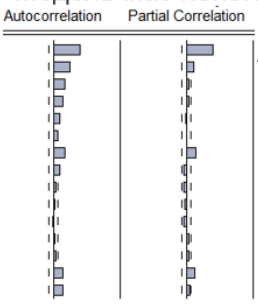
АКФ та ЧАКФ

Variable	Coefficient	Std. Error	t-Statistic	Prob.
PAGE_LOADS(-1)	0.632827	0.027353	23.13517	0.0000
PAGE_LOADS(-2)	-0.118987	0.019135	-6.218208	0.0000
PAGE_LOADS(-6)	-0.008884	0.022351	-0.397464	0.6911
PAGE_LOADS(-7)	0.435462	0.026014	16.73966	0.0000
C	1161.240	97.35979	11.92730	0.0000
DAY_OF_WEEK	-225.5442	14.66893	-15.37564	0.0000

R-squared	0.897175	Mean dependent var	4164.052
Adjusted R-squared	0.896522	S.D. dependent var	1315.060
S.E. of regression	423.0289	Akaike info criterion	14.94030
Sum squared resid	1.41E+08	Schwarz criterion	14.97567
Log likelihood	-5917.827	Hannan-Quinn criter.	14.95389
F-statistic	1373.357	Durbin-Watson stat	1.790892
Prob(F-statistic)	0.000000		

1 побудована модель

Модель має непогані оцінки, спробуємо ще їх покращити



АКФ та ЧАКФ залишків. Є автокореляція залишків, Побудуємо УАРУГ модель

Variable	Coefficient	Std. Error	z-Statistic	Prob.
PAGE_LOADS(-1)	0.633206	0.031931	19.83053	0.0000
PAGE_LOADS(-2)	-0.123662	0.019155	-6.455859	0.0000
PAGE_LOADS(-7)	0.433176	0.026305	16.46768	0.0000
C	1142.842	78.74843	14.51257	0.0000
DAY_OF_WEEK	-223.5243	12.69249	-17.61076	0.0000
Variance Equation				
C	96017.16	26465.63	3.627994	0.0003
RESID(-1)^2	0.138385	0.047411	2.918837	0.0035
RESID(-2)^2	0.013620	0.056899	0.239373	0.8108
RESID(-3)^2	-0.051190	0.030570	-1.674519	0.0940
RESID(-4)^2	0.017466	0.037042	0.471517	0.6373
RESID(-5)^2	0.001588	0.034962	0.045421	0.9638
RESID(-6)^2	-0.021424	0.036992	-0.579148	0.5625
RESID(-7)^2	0.231374	0.072975	3.170586	0.0015
GARCH(-1)	0.207038	0.238180	0.869253	0.3847
GARCH(-2)	-0.011483	0.199862	-0.057454	0.9542
GARCH(-3)	-0.016707	0.193439	-0.086367	0.9312
GARCH(-4)	-0.031783	0.166974	-0.190347	0.8490
GARCH(-5)	-0.005187	0.148080	-0.035031	0.9721
GARCH(-6)	0.005064	0.165540	0.030589	0.9756
GARCH(-7)	0.003426	0.097917	0.034991	0.9721

R-squared	0.897128	Mean dependent var	4164.052
Adjusted R-squared	0.896606	S.D. dependent var	1315.060
S.E. of regression	422.8561	Akaike info criterion	14.81346
Sum squared resid	1.41E+08	Schwarz criterion	14.93138
Log likelihood	-5853.535	Hannan-Quinn criter.	14.85878
Durbin-Watson stat	1.785675		

УАРУГ(7, 7)

GARCH = C(6) + C(7)*RESID(-1)^2 + C(8)*RESID(-2)^2 + C(9) *GARCH(-1) + C(10)*GARCH(-2)

Variable	Coefficient	Std. Error	z-Statistic	Prob.
PAGE_LOADS(-1)	0.427943	0.028662	14.93064	0.0000
PAGE_LOADS(-2)	-0.095817	0.014906	-6.428055	0.0000
PAGE_LOADS(-7)	0.620233	0.021950	28.25623	0.0000
C	804.5058	69.11757	11.63967	0.0000
DAY_OF_WEEK	-142.4968	12.11827	-11.75884	0.0000
Variance Equation				
C	11919.89	16734.03	0.712315	0.4763
RESID(-1)^2	0.377295	0.064992	5.805219	0.0000
RESID(-2)^2	-0.314404	0.099190	-3.169714	0.0015
GARCH(-1)	0.916856	0.278210	3.295549	0.0010
GARCH(-2)	-0.049625	0.122876	-0.403859	0.6863

R-squared	0.886766	Mean dependent var	4164.052
Adjusted R-squared	0.886192	S.D. dependent var	1315.060
S.E. of regression	443.6421	Akaike info criterion	14.77633
Sum squared resid	1.55E+08	Schwarz criterion	14.83530
Log likelihood	-5848.817	Hannan-Quinn criter.	14.79900
Durbin-Watson stat	1.070784		

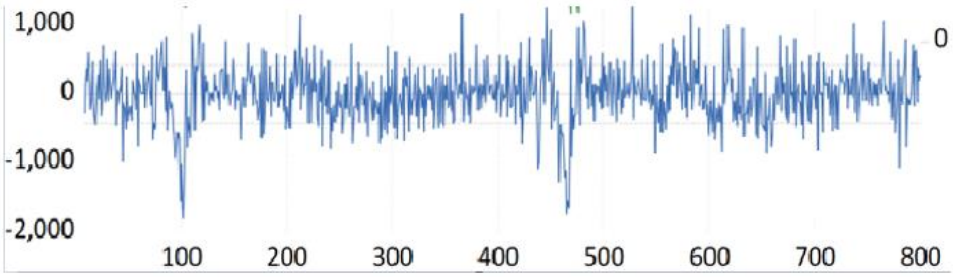
УАРУГ(2, 2)

C	113059.1	31167.60	3.627455	0.0003
RESID(-1)^2	0.390529	0.065191	5.990532	0.0000
RESID(-2)^2	0.145219	0.121128	1.198888	0.2306
GARCH(-1)	-0.182285	0.299659	-0.608308	0.5430

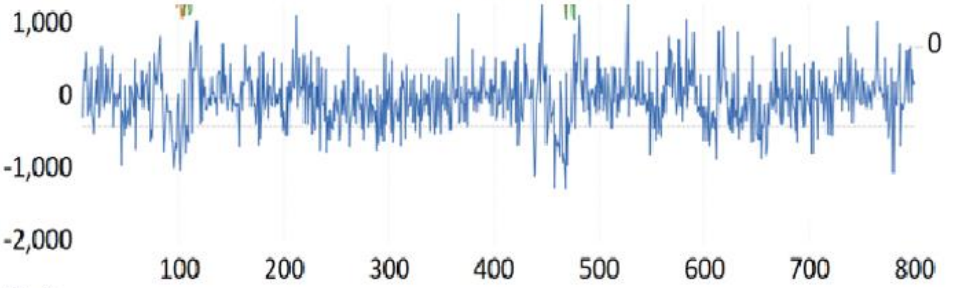
-squared	0.885466	Mean dependent var	4164.052
djusted R-squared	0.884885	S.D. dependent var	1315.060
.E. of regression	446.1814	Akaike info criterion	14.77581
um squared resid	1.57E+08	Schwarz criterion	14.82888
og likelihood	-5849.610	Hannan-Quinn criter.	14.79621
urbin-Watson stat	1.034143		

УАРУГ(2, 1)

На основі корелограм була побудована УАРУГ(7,7). Після чого була спроба її покращити, проте обидві моделі показали гірший результат.



Через викиди раз на рік спостерігається різке збільшення помилки моделі.
З огляду на це до моделі було введено корегуючу функцію.



Як бачимо, це допомогло.

PAGE_LOADS(-1)	0.543631	0.026017	20.89501	0.0000
PAGE_LOADS(-2)	-0.112861	0.016628	-6.787402	0.0000
PAGE_LOADS(-7)	0.503093	0.021497	23.40259	0.0000
DAY_OF_WEEK	-195.1611	11.28804	-17.28919	0.0000
C	1087.756	69.46752	15.65849	0.0000
SERIES01	2133.440	177.6593	12.00860	0.0000

R-squared	0.913081	Mean dependent var	4164.052
Adjusted R-squared	0.912529	S.D. dependent var	1315.060
S.E. of regression	388.9358	Akaike info criterion	14.77224
Sum squared resid	1.19E+08	Schwarz criterion	14.80762
Log likelihood	-5851.194	Hannan-Quinn criter.	14.78584
F-statistic	1653.483	Durbin-Watson stat	1.802714
Prob(F-statistic)	0.000000		

Побудова ARIMA моделі. Порівняння моделей

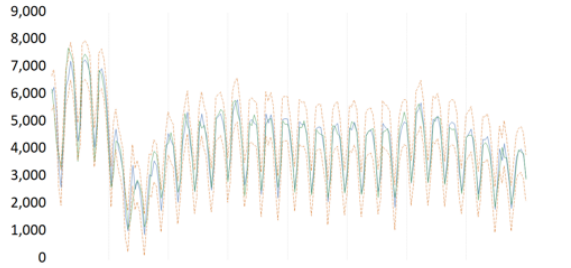
Variable	Coefficient	Std. Error	t-Statistic	Prob.
C	1969.840	56.98075	34.57026	0.0000
DAY_OF_WEEK	-490.3467	14.35749	-34.15268	0.0000
SERIES01	641.9275	74.63227	8.601205	0.0000
AR(1)	-0.182373	0.021778	-8.374019	0.0000
AR(2)	-0.197934	0.020845	-9.495319	0.0000
AR(3)	-0.161013	0.020815	-7.735498	0.0000
AR(4)	-0.132045	0.021801	-6.056813	0.0000
AR(5)	-0.113095	0.022015	-5.137201	0.0000
AR(6)	-0.145678	0.022433	-6.493999	0.0000
AR(7)	0.762654	0.027816	27.41778	0.0000
AR(183)	-0.065355	0.020591	-3.173916	0.0016
MA(7)	-0.580453	0.035040	-16.56552	0.0000
SIGMASQ	95866.36	3988.269	24.03708	0.0000

R-squared	0.909789	Mean dependent var	5.244055
Adjusted R-squared	0.908412	S.D. dependent var	1031.515
S.E. of regression	312.1729	Akaike info criterion	14.35647
Sum squared resid	76597219	Schwarz criterion	14.43267
Log likelihood	-5722.409	Hannan-Quinn criter.	14.38574
F-statistic	660.5764	Durbin-Watson stat	2.069522
Prob(F-statistic)	0.000000		

Augmented Dickey-Fuller test statistic	-8.591618	0.0000	

Тест Дікі-Фулера для різниці 1 порядку

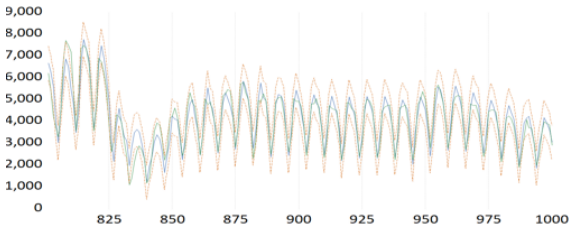
	RMSE	R-Sum	Akaike	Durbin-Watson
Функ модель	0.913	$1.19*10^8$	14.77	1.80
Модель	0.897	$1.41*10^8$	14.93	1.79
УАРУГ(7, 7)	0.897	$1.41*10^8$	14.81	1.79
УАРУГ(2, 2)	0.887	$1.55*10^8$	14.78	1.07
АРУГ(2, 1)	0.885	$1.57*10^8$	14.78	1.03
ARIMA(7,7,1)	0.91	$7.65*10^7$	14.36	2.06



Forecast sample: 801 1000
Included observations: 200

Root Mean Squared Error	326.1240
Mean Absolute Error	243.2772
Mean Abs. Percent Error	6.857827

ARIMA



1 модель

Forecast sample: 801 1000
Included observations: 200

Root Mean Squared Error	432.3394
Mean Absolute Error	317.9497
Mean Abs. Percent Error	8.797532

Модель ARIMA виявилась точнішою. Її рівняння

$y(k) = 1969.840 – 490*x_1(k) + 641.9275*x_2(k) – 0.182372*y(k-1) - 0.197934*y(k---2) –$
$-0.161013*y(k-3) - 0.132045*y(k-5) - 0.113095*y(k-6) – 0.145678*y(k-7) –$
$- 0.065355*y(k -183) – 0.580453*o(t-7) + o(t)$

де $o(t) \sim N(0;95866.36)$

Висновки

- В даній роботі досліджено методи моделювання і прогнозування нелійнійних нестаціонарних процесів, що є розповсюдженою задачею у багатьох сферах.
- Виконано моделювання та прогнозування декількох процесів. Для кожного процесу було змодельовано кілька моделей та серед них обрано найкращу, за допомогою якої проводилося прогнозування.
- Прогнози виявлялися достатньо точними

Перспективи для подальших досліджень

- Створення власної комп’ютерної системи для моделювання та прогнозування процесів різної природи із залученням систем штучного інтелекту, а також розробка нових теоретичних прийомів на основі уже відомих

Дякую за увагу