Міністерство освіти і науки України

Національний університет «Чернігівський колегіум» ім. Т.Г. Шевченка

Природничо-математичний факультет

Кафедра математики та економіки

Кваліфікаційна робота

Освітній ступень – магістр

на тему: «**Часові ряди та їх застосування**»

Виконала студентка 2-го року навчання,
Напрямку підготовки
111 Математика

Іванцова Альона Владиславівна

Керівник: Балюнов О. О.

Кваліфікаційна робота захищена
з оцінкою «_________________»

«___» ________________2021 р.

Чернігів – 2021

Зміст

Вступ

Актуальність дослідження. У промисловості й торгівлі, а також у більшості кількісних навчальних предметів багато змінних вимірюються послідовно в часі. Наприклад, в економіці відсоткові ставки та обмінні курси записуються щодня. Різні змінні можуть бути виміряні протягом різних проміжків часу, наприклад: кількість опадів, зафіксованих датчиком, може бути записана для кожного календарного місяця; температури можна вимірювати щогодини на метеостанції; або валовий внутрішній продукт, що розраховується на щорічній основі урядовим департаментом статистики. Коли змінну вимірюють послідовно в часі протягом фіксованого інтервалу (або інтервалу вибірки), дані утворюють часовий ряд.

Часовий ряд – це зібрані в різні моменти часу статистичні дані про значення будь-яких параметрів (одного або декількох) досліджуваного процесу. Кожна одиниця статистичних даних називається виміром або відліком, також можна називати його рівнем на вказаний з ним момент часу.

Часовий ряд суттєво відрізняється від звичайної вибірки даних, так як при аналізі враховується взаємозв’язок вимірювань з часом, а не тільки статистичне розмаїття та статистичні характеристики вибірки.

Прогнозування часових рядів – це техніка для передбачення подій через послідовність часу. Прогнозування передбачає майбутні події, аналізуючи тенденції минулого, виходячи з припущення, що майбутні тенденції будуть подібними до історичних тенденцій.

Однією з головних цілей аналізу даних часових рядів є сформулювання та підгонка відповідної математичної моделі для ряду. В данній роботі використано статистичний підхід, у якому часові ряди розглядаються як реалізація послідовності випадкових величин. Послідовність випадкових величин, узятих за фіксовані інтервали вибірки, іноді називають стохастичним процесом із дискретним часом.

Важливою особливістю більшості часових рядів є те, що спостереження послідовно залежать і корелюють у часі. Значна частина методології аналізу часових рядів спрямована на пояснення цієї кореляції та основних ознак даних, використовуючи відповідні статистичні моделі та описові методи. Як тільки хороша модель знайдена та адаптована до даних, аналітик може перейти до використання моделі; наприклад, для прогнозування майбутніх значень для ряду, що може допомогти з будь-якими довгостроковими плановими рішеннями, які необхідно прийняти. Підібрана модель також може бути використана як основа для статистичних тестів; наприклад, з місячними показниками продажів керівництво, ймовірно, захоче знати, чи є якісь статистичні докази зростання продажів. Нарешті, підібрана статистична модель також надає стислий виклад основних характеристик часового ряду, забезпечуючи таким чином корисний спосіб подання даних.

Аналіз часових рядів враховує той факт, що точки даних, отримані з часом, можуть мати внутрішню структуру (наприклад, автокореляцію, тенденцію або сезонні зміни), яку слід враховувати.

В даній рообті буде наведено короткий огляд деяких найбільш широко використовуваних методів у багатій і швидко зростаючій області моделювання та аналізу часових рядів.

Є багато хороших книг, які надають детальний розгляд теорії, яка лежить в основі методів, що використовуються тут.

Chatfield (1989) є популярним стислим вступним текстом, в якому математичні деталі не домінують у дискусії.

З іншого боку, для цілого ряду застосувань і подальших моделей варто прочитати Shumway і Stoffer (2000), або, для більш спеціалізованих економічних моделей часових рядів і застосувань, розглянути Enders (1995).

Наступні посилання також можуть бути корисними у розгляданні данної теми : Diggle (1990); Кендалл і Орд (1990); Броквелл і Девіс (1996); Bowerman et al. (2005).

На основі досить важливих факторів мною була вибрана така тема магістерської роботи, яка на мою думку досить популярна у наш час: «Часові ряди та їх застосування».

Об'єкт дослідження – часові ряди та моделі для їх прогнозування.

Предмет – засоби аналізу та моделі прогнозування часових рядів у різних сферах.

Мета роботи: дослідити основні концепції застосування сучасних методів аналізу часових рядів.

Відповідно до мети були поставлені наступні **завдання**:

1. Дати загально-теоретичну характеристику особливостей застосування часових рядів;
2. проаналізувати основні методи використання часових рядів;
3. провести практичне дослідження особливостей використання часових рядів.

Під час дослідження та розв'язування поставлених завдань в роботі були використані різні методи, такі як: кореляційний аналіз, спектральний аналіз, методи авторегресії та ковзних середніх, методи прогнозування, методи для перевірки даних на стаціонарність. А також до кожного з них були зроблені висновки.

Робота може бути використана для студентів та викладачів різних навчальних закладів.

Структура роботи. Робота складається зі вступу, трьох розділів, практичного дослідження, висновків, списку використаних джерел, яка містить 44 найменувань. Повний обсяг роботи становить: 80 сторінок.

Розділ 1. Загальна характеристика поняття часових рядів

1.1Загальні ідеї утворення поняття часового ряду

1. Час, можливо, являє собою найбільшу таємницю в таємничому всесвіті. Ми дивимося на якийсь час так само, як Ісак Ньютон, як на плавно поточний потік, який несе цей незвичайний світ із постійною швидкістю. Можна легко визначати моменти часу та з великою точністю вимірювати інтервали між ними. Насправді, незважаючи на те, що помилки та випадкові обурення більшості змінних, з якими ми маємо справу, будуть частими і значними, помилки у вимірі часу нам потрібно розглядати лише у виняткових випадках. Подібні труднощі, що виникають при вимірі інтервалів часу, мають штучний характер (так, наприклад, абсурдно, що християнський світ суворо зафіксував Різдво, а великодня неділя змінюється у досить широких межах)[28, с.8]
2. З давніх-давен людина вимірювала плин часу зі свічок, водяного годинника, створювали календарі, іноді зі значною точністю, і записував послідовність свого життя у вигляді літописів. Проте аналіз часових рядів на науковому рівні почав розвиватися зовсім недавно. Запис подій вздовж горизонтальної осі, на якій рівні інтервали відповідають рівним проміжкам часу, мабуть, мав місце тисячу років тому: наприклад, давні чернечі псалми, записані на одинадцяти лінійному нотному стані, є різновидом часових рядів.

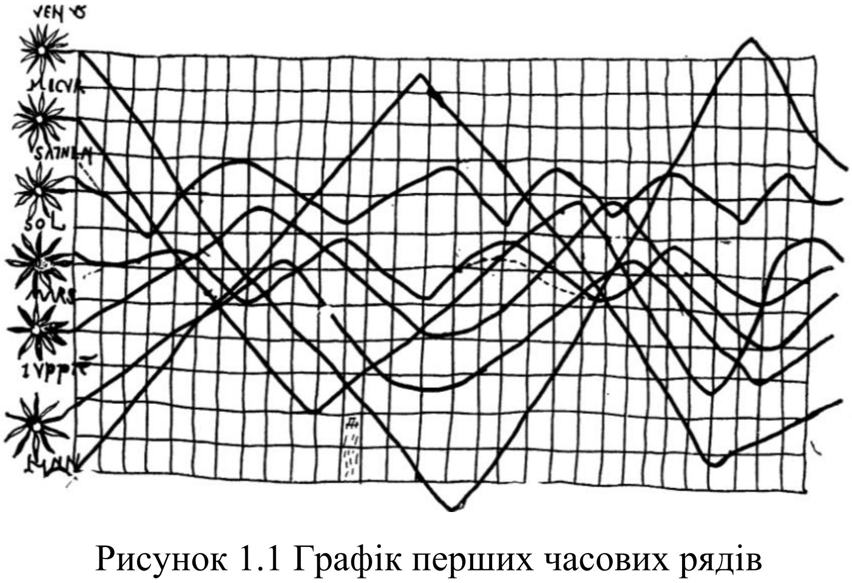


Рисунок 1.1 Графік перших часових рядів

Рисунок 1.1 може представляти інтерес як найдавніша відома у західному світі діаграма, в якій знайшли явне відображення основні ідеї тимчасової діаграми [1]. Вона датується десятим, можливо, одинадцятим століттям і є в рукописі, що містить коментар Макробіуса до «Сну Сципіона» Цицерона. Діаграма повинна була, мабуть, представляти графік нахилу планетарних орбіт як функція часу. Зона зодіаку задана на площині горизонтальною (часовою) віссю, розділеної на тридцять частин, ширина пояса зодіаку представлена ординатою. Навіть у ті далекі часи використовувалися час — абсциса і змінна — ордината, хоча примітивним обмеженим чином.

3. Не дивлячись на розробку Декартом координатної геометрії, образотворче уявлення часового ряду було пізнішим винаходом. Навіть у 1879р. Стенлі Джевонс, книга якого «Основи науки» [29] аж ніяк не призначалася для школярів, вважав за необхідне приділити увагу використанню діаграмного паперу. Очевидно, першим (чи, безумовно, однією з перших), хто представив часові діаграми сучасного типу, був Вільям Плейфер [9, с. 29], одна з діаграм якого відтворена на рисунку 1.2 (ця діаграма була опублікована в 1821 році) Плейфер, брат математика, відомого геометрам як автор аксіоми Плейфера про паралельні прямі, неодноразово пред'являв претензії на пріоритет у побудові діаграм, справедливі вони чи ні, але свідчать про те, що ця процедура була невідома.

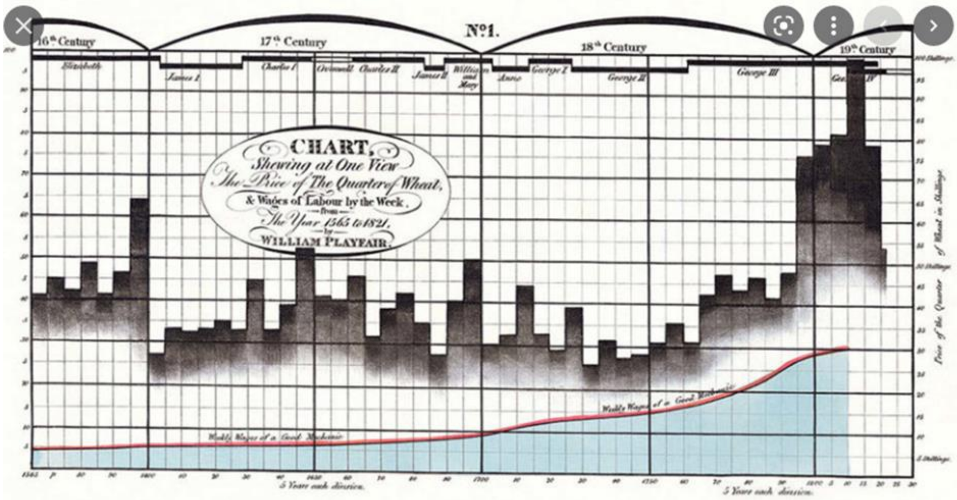


Рисунок 1.2 Діаграма з роботи Плейфера, з'єднуюча графік і гістограму

4. У дев'ятнадцятому столітті теорія статистики не була цілісною дисципліною, якою вона стала тепер. Дослідження дуже широко велися у фізичних науках незалежно від досліджень в економіці та соціології, а в той період ідеї у фізиці були повністю детерміністськими (тобто розвиток явища в часі представлявся як той, що підкоряється лише детерміністським законам). Будь-які недосконалості теорії та її невідповідності фактам або усувалися модифікацією теорії детерміністським чином (як це було, наприклад, з відкриттям планети Нептун), або належали до помилок спостереження. Коли в другій половині дев'ятнадцятого століття були зроблені спроби застосувати в економічних, соціологічних та біологічних науках методи, що настільки успішно використовувалися у фізичних науках, цей детерміністський підхід був перейнятий разом з усім розробленим математичним апаратом. Сучасна теорія статистики зародилася в момент усвідомлення того, що хоча індивідууми можуть поводитися недетерміністський, сукупність індивідуумів підпорядковується законам, які часто досить легко можуть бути виражені мовою математики.
5. У 1927 р. плідною ідеєю, що послужила відправним моментом для багатьох наукових праць в аналізі часових рядів, виконаних з того часу, Едні Юл [22] відкрив новий етап. Досліджуючи число плям на сонці, характер коливань якого явно не може вважатися повністю випадковим, Юл був вражений тим фактом, що амплітуди ряду і відстані між западинами були нерегулярними. Приклад, яким він користувався для пояснення нового підходу, став класичним: якщо розглядається вільне хитання маятника, що відхиляється на малий кут під дією сили тяжіння, то добре відомо, що його рух є гармонійним, тобто він може бути представлений синусоїдальною або косинусоїдальною хвилею з постійними амплітудами та періодами коливань. Але, якщо хлопчик буде розхитувати маятник нерегулярним чином, то він буде гойдатись, але з нерегулярними амплітудами і періодами коливань. Фактично замість такої

поведінки, при якому розбіжність між теорією і спостереженням можна віднести на рахунок незначної помилки, горох викликає ряд поштовхів, які впливають на майбутній рух системи. Ця концепція приводить нас до теорій стохастичних процесів, важливим розділом якої є теорія стохастичних часових рядів.

6. Характерна особливість часових рядів на відміну від інших статистичних об'єктів полягає в тому, що спостереження проводяться послідовно в часі, а це не таке банальне зауваження, як може здатися. Якщо розглядаються кілька рядів як багатовимірний комплекс, то крім зростання обсягу обчислень при багатовимірному аналізі, сам об'єкт набуває певного ступеня складності. У багатовимірному аналізі, як це звичай прийнято, мають справу зі зв'язками або взаємозв'язками змінних незалежно від будь-яких суб'єктивних моментів. У багатовимірних часових рядах необхідно, крім усього іншого, досліджувати кореляції і крос-кореляції рядів, коли одні з них випереджають інші або відстають від них [28, с.11]
7. Мимохідь зауважимо, що можуть зустрітися ще більш складні випадки, які потребують статистичного дослідження. Методи, які ми застосовуємо для аналізу головним чином одновимірних часових рядів, можуть бути узагальнені на випадок дослідження просторових явищ, коли є більш ніж один вимір, подібний, так би мовити, часу. Наприклад, нас може цікавити інтенсивність зараження поля жуком-лускотом (двовимірний випадок); або зміна швидкості вітру у верхніх шарах атмосфери (тривимірний випадок). Загальні теорії в таких ситуаціях стають дуже складними і виникає небезпека втрати зв'язку з реальністю.

1.2 Основні поняття та елементи часового ряду

Існують дві основні мети вивчення часових рядів :

- статистичний аналіз часових рядів широко застосовують в економічному прогнозуванні.

- властивості часових рядів визначають, яку економетричну модель буде обрано для дослідження залежності між різними змінними.

Деякі моделі часових рядів використовують також для описання автокорельованих збурень у рівняннях регресії.

Під часовим рядом (time series) розуміють як набір значень, що їх мала деяка змінна за послідовні і зазвичай за однакові проміжки часу.

Довжину такого проміжку часу можна прийняти за(рік, квартал, день, тощо), то можна вважати, що послідовні спостереження $y_1, y_2, \dots, y_n$ здійснено в моменти $t = 1,2, \dots, n$.

Особливістю статистичного аналізу, що характеризує часові ряди це, є послідовність спостережень $y_1, y_2, \dots, y_n$ розглядають як траєкторію випадкового процесу з дискретним часом, тобто як реалізацію послідовності загалом статистично залежних випадкових величин $Y_1, Y_2, \dots, Y_n$, що мають певний спільний розподіл [41, с. 182].

В загальному випадку при вивченні часових рядів економічних показників виділяють чотири основні складові:

- стійку систематично змінну довготривалу складову – тренд T;
- циклічні коливання – C;
- сезонна складова – S;
- випадкова складова – U.

Перші три складові часто об'єднують в одну детерміновану і розглядають модель ряду у вигляді: $y_t = f(t) + u_t, \forall t$. Зміна рівня $f(t)$ в залежності від часу називається трендом.

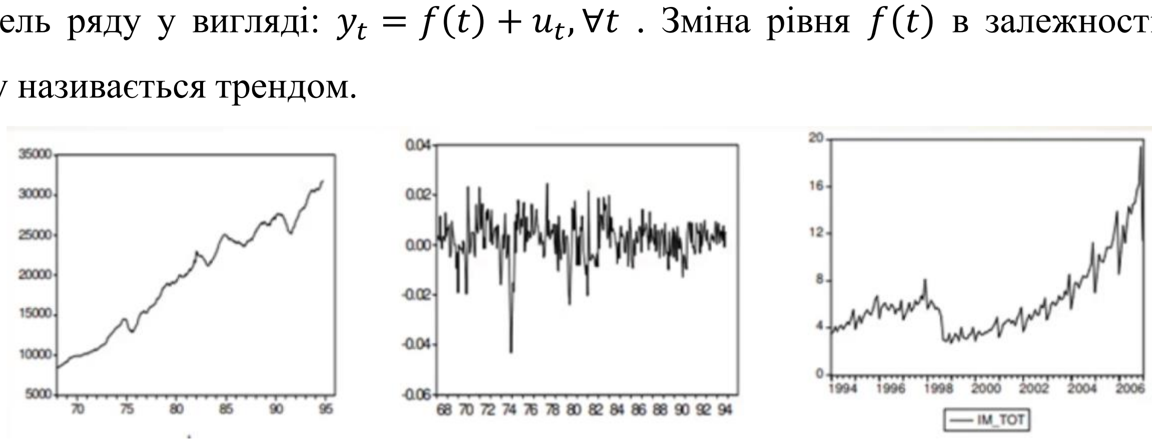


Рис. 1.1 Приклад часових рядів

На першому графіку видно явний лінійний тренд, на другому випадкові коливання, на третьому складний цикл.

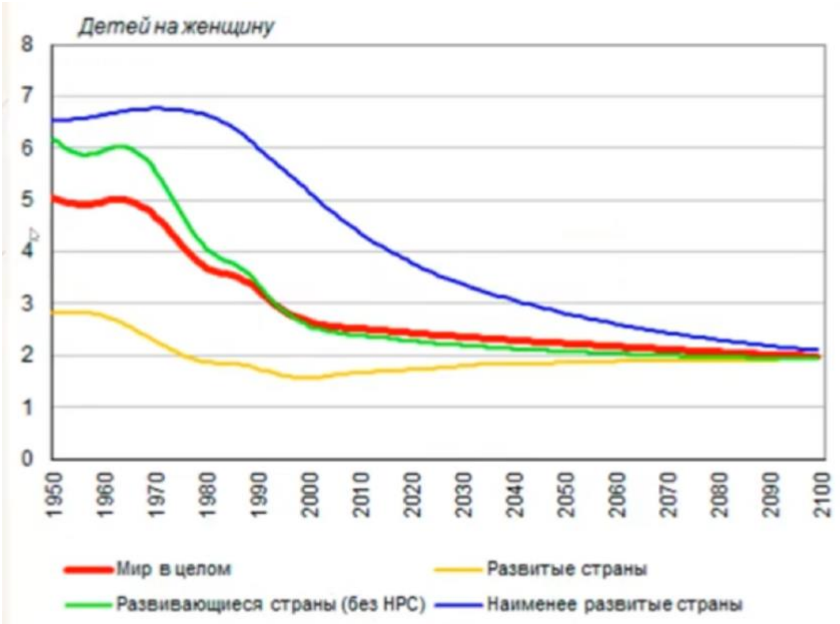


Рис. 1.2. Коефіцієнт сумарної народжуванності дітей на одну жінку, світ в цілому і основні групи країн, 1950-2100 роки.

Тренд є довготривалою тенденцією зміни, обумовленої зростанням популяції, технологічними змінами та іншими довготривалими впливами, Циклічна (кон'юнктурна) складова проявляється протягом тривалого часу і є результатом дії факторів, які мають великі післядії, або циклічно змінюються з часом. Характерним прикладом служать цикли ділової активності, демографічні та астрофізичні цикли. Циклічне змінення необов'язково є періодичним.

Випадкова складова обумовлена дією деякого періодично повторюваного і певний час року механізму.

Випадкова складова не піддається обліку та реєстрації, утворена в результаті суперпозиції великого числа зовнішніх факторів, що не беруть участь у формуванні детермінованої складової.

Предметом аналізу часового ряду є виділення і вивчення зазначених компонент ряду. Як правило - в рамках однієї з моделей ряду або адитивної $Y = T + C + S + U$, або мультиплікативної $Y = T \times C \times S \times U$.

На рисунках зображено (рис.1.3) графіки зростаючої тенденції часового ряду, сезонної компоненти та випадкової компоненти часових рядів.

Деякі складові можуть бути відсутні в тих або інших рядах. В результаті аналізу часового ряду необхідно визначити який з невипадкових складових присутні в розкладах ряду, побудувати для них хороші оцінки, підібрати модель, що описує поведінку залишків і оцінити її параметри [30, с.2].

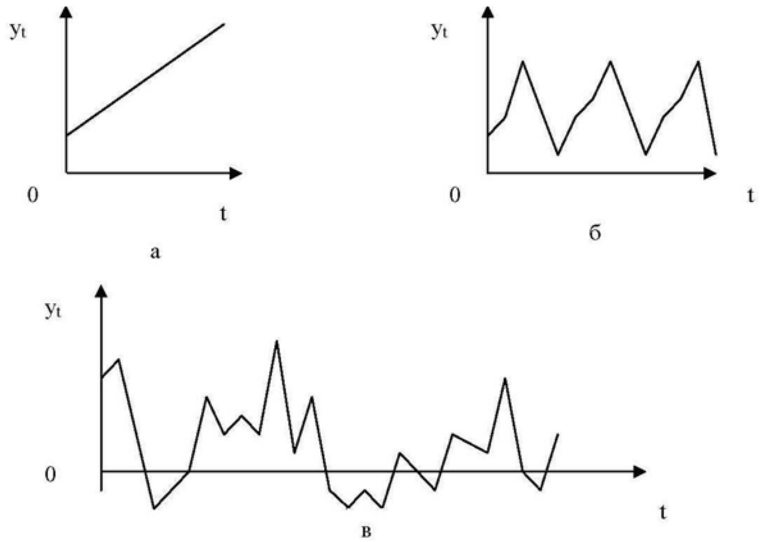


Рис. 1.3. Основні компоненти часового ряду:

а – зростаюча тенденція; б – сезонна компонента; в – випадкова компонента

При дослідженні рядів динаміки виникає низка труднощів.

- Характерною рисою часових рядів є суттєвість порядку спостережень, тоді як у випадковій вибірці порядок слідування елементів не важливий.
- Ряди економічних показників містять довготривалі тенденції - тренди.
- Вибірка містить порівняно трохи елементів, що ускладнює прогнозування досліджуваних явищ.
- Економічні ряди динаміки часто є сильно автокорелюваними.
- Можливе існування часових лагів між рядами.
- При побудові багатофакторних регресійних моделей за рядами динаміки часто виникає проблема мультиколпінеарності.

- Розвиток економічних процесів і явищ відбувається безперервно, але реально досліджувати можна лише дискретні за часом значення процесу.
- Члени часового ряду не є однаково розподіленими [46, с.12].

Метою аналізу часового ряду є досягнення розуміння причинних механізмів, що зумовили появу цього ряду. Оскільки в реальності ряд може відображати один із аспектів складного явища, що породжує кілька динамічних рядів, то на практиці обмежуються вивченням типу поведінки окремого ряду та побудовою моделей які пояснюють цей тип поведінки.

1.3 Основні методи дослідження та класифікація часових рядів

Для дослідження часових рядів найбільш часто використовують такі методи аналізу:

1. Спектральний аналіз — дозволяє знаходити періодичні та квазіперіодичні складові компоненти часового ряду.
2. Кореляційний аналіз — дозволяє виявити важливі періодичні кореляції та відповідні затримки (лаги) всередині одного ряду (автокореляція) та між кількома рядами (кроскореляція).
3. Авторегресивні моделі та моделі ковзного середнього зосереджені на описі процесу, що виявляють однорідні коливання, порушених випадковими впливами. Дозволяє передбачити майбутнє значення ряду.
4. Сезонна модель Бокса-Дженкінса — використовується, коли часовий ряд містить очевидні лінійні тенденції та сезонні компоненти. Дозволяє передбачити майбутню вартість ряду. Ця модель запропонована для аналізу повітряних перевезень.
5. Експоненціально зважене прогнозування ковзної середньої є найпростішою моделлю для прогнозування часових рядів. Застосовується до багатьох ситуацій. Включаючи моделі ціноутворення на основі випадкових блукань.

6. Методи згладжування та фільтрації, призначені для перетворення часових рядів для усунення високочастотних та сезонних коливань.
7. Метод прогнозування-дають можливість оцінити найбільш вірогідне значення в майбутньому відповідно до обраної моделі часового ряду.

Розрізняють два види часових рядів. Вимірювання деяких величин (температури, напруги і т.д.) проводиться безперервно принаймні теоретично. При цьому спостереження можна фіксувати у вигляді графіка. Але навіть у тому випадку, коли досліджувані величини реєструються (або можуть реєструватися) безперервно, практично при їх обробці використовуються тільки ті значення, які відповідають дискретній кількості моментів часу.

Отже, якщо час вимірюється безперервно, часовий ряд називається безперервним, якщо час фіксується дискретно (тобто через фіксований інтервал часу), то часовий ряд дискретний. Надалі ми будемо мати справу лише з дискретними часовими рядами. Дискретні часові ряди можна отримати двома способами:

- Вибіркою з неперервних часових рядів через регулярні проміжки часу (наприклад, чисельність населення, величина власного капіталу фірми, обсяг грошової маси, курс акції), такі часові ряди називаються моментними;
- Нагромадженням змінної протягом деякого періоду часу (приклади: обсяг виробництва якогось виду продукції, кількість опадів, обсяг імпорту), - у цьому випадку часові ряди називаються інтервальними.

В економетрії прийнято моделювати часовий ряд як випадковий процес, званий також стохастичним процесом, під яким розуміється статистичне явище, що розвивається у часі згідно із законами теорії ймовірностей. Випадковий процес – це випадкова послідовність. Зазвичай припускають, що ця послідовність йде від мінус до плюс нескінченності: $\{X_t\}_t = -\infty, \dots, +\infty$.

Для зручності можна провести класифікацію випадкових процесів та відповідних їм часових рядів на детерміновані та випадкові процеси (часові ряди). Детермінованим називають процес, який набуває заданого значення з ймовірністю одиниця.

Стохастичні процеси поділяються на стаціонарні та нестаціонарні. Стохастичний процес є стаціонарним, якщо він перебуває у сенсі у статистичній рівновазі, так як його властивості з імовірнісної точки зору не залежать від часу. Процес нестаціонарний, якщо ці умови порушуються.

Як правило, термін часовий ряд і сам має на увазі, що цей ряд є одномірним (скалярним). Часто буває важливо розглянути спільну динаміку набору часових рядів $x_t = (x_{1t}, \ldots, x_{kt})$. Такий набір називають багатовимірним часовим рядом або векторним часовим рядом. Відповідно, говорять також про багатовимірні (векторні) випадкові процеси.

Приклад: Проаналізувати виручку підприємства за 11 місяців та скласти прогноз на 12 місяців. На малюнку зображено дані за кожен місяць і його виручку у доларах.

1		
2	**Месяц**	**Выручка, $**
3	1	7 600
4	2	8 200
5	3	8 460
6	4	8 200
7	5	7 900
8	6	7 600
9	7	7 700
10	8	8 300
11	9	8 900
12	10	9 360
13	11	9 500
14	12	

Сформуємо згладжені часові ряди методом ковзного середнього за допомогою функції СРЗНАЧ. Знайдемо середні відхилення згладжених часових рядів від заданого часового ряду.

1. За значеннями вихідного часового ряду будуємо згладжений ряд часу методом ковзного середнього за даними за 2 попередні місяці. Формула

ковзної середньої в Excel. Використовуючи маркер автозаповнення, копіюємо формулу на діапазон осередків C6: C14.

C5 =СРЗНАЧ(B3;B4)

	A	B	C
1			Скс
2	**Месяц**	**Выручка, $**	*По 2 месяцам*
3	1	7 600	
4	2	8 200	
5	3	8 460	7 900
6	4	8 200	8 330
7	5	7 900	8 330
8	6	7 600	8 050
9	7	7 700	7 750
10	8	8 300	7 650
11	9	8 900	8 000
12	10	9 360	8 600
13	11	9 500	9 130
14	12		9 430

2. За таким же принципом формуємо ряд значень тримісячного та чотиримісячного ковзного середнього.

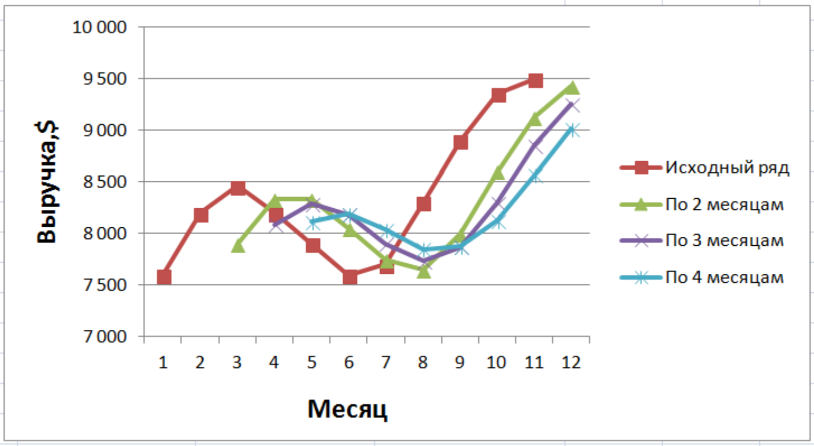
Буфер обмена Шрифт Выравнивание Число Стили Ячейки

E7 =СРЗНАЧ(B3:B6)

	A	B	C	D	E
1			**Скользящее среднее**		
2	**Месяц**	**Выручка, $**	*По 2 месяцам*	*По 3 месяцам*	*По 4 месяцам*
3	1	7 600			
4	2	8 200			
5	3	8 460	7 900		
6	4	8 200	8 330	8 087	
7	5	7 900	8 330	8 287	8 115
8	6	7 600	8 050	8 187	8 190
9	7	7 700	7 750	7 900	8 040
10	8	8 300	7 650	7 733	7 850
11	9	8 900	8 000	7 867	7 875
12	10	9 360	8 600	8 300	8 125
13	11	9 500	9 130	8 853	8 565
14	12		9 430	9 253	9 015

3. Побудуємо графік заданого часового ряду та розраховані щодо його значень прогнози за цим методом. На малюнку видно, що лінії тренду ковзного середнього зрушені щодо лінії вихідного часового ряду. Це тому, що розраховані значення згладжених часових рядів запізнюються

порівняно з відповідними значеннями заданого ряду. Адже розрахунки базувалися на даних попередніх спостережень.



4. Розрахуємо абсолютні, відносні та середні квадратичні відхилення по згладжених часових рядах.

	B	C	D	E	F	G	H	I	J	K	L	M	N
1		Скользящее среднее			Абсолютні відхилення			Відносні відхилення			Середнє квадратичне відхилення		
2	Выручка, $	*По 2 месяцам*	*По 3 месяцам*	*По 4 месяцам*	По 2 мес.	По 3 мес.	По 4 мес.	По 2 мес.	По 3 мес.	По 4 мес.	По 2 мес.	По 3 мес.	По 4 мес.
3	7 600										577,79	703,50	771,13
4	8 200												
5	8 460	7 900											
6	8 200	8 330	8 087										
7	7 900	8 330	8 287	8 115	430	387	215	5,44%	4,89%	2,72%			
8	7 600	8 050	8 187	8 190	450	587	590	5,92%	7,72%	7,76%			
9	7 700	7 750	7 900	8 040	50	200	340	0,65%	2,60%	4,42%			
10	8 300	7 650	7 733	7 850	650	567	450	7,83%	6,83%	5,42%			
11	8 900	8 000	7 867	7 875	900	1033	1025	10,11%	11,61%	11,52%			
12	9 360	8 600	8 300	8 125	760	1060	1235	8,12%	11,32%	13,19%			
13	9 500	9 130	8 853	8 565	370	647	935	3,89%	6,81%	9,84%			
14		9 430	9 253	9 015	516	640	684	6,00%	7,40%	7,84%			

Абсолютні відхилення: =ABS(B7-C7), =ABS(B7-D7), =ABS(B7-E7)

Відносні відхилення: =ABS(B7-C7)/B7

Середні квадратичні відхилення:

=КОРЕНЬ(СУММКВРАЗН(B7:B13;C7:C13)/СЧЕТ(B7:B13))

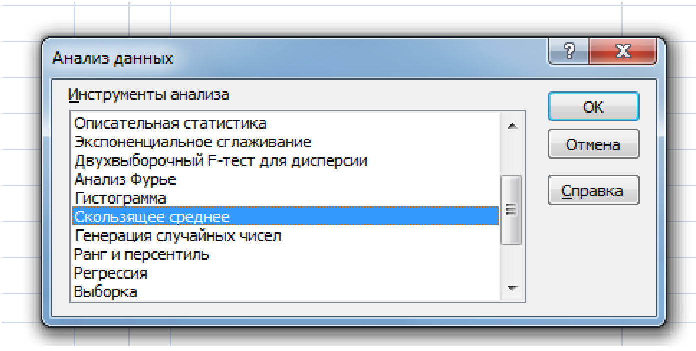
При розрахунку відхилень брали однакову кількість спостережень. Це потрібно для того, щоб провести порівняльний аналіз похибок.

Після зіставлення таблиць з відхиленнями стало видно, що для складання прогнозу за методом ковзної середньої в Excel про тенденцію зміни виручки підприємства краще модель двомісячного ковзного середнього. У неї мінімальні помилки прогнозування (порівняно з три- та чотиримісячною).

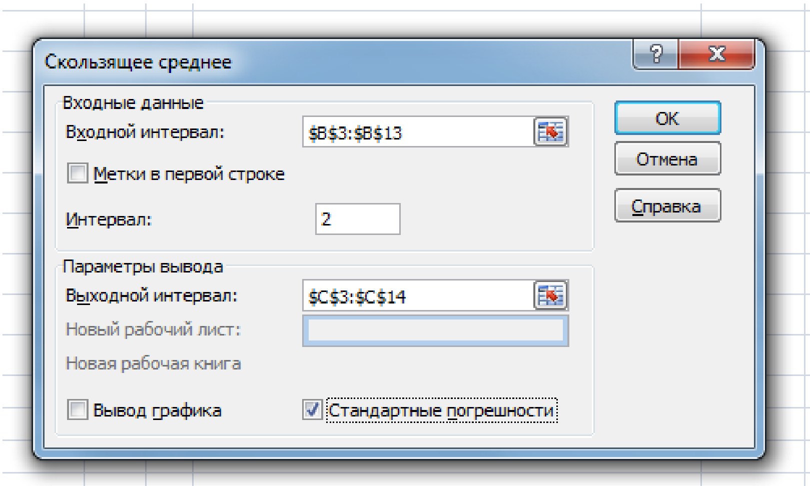
Прогнозне значення виручки на 12 місяць - 9430 у.о.

Також можна розвязати цю задачу за допомогою панелі пакет аналіза.

На вкладці Дані знаходимо команду Аналіз даних. У діалоговому вікні, що відкрилося, вибираємо Ковзне середнє:



Заповнюємо. Вхідний інтервал – вихідні значення часового ряду. Інтервал - кількість місяців, що включається в підрахунок ковзного середнього. Так як спочатку будуватимемо згладжений часовий ряд за даними двох попередніх місяців, у поле вводимо цифру 2. Вихідний інтервал - діапазон осередків для виведення отриманих результатів.



Встановивши прапорець у полі Стандартні похибки, ми автоматично додаємо до таблиці стовпець зі статистичною оцінкою похибки.

Так само знаходимо ковзне середнє за трьома місяцями. Змінюється лише інтервал (3) та вихідний діапазон.

D5 =КОРЕНЬ(СУММКВРАЗН(B4:B5;C4:C5)/2)

Месяц	Выручка, $	По 2 мес.	Стандартная погрешность	По 3 мес.	Стандартная погрешность
1	7 600	#Н/Д	#Н/Д	#Н/Д	#Н/Д
2	8 200	7900,00	#Н/Д	#Н/Д	#Н/Д
3	8 460	8330,[illegible]	231,19	8086,67	#Н/Д
4	8 200	8330,00	130,00	8286,67	#Н/Д
5	7 900	8050,00	140,36	8186,67	276,33
6	7 600	7750,00	150,00	7900,00	244,74
7	7 700	7650,00	111,80	7733,33	240,34
8	8 300	8000,00	215,06	7866,67	304,90
9	8 900	8600,00	300,00	8300,00	427,74
10	9 360	9130,00	267,30	8853,33	517,84
11	9 500	9430,00	170,00	9253,33	475,24
12					

1.4 Етапи аналізу часових рядів

На практиці вивчення часових рядів пов'язане з виявленням властивостей часового ряду та висновками про механізм ймовірності, що призводить до такого ряду. Основні цілі дослідження часових рядів:

- коротко описати характерні риси ряду;
- побудувати модель часового ряду;
- прогнозувати майбутні цінності на основі минулих спостережень;
- для управління процесом, який генерує часові ряди шляхом виявлення сигналів про несприятливу подію, що насувається.

Цілі не завжди можуть бути досягнуті через брак базової інформації (через недостатню тривалість спостережень) або через те, що статистичний склад ряду змінюється в часі. На основі цих цілей можна сформулювати процедуру аналізу часових рядів:

1) Побудова графіків та опис поведінки серії;

Якщо нестаціонарність часового ряду очевидна, першим кроком є виділення нестаціонарних компонентів ряду. Процес виявлення тенденцій та інших компонентів послідовності може здійснюватися в кілька етапів, тим самим порушуючи стаціонарність. У кожному з них розглядається ряд залишків, які є результатом віднімання підігнаної моделі тренду з вихідної послідовності або результатом різниць послідовності та інших перетворень. Крім графѕка, знаком нестаціонарності часового ряду може бути автокореляційна функція, яка не стає нульовою (за винятком дуже великих значень лагів), а на періодограмі є очевидні піки на низьких частотах. За допомогою автокореляційної функції також вивчаються внутрішні зв'язки між елементами часових рядів.

У вибіркових дослідженнях найпростіші числові ознаки описової статистики (середнє, медіана, дисперсія, стандартне відхилення, асиметрія та коефіцієнт ексцесу) зазвичай дають досить багато інформації про ряд.

Табличне представлення часових рядів і описових статистик зазвичай не дозволяє зрозуміти природу процесу, а з діаграми часових рядів можна зробити багато висновків. За графіком часових рядів можна визначити: наявність і характер тенденції; наявність сезонних і циклічних компонентів; ступінь повільних або перериvчастих змін безперервного значення послідовності після усунення тренду.

2) Визначити та виключити природні, невипадкові компоненти серії, що залежать від часу.
3) Вивчити випадковий компонент часового ряду, що залишився після виключення регулярного компонента;
4) Сформувати (вибір) математичної моделі для опису випадкових складових та перевірки її адекватності;

Модель може вважатися підібраною, якщо залишкова компонента ряду є процесом типу, як правило, білого шуму.

5) Виконати прогноз майбутніх значень ряду.

Останнім етапом аналізу часових рядів може бути прогнозування майбутнього (екстраполяція) або відновлення відсутніх (інтерполяція) значень, а також визначення точності прогнозування на основі підібраної моделі. Неможливо знайти ідеальну математичну модель для кожного часового ряду. Часто бувають ситуації, коли для опису підходять кілька моделей одночасно. Неоднозначність вибору моделі можна спостерігати при виділенні компонентної стадії детермінованої послідовності та виділенні структури залишкової послідовності. Тому для створення кількох прогнозів часто використовуються різні моделі[40, р.1,п.2].

Розділ 2. Основні характеристики динаміки розвитку соціально-економічних процесів

2.1 Характеристики динаміки часовго ряду

Для аналізу різних економічних і соціальних показників часто виникає необхідність перетворення абсолютних і середніх рівнів, моментальних або інтервальних рядів у відносні значення. Найпоширеніші характеристики динаміки розвитку соціально-економічних процесів та їхні розрахунки наведено в табл. 2.1.

Характеристики	Розрахункові формули
1. Абсолютний приріст	$\Delta y_i = y_i - y_{i-k}$
2. Коефіцієнт зростання	$K_{i(\text{зр})} = \frac{y_i}{y_{i-k}}$
3. Коефіцієнт приросту	$K_{i(\text{пр})} = \frac{y_i - y_{i-k}}{y_{i-k}}$
4. Темп зростання	$T_{i(\text{зр})} = \frac{y_i}{y_{i-k}} \times 100\% = K_{i(\text{зр})} \times 100\%$
5. Темп приросту	$T_{i(\text{пр})} = T_{i(\text{зр})} - 100\%$, або $T_{i(\text{пр})} = \frac{y_i - y_{i-k}}{y_{i-k}} \times 100\%$
6. Середня арифметична	$\bar{y} = \frac{\sum y_i}{n}$
7. Середня хронологічна	$\bar{y} = \frac{\sum y_i t}{\sum t}$
8. Середній абсолютний приріст	$\overline{\Delta y_k} = \frac{y_i - y_{i-k}}{k}$
9. Середній темп зростання	$\overline{T_{(\text{зр})}} = \sqrt[n-1]{\frac{y_n}{y_1}}$

10.Середній темп приросту	$\overline{T_{(\text{пр})}} = \overline{T_{(\text{зр})}} - 100\%$

Таблиця 2.1 Характеристики динаміки часових рядів

Для того щоб визначити зміну досліджуваного явища, необхідно спочатку розрахувати швидкість розвитку цього явища в часі. Показником швидкості є абсолютне збільшення, що характеризує величину зміни індексу в інтервалі часу між порівнювальними періодами, і розраховується за формулою:

$$\Delta y_i = y_i - y_{i-k} \, , \qquad (2.1)$$

де y_t — i-й рівень часового ряду ($i = 2,3, \dots, n$);

k — індекс початкового рівня; $k = 1,2, \dots, n-1$ і може бути обраний будь-яким залежно від мети дослідження: за $k = 1$ отримують ланцюгові показники, за $k = i - 1$ отримують базові показники із базовим початковим рівнем ряду і так далі.

Більш точно можна сказати, що швидкість розподілу показника характеризує приріст за одиницю часу, це значення називається середнім абсолютним приростом:

$$\overline{\Delta y_k} = \frac{y_i - y_{i-k}}{k}. \qquad (2.2)$$

Причому, середнє абсолютне зростання даного ряду часу за весь період спостереження дорівнює:

$$\overline{\Delta y} = \frac{y_n - y_1}{n-1} \qquad (2.3)$$

і характеризує середню швидкість зміни часового ряду, де $n -$ індекс останнього спостереження.

Для визначення відносної швидкості зміни економічного явища як одиницю часу використовують відносні показники: коефіцієнти зростання й приросту. Відмітимо, що в усіх наступних формулах індекс початкового рівня, стосовно якого здійснюють порівняння, також визначають за допомогою індексу k, як і раніше для показника абсолютного приросту.

Коефіцієнт зростання для i-го періоду обчислюють за формулою:

$$K_{i(\text{зр})} = \frac{y_i}{y_{i-k}} \, , \qquad (2.4)$$

$K_{i(\text{зр})} > 1$, якщо рівень підвищується; $K_{i(\text{зр})} < 1$, якщо рівень зменшується; за $K_{i(\text{зр})} = 1$ рівень не змінюється.

Коефіцієнт приросту дорівнює:

$$K_{i(\text{пр})} = K_{i(\text{зр})} - 1 \text{ або } K_{i(\text{пр})} = \frac{y_i - y_{i-k}}{y_{i-k}} \qquad (2.5)$$

На практиці часто застосовують показники темпу зростання й темпу приросту:

$$T_{i(\text{зр})} = \frac{y_i}{y_{i-k}} \times 100\%, \qquad (2.6)$$

де $T_{i(\text{зр})}$ — темп зростання для i-го періоду;

$$T_{i(\text{пр})} = T_{i(\text{зр})} - 100\%, \text{ або } T_{i(\text{пр})} = \frac{y_i - y_{i-k}}{y_{i-k}} \times 100\% , \qquad (2.7)$$

де $T_{i(\text{пр})}$ — темп приросту для i-го періоду. Темп зростання вказує на відсоткове збільшення рівня одного періоду відносно рівня іншого періоду, тобто відносне значення відсоткового збільшення цього показника.

Порівняння абсолютних темпів приросту з темпами зростання за цей же період показує, що в процесі реальної економіки уповільнення темпів зростання часто не супроводжується зниженням абсолютного темпу зростання.

Абсолютне значення одного відсотка приросту визначають як відношення абсолютного приросту Δy_i до темпу приросту у відсотках $T_{i(\text{пр})}$.

Середню швидкість зміни показників дослідження за певний проміжок часу характеризується також середнім темпом зростання. Він розраховується за формулою середнього геометричного:

$$\overline{T_{(\text{зр})}} = \sqrt[n-1]{T_{1(\text{зр})} T_{2(\text{зр})} \dots T_{n(\text{зр})}} = \sqrt[n-1]{\frac{y_n}{y_1}} \times 100\%, \qquad (2.8)$$

Де $T_{1(\text{зр})}, T_{2(\text{зр})}, \dots, T_{n(\text{зр})}$ — середні темпи зростання за окремі інтервали часу.

Відповідно середній темп приросту визначають як:

$$\overline{T_{(\text{пр})}} = \overline{T_{(\text{зр})}} - 100\%. \qquad (2.9)$$

Середній індекс темпів зростання, розрахований за формулою середнього геометричного (2.8), має очевидні недоліки, оскільки базується на порівнянні кінцевого та початкового рівнів часового ряду, а не враховує проміжний рівень.

Якщо рівень використання середньогеометричного темпу зростання значно коливається, статистичний аналіз призведе до серйозних помилок, які спотворять справжній тренд часового ряду.

Сучасний метод розрахунку середньої швидкості зростання певною мірою не має недоліків середнього геометричного.

Наприклад, для розрахунків середнього темпу зростання рекомендується використовувати формулу:

$$\overline{T_{(\text{зр})}} = \sqrt[n-1]{\frac{\hat{y}_n}{\hat{y}_1}}, \tag{2.10}$$

де $\hat{y}_n$, $\hat{y}_1$ — це перший і останній рівні часового ряду, згладженого відповідно до рівняння тренду (рівняння кривої зростання). Трендова модель враховує коливання проміжного рівня часового ряду, тому розраховані значення $\hat{y}_1$ та $\hat{y}_n$ і середній темп зростання (2.10) для неї точніше описуватимуть економічне явище в інтервалі дослідження.

Якщо тенденція часового ряду не змінюється, використовується середній показник рівня ряду. В інтервальному ряду динаміки з однаково розташованими в часі рівнями середній рівень ряду можна обчислити за формулою середнього арифметичного:

$$\bar{y} = \frac{\sum y_i}{n}\,. \tag{2.11}$$

Якщо інтервальна послідовність має рівень нерівності в часі, для обчислення середнього рівня послідовності (так званого середнього хронологічного) використовується формула середньозваженого арифметичного, де вага — це тривалість часу, протягом яких рівень залишається незмінним:

$$\bar{y} = \frac{\sum y_i t}{\sum t}\,, \tag{2.12}$$

де t — кількість періодів часу, для яких значення рівня y_i не змінюється.

Для моментального ряду з рівновіддаленими інтервалами часу середня часова послідовність обчислюється за такою формулою:

$$\bar{y} = \frac{{}^{1}\!/_{2}y_1 + y_2 + y_3 + \cdots + y_{n-1} + {}^{1}\!/_{2}y_n}{n-1}\,, \tag{2.13}$$

де n — кількість рівнів ряду.

Середню хронологічну для моментального часового ряду з неоднаково розташованими в часі рівнями розраховують за формулою:

$$\bar{y} = \frac{(y_1+y_2)t_1+(y_2+y_3)t_2+\cdots+(y_{n-1}+y_n)t_{n-1}}{2\sum t}. \qquad (2.14)$$

Тут n — кількість рівнів ряду, а t — період часу, що відокремлює 1-й рівень ряду від $(t\ +\ 1)$-го рівня.

2.2 Коригування рівнів часового ряду

Часовий ряд правильно показує об'єктивний закон зміни економічного показника, коли рівень ряду є порівнянним, однорідним, постійним і достатнім спостереженням. Якщо одна з цих умов не виконується, аналізувати часові ряди за допомогою математичних інструментів неправильно [26, с.43].

Проблема порівняності даних особливо гостро стоїть у часових рядах, оскільки вони можуть охоплювати значні періоди часу, протягом яких могли статися зміни, які призводять до непорівнянності статистичних рядів. Розглянемо основні причини несумісності рівнів часового ряду.

Неспівставність рівнів ряду може виникнути внаслідок зміни одиниць виміру та одиниць рахунку. Не можна порівнювати та аналізувати цифри про виробництво тканин, якщо за одні роки воно дано в метрах, а за інші – у квадратних метрах.

На сумісність рівнів часового ряду безпосередньо впливає методологія обліку чи розрахунку показників. Наприклад, якщо в роки середню врожайність вважали з засіяної площі, а в інші – з прибраної, то такі рівні будуть непорівнянні.

У процесі розвитку у часі, передусім, відбуваються кількісні зміни явищ, та відбуваються якісні стрибки, які призводять до зміни закономірностей явища. Тому науковий підхід до вивчення часових рядів полягає в тому, щоб ряди, що охоплюють великі періоди часу, розчленовувати на такі, які об'єднували лише одноякісні періоди розвитку сукупності, що характеризується однією закономірністю розвитку.

Однорідність означає відсутність нетипових, аномальних спостережень, а також викривлень тенденції. Полегшити аналіз часового ряду можна наперед звільнивши його від аномальних, нехарактерних спостережень. Аномальність спостереження є серйозною проблемою і нелегко піддається виявленню та усуненню. Аномальним рівнем називатимемо окреме значення рівня часового ряду, яке не відповідає потенційним можливостям досліджуваної економічної системи та істотно впливає на значення основних характеристик тимчасового ряду.

Зазвичай аномальні значення можна виявити візуально за допомогою графіка ряду, але перш ніж усунути виявлені таким чином їх значення потрібно піддати подальшому кількісному і якісному аналізу.

Нехарактерні точки можна поділити на три групи:

1) значення, що відображають об'єктивний розвиток процесу, але сильно відхиляються від загальної тенденції;
2) значення, що виникають унаслідок зміни методики розрахунків;
3) значення, що виникають через помилки при вимірі показника, при записі, передачі інформації і т.д.

Нехарактерні точки першого типу можуть бути корисними вивчення механізму розвитку процесу. Значення другого типу мають бути прийняті за поворотні (з них усі попередні значення слід перерахувати за новою методикою). Аномальні значення третього типу виключаються із розгляду у будь-якому випадку. Для цього розроблено низку аналітичних методів.

Засоби описової статистики та обчислення їх за даними вибіркових спостережень наведено в табл. 2.2.

Характеристики	Оцінки вибіркових значень
1. Середні значення: арифметичне геометричне	$\bar{y} = \frac{\sum_{i=1}^{n} y_i}{n}$ $y_G = \sqrt[n]{y_1 y_2 \dots y_n} = (y_1 y_2 \dots y_n)^{1/n}$

гармонійне	$\frac{1}{y_H} = \frac{1}{n}\sum_{i=1}^{n}\frac{1}{y_i}$
2. Дисперсія Середньоквадратичне відхилення	$\hat{\sigma}^2 = \frac{\sum_{i=1}^{n}(y_1-\bar{y})^2}{n-1}$ (незміщена оцінка) $\hat{\sigma}^2 = \sqrt{\hat{\sigma}}$
3. Середнє абсолютне лінійне відхилення	$MAD = \left(\sum_{i=1}^{n}\lvert y_i - \bar{y}\rvert\right)/n$
4. Початкові моменти: другого, третього, четвертого порядку	$H_2 = \frac{(\sum_{i=1}^{n} y_i{}^2)}{n}$; $H_3 = \frac{(\sum_{i=1}^{n} y_i{}^3)}{n}$; $H_4 = \frac{(\sum_{i=1}^{n} y_i{}^4)}{n}$
5. Моменти центральні: другого, третього, четвертого порядку	$m_2 = \left(\sum_{i=1}^{n}(y_i - \bar{y})^2\right)/n$ $m_3 = \left(\sum_{i=1}^{n}(y_i - \bar{y})^3\right)/n$ $m_4 = \left(\sum_{i=1}^{n}(y_i - \bar{y})^4\right)/n$
6. Коефіцієнт асиметрії відхилення	$A = {m_3}/{m_2}^{\frac{3}{2}}$
7. Показник ексцесу	$E = \frac{m_4}{(m_2)^2} - 3$
8. Коефіцієнт варіації: за розмахом за середнім абсолютним лінійним відхиленням за середнім абсолютним лінійним відхиленням медіана мода мінімальне значення ряду максимальне значення ряду розмах	$\frac{R}{\bar{y}}$ $\frac{MAD}{\bar{y}}$ ${\hat{\sigma}}/{y}$ $m_e = {y_n}/{2}$ $m_0 -$ характеризує величину, яка найчастіше спостерігається y_{min} y_{max} $R = y_{max} - y_{min}$

Таблиця 2.2 Основні характеристики випадкової вибірки

2.3 Метод Ірвіна

У процесі отримання даних, наприклад, при вимірюваннях, автоматичній обробці та передачі сигналів та ін, у вибірці іноді виходять аномальні значення (грубі помилки, промахи, викиди), обумовлені зміною умов, збій засобів вимірювань, помилки оператора. Аномальними називають значення, що різко відрізняються за величиною і статистичними властивостями від основної групи значень [42, с.9]. Частка аномальних значень може сягати 10 % і більше. Наявність їх може призводити до суттєвих спотворень результатів обробки даних. Для відбракування аномальних значень можна використовувати різні статистичні критерії. Більшість таких критеріїв ґрунтується на припущенні, що вибірку отримано з нормально розподіленої випадкової величини. До таких критеріїв належить, зокрема, критерій Ірвіна. При цьому нульова гіпотеза полягає в тому, що всі значення вибірки належать до того самого нормального розподілу. Конкуруюча гіпотеза полягає в тому, що сумнівні значення належать до іншого розподілу.

Критерій Ірвіна запропонований у 1925 р. у роботі [43] для відбракування аномальних значень із вибірки, отриманої з нормально розподіленої випадкової величини. В даний час критерій застосовується при статистичній обробці результатів вимірювань та випробувань, зокрема, при попередньому аналізі часових рядів, що використовуються при вирішенні соціально-економічних, технічних та природничо-наукових завдань.

Порядок застосування критерію Ірвіна, виходячи з [43], наступний. Значення вибірки розташовують у варіаційний ряд: $x_1, x_2, \ldots, x_i, \ldots, x_n$. Знаходять значення статистики Ірвіна λ для k-го значення варіаційного ряду за формулою

$$\lambda_{\text{расч}} = \frac{|x_k - x_{k+1}|}{\sigma} \qquad (2.15)$$

де k – номер значення x_i у варіаційному ряду, рахуючи з одного з кінців ряду, σ – генеральне середньоквадратичне відхилення.

Оскільки σ зазвичай невідомо з достатньою точністю, у [43] рекомендується використовувати вибіркове середньоквадратичне відхилення s як найкраще наближення до нього. Ця рекомендація зазвичай використовується практично. Значення s розраховують за формулою $s = \sqrt{\frac{1}{n-1}\sum_{i=1}^{n}(x_i - \bar{x})^2}$,

де $\bar{x}$ – середнє арифметичне значення вибірки, n – обсяг вибірки.

Перевірку починають із першого (з одного з кінців ряду) значення ряду, тобто, при $k = 1$, і продовжують, переходячи до наступних значень ряду, при $k = 2$ потім k = 3 і т.д. Якщо для k-го значення ряду

$$\lambda_{\text{рас}} > \lambda_{k,\alpha},$$

де $\lambda_{k,\alpha}$ – табличне значення критерію Ірвіна для k-го елемента ряду при вибраному рівні значимості α, то всі значення ряду від першого до k-го вважають аномальними і відкидають. Після цього наступні значення ряду перевіряють у тому порядку, приймаючи для найближчого наступного значення $k = 1$ і заново перераховуючи s. Такий підхід виглядає обґрунтованим: якщо виключені промахи, вибірку, що залишилася, логічно розглядати як випадкову і репрезентативну.

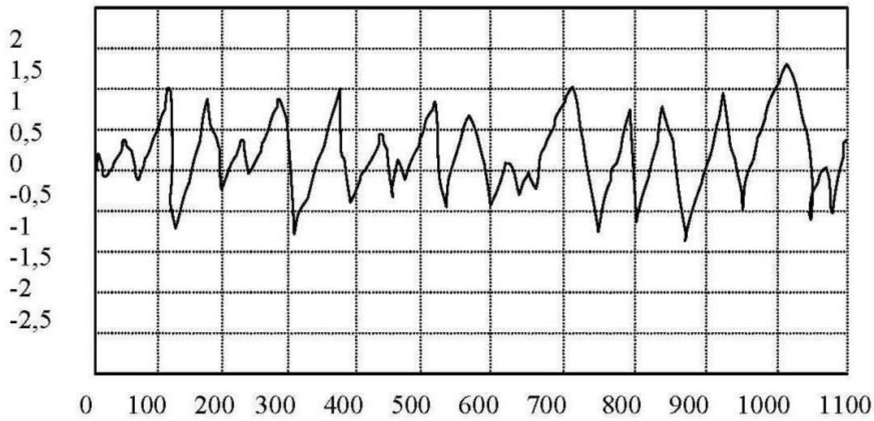
Перевіряти всі значення ряду, не потрібно, так як це може призвести до помилкового відбракування великої кількості значень. Доцільно перевіряти деяку кількість значень $k_{\text{пр}}$. Потім переходити до іншого кінця ряду та перевіряти також $k_{\text{пр}}$ значень. Далі чергувати перевірку з різних кінців ряду до тих пір, поки на обох кінцях не будуть відсутні аномальні значення. У цьому загальна кількість перевірених значень кожному кінці ряду має перевищувати $k_{\text{пр}}$.

Розділ 3. Стаціонарні часові ряди

3.1 «Білий шум»

Для того, щоб реально реалізувати задачу статистичного аналізу часових рядів, необхідно певним чином обмежити клас моделей часових рядів, що розглядаються, з точки зору структури послідовності та її ймовірнісних характеристик. Пошук відповідної моделі часового ряду зазвичай здійснюється в категорії стаціонарних часових рядів.

Узагальнений стаціонарний часовий ряд – це процес з математичним сподіванням і дисперсією. Це постійне значення, яке не змінюється з часом. Функція автоковаріації залежить лише від різниці між двома $t_1 - t_2 = \tau$ і не залежить від певного періоду час. Іншими словами, для досягнення випадкового



Таблиця 3.1 Стаціонарний часовий ряд

процесу $y_{t_1}, y_{t_2}, \dots, y_{t_n}$ основні моменти залишаються постійними й обмеженими у разі зміні часу t, для якого вони розраховуються, а саме:

- математичне сподівання: $M(y_t) = \mu_y < \infty$, для всіх t ;
- дисперсія:$Var(y_t) = D(y_t) = M(y_t - M(y_t))^2 = \sigma^2 < \infty$, для всіх t ;
- автоковаріація порядку τ:
- $Cov(y_t, y_{t+\tau}) = M((y_t - M(y_t))(y_{t+\tau} - M(y_{t+\tau}))) = y_\tau < \infty, \tau = 1,2,3, \dots$ для всіх t .

  Щоб отримати фактичну оцінку часового ряду, скористаємося такими формулами:

математичне сподівання:

$$\hat{\mu}_y = \frac{\sum_{t=1}^{n} y_t}{n};$$

дисперсія:

$$\hat{\sigma}_y^2 = \hat{\gamma}_0 = \frac{\sum_{t=1}^{n}(y_t - \hat{\mu}_y)^2}{n}; \qquad (3.1)$$

автоковаріація порядку τ:

$$\hat{\gamma}_\tau = \frac{\sum_{t=1}^{n-\tau}(y_t - \hat{\mu}_y)(y_{t+\tau} - \hat{\mu}_y)}{n-\tau}.$$

Часовий зсув τ називають ***часовим лагом.*** Зауважимо, що $Cov(y_t, y_{t+\tau})$, коли τ= 0, дорівнює дисперсії: $Cov(y_t, y_{t+0}) = Cov(y_t{}^2) = D(y_t) = \gamma_0$.

Крім того $|\gamma_k| \leq \gamma_0$. Можна розглядати функцію γ_k як усі можливі значення автоковаріацій, де τ перебирає цілочисельні значення від $-\infty$ до ∞. Сукупність значень автоковаріацій за всіх можливих значень τ називають ***автоковаріаційною функцією*** випадкового процесу. Автоковаріаційна функція стаціонарного часового ряду залежить лише від різниці в часі $(t_1 - t_2 = \tau)$. Ця функція є парною, і достатньо розглядати невід'ємні τ [37, с.29].

Коефіцієнт автокореляції між рівнями часового ряду зміщеного на τ, є автоковаріацією, розділена на корінь із добутку двох дисперсій. Оскільки дисперсія стала, отримуємо просто σ^2 або γ_0. Формула розрахунку коефіцієнта автокореляції: $\rho_\tau = \frac{\gamma_k}{\gamma_0}$. (3.2)

Вираз (3.2) визначає ***автокореляційну функцію*** $\boldsymbol{(ACF)}$ часового ряду, яка показує статистичну кореляцію значень часового ряду до різних зсувів у часі τ (наприклад, для річних спостережень один рік або два роки). Автокореляційна функція стаціонарного часового ряду залежить лише від різниць моментів часу $t_1 - t_2 = \tau$, і є парною функцією, тобто $\rho(\tau) = \rho(-\tau)$. Задаючи різні значення $\tau = 1,2,3, \ldots,$ отримують ряд значень ρ_1, ρ_2, ρ_3 ...

Графік автокореляційної функції називають ***корелограмою.*** За корелограмою можна визначити, що зміна показника y_t впливає на затримку його наступного значення.

Загалом, оцінки цих статистичних даних є несуперечливими, тобто для них існує межа ймовірності збігу з їх істинним значенням для генеральної сукупності.

Інтуїтивно можна очікувати, що деякі часові ряди соціально-економічних показників (якщо такі є) будуть стаціонарними, оскільки зростання та падіння значень є основними характеристиками соціально-економічних показників.

Для будь-яких часових рядів можна створити деякі моделі, які відображають розвиток процесу. Але яка б модель не була, неможливо безпомилково передбачити значення часового ряду. Звичайно, були розроблені деякі методи, щоб спробувати зменшити цю похибку, але ніколи не можна бути впевненим в абсолютній точності передбачення. Іншими словами, ніколи не було, і не буде моделі, яка б могла робити абсолютно точні прогнози. Тому будь-яка модель повинна мати доданок похибку, за величиною якої вимрюється точність моделі:

$$y_t = f(x_t) + \varepsilon_t$$

де x_t − набір екзогенних змінних.

Звичайно, така похибка може мати довільний розподіл, але при прогнозуванні стаціонарних часових рядів припускається, що похибка ε_t у кожний період часу t має властивості:

1) нульове математичне сподівання, тобто $\mu = E\varepsilon_t = 0$;
2) постійну дисперсію, тобто $\gamma_0 = var(\varepsilon_t) = \sigma^2$;
3) некорельованість елементів $\gamma_j = cov(\varepsilon_t - \varepsilon_{t-j}) = 0, j \neq 0$.

Будь-який процес, який задовольняє цим властивостям, називається **білим шумом [** 40].

3.2 Аналіз часових рядів за допомогою Бокса-Дженкінса

Г. Бокс і Г. Дженкінс вперше розглянули можливість використання моделі ARMA у 1976 році. Вони запропонували алгоритм для вибору моделі та об'єднали кілька напрямів у літературі [35].

Першим кроком є перегляд даних і визначення, чи є вибраний ряд стаціонарним. Якщо часовий ряд не є стаціонарним, Бокс і Дженкінс пропонують знайти його різницю до отримання стаціонарного процесу, оскільки перехід до першої різниці усуває лінійну тенденцію, друга різниця – квадратичну і так далі. Взагалі кажучи, щоб отримати стаціонарний ряд, можна знайти різницю d разів. При перетворенні змінних для отримання стаціонарного ряду важливо уникати знаходження різниці вищого порядку, ніж потрібно. Це трапляється, коли знаходимо різницю в стаціонарному ряді. Простою перевіркою є перевірка дисперсії ряду на зростання. Знаходження різниць нестаціонарних рядів дає – здебільшого ряди з меншою дисперсією.

Після отримання стаціонарної послідовності знаходять модель $ARMA$ на основі вибіркових даних. На цьому етапі дуже корисний графічний метод, а також порівняння функцій автокореляції та часткової автокореляції з відповідними функціями відомого процесу $ARMA$, наведеними в таблиці. 3.1.

Модель	ACF	$PACF$
Білий шум	Всі нулі	Всі нулі
$MA(1)$	Нулі після p_1	Спадна після p_{11}
$MA(2)$	Нулі після p_2	Спадна після p_{22}
$MA(q)$	Нулі після p_q	Спадна після p_{qq}
$AR(1)$	Спадна після p_1	Нулі після p_{11}
$AR(2)$	Спадна після p_2	Нулі після p_{22}
$AR(p)$	Спадна після p_p	Нулі після p_{pp}
$ARMA(1,1)$	Спадна після p_1	Спадна після p_{11}
$ARMA(p,q)$	Спадна після p_p	Спадна після p_{qq}

Таблиця 3.1. Характеристика *ARMA*- моделей

На практиці найчастіше використовують модель $ARMA$ першого порядку. Цей метод передбачає наступну послідовність процедур:

1) ідентифікація моделі часових рядів.
2) Оцінка параметрів моделі.
3) Діагностика побудованої моделі.
4) Використання моделі для прогнозування майбутніх значень часового Ряду [37, с.41].

Ці процедури можна повторювати багато разів у процесі уточнення моделі. Розглянемо кожен етап алгоритму детальніше.

- **Ідентифікація моделі.** При побудові моделі аналізу часових рядів однією з проблем з його визначенням є те, що кількість параметрів є найменшою. Ця проблема називається ідентифікацією.

Взагалі кажучи, якщо ми використовуємо вибірку спостереження, її розмір зазвичай відносно невеликий, і можна очікувати, що не буде точної відповідності між даними та теоретичною моделлю. Це призведе до вибору двох або трьох тестових моделей $ARMA\ (p, q)$ на цьому кроці, які мають кілька пар часових лагів p в авторегресійному процесі та лагову змінну q у моделі ковзного середнього. Вибір кількох моделей найбільш придатний для подальшого аналізу та прогнозування за допомогою методів діагностичної перевірки, про які йтиметься нижче.

- **Оцінка параметрів моделі.** Після процесу ідентифікації, який визначив початковий варіант стаціонарної моделі моделі $ARMA$, яка адаптується до цих спостережень шляхом знаходження оцінок параметрів $a = (a_1, \dots, a_p)$ та $b = (b_1, \dots, b_q$. Раніше було показано, що модель порядку $ARMA\,(p, d, q)$, що розглядає нестаціонарні процеси, зводиться до моделі стаціонарного порядку $(p, 0, q)$ через різницю першого порядку. Тому в процесі розрахунку коефіцієнтів розглядається лише стаціонарна модель.

Параметри моделі AR можна оцінити за допомогою звичайного методу найменших квадратів (виходять зсунуті, але консистентні оцінки), і його не можна застосувати до моделей MA або $ARMA$. Наприклад, для моделі MA (1):

$$y_t = \mu + \varepsilon_t - b_1 \varepsilon_{t-1}$$

Параметр не можна оцінити, використовуючи лише спостережуване значення y_t, оскільки значення параметра, використане для обчислення, невідоме.

Метод Бокса-Дженкінса рекомендує використовувати програму нелінійної оптимізації: програму пошуку сітки. Це ітераційний процес, у якому оцінки параметрів мінімізують суму квадратів залишків. Запишемо модель МА (1) як $\varepsilon_t = y_t - \mu + b_1 \varepsilon_{t-1}$. Аналізуючи розрахункові значення ACF і $PACF$, можна зробити припущення про значення параметрів. Можна використовувати вибіркове середнє (для μ) і першу автокореляцію (для b). Припустимо, що вони 100 і 0,2. Тоді модель виглядає як $\varepsilon_t = y_t - 100 + 0{,}2\varepsilon_{t-1}$. Припускаючи, що ε_0 дорівнює нулю, ми можемо отримати оцінку ε_t для t від 1 до n і обчислити суму квадратів залишку s_1. Вибір початкових значень μ і b дає нове значення s_2 залишкової суми квадратів. Потім перевіряються інші вихідні дані і остаточними оцінками стають значення коефіцієнтів моделіі, за якими s є мінімальним.

В епоху Боксу і Дженкінса через великі обмеження на використання комп'ютерів, для оцінювання коефіцієнтів розроблялись окремі методи для кожної моделі. Зараз вчені розробили загальний метод максимальної правдоподібності, який дозволяє отримання консистентних та асимптотично ефективних оцінок коефіцієнтів для будь-якої моделі.

Основна ідея використання цього методу полягає в тому, щоб припустити, що дані мають певний розподіл ймовірностей, і розрахувати ймовірність очікуваної події. Зазвичай це залежить від якихось невідомих параметрів. Використовуючи дані, ви можете максимально збільшити ймовірність цієї події. Коефіцієнт досягнення максимальної ймовірності відповідної події є необхідною оцінкою параметра. Іноді буває важко знайти ці оцінки в аналітичній формі. У цьому випадку для оптимізації функції правдоподібності використовуються чисельні методи.

Будемо виходити зі спостережень $z = (z_1, z_2, \dots, z_n)$ (для цього нам потрібно $N = n + d$ спостережень $y_{-d+1}, \dots, y_0, y_1, \dots, y_n$ над початковим рядом y_t). Запишемо модель *ARMA*(p, q)-процес у вигляді:

$$z_t = \sum_{i=1}^{p} a_i y_{t-i} + \varepsilon_t + \sum_{j=1}^{q} b_j \varepsilon_{t-j} \tag{3.1}$$

Цей процес містить $p + q + 1$ невідомих параметрів: $a_i, b_j, \sigma_\varepsilon^2$.

Нехай $A = (a_1, a_2, \dots, a_p)^T$, $B = (b_1, b_2, \dots, b_q)^T$, $Z = (z_1, z_2, \dots, z_n)^T$, визначимо матрицю V(A, B) таким чином, що $Dz_t = \sigma^2 V(A, B)$. Також нехай ε_t мають нормальний розподіл. Тоді логарифм функції правдоподібності має вигляд:

$$lnL(\sigma^2, A, B) = -\frac{1}{2}\left(n\, ln\sigma^2 + \ln V(A, B) + \frac{Z^T \left(V(A,B)\right)^{-1} Z}{\sigma^2}\right. \tag{3.2}$$

Оцінки σ^2, A, B отримують шляхом максимізації заданого логарифма функції правдоподібності. Існують більш ефективні способи обчислення функції правдоподібності.

Функція правдоподібності. Розглянемо побудову функції правдоподібності більш детально. Нехай $y_t, t = 1,2,\dots,n$ — вибірка, з розподілом ймовірностей $P(y_t \backslash A)$, де A — набір невідомих параметрів. Припускаємо, що y_t є незалежними, кожен має розподіл ймовірностей $P(y_t \backslash A)$, а сумісний розподіл всієї сукупності $y_t, t = 1,2,\dots,n$ виражається формулою:

$$P(y_1, y_2, \dots, y_t \backslash A) = P(y_1 \backslash A)\, P(y_2 \backslash A) \dots P(y_t \backslash A) = \prod_{t=1}^{n} P(y_t \backslash A) \quad (3.3)$$

Щоб відповісти на це питання, яке значення А максимізує ймовірність породження моделлю саме вибірки $y_t, t = 1,2,\dots,n$ потрібно максимізувати функцію правдоподібності [37, с.44]:

$$L(A) = \prod_{t=1}^{n} P(y_t \backslash A) \quad (3.4)$$

Для подальшої оптимізації потрібно точно знати розподіл вибірки.

Припустимо, що аналізується модель

$y = f(X, a) + \varepsilon,$

де y — часовий ряд,

X — матриця екзогенних змінних,

ε — вектор збурень, який має нормальний розподіл із нульовим вектором математичних сподівань та коваріаційною матрицею θ. Тоді функція правдоподібності буде виглядати так:

$$L(a) = (2\pi)^{-\frac{n}{2}} \theta^{-\frac{1}{2}} e^{-\frac{1}{2}(y - f(X,a))^t \theta^{-1} (y - f(X,a))} \quad (3.5)$$

Загалом функція правдоподібності та її логарифм не є лінійними, тому важко знайти максимальне значення функції правдоподібності в аналітичному вигляді. Тому для знаходження максимального значення функції потрібно використовувати числовий метод, наприклад метод Гауса. Загальний алгоритм включає наступні кроки [37, с.45]:

1. Встановити початкове значення для вектора A_i .
2. Визначити напрям руху для A_i , в якому значення L(A_i) збільшується.
3. Визначити довжину кроку й обчислити нове значення A_{i+1}.
4. Перевірити критерій зупинки.

Якщо алгоритм потребує продовження, то задаємо $i = i + 1$ і повертаємося до кроку 2.

Звичайним критерієм зупинки є $|L(A_{I+1}) - L(A_I)| < \delta$, де δ — наперед задане мале число.

Використаємо відповідні функції правдоподібності, щоб перевірити гіпотезу. Розглянемо критерії перевірки гіпотези H^0 проти альтернативної H^1 у загальному випадку. Тестову статистику в основному поділяють на три категорії: тест Вальда, тест з використанням множника Лагранжа та тест на основі відношення функції правдоподібності. Усі ці критерії засновані на максимізації функцій правдоподібності. Вони асимптотично еквівалентні. Основна відмінність цих трьох методів полягає у виборі оцінки для розрахунків оціночної вартості. Метод відношення правдоподібності є найстарішим із цих тестів, розробленим у 1928 році Нейманом і Пірсоном. Суть цього методу полягає в порівнянні значень функції правдоподібності за умов (з обмеженнями) і без урахування (без обмежень). Наприклад, нехай без урахування H^0 оцінкою є $A^\wedge$, при врахуванні умови H^1 оцінкою буде $A^\wedge$. Тоді

$$LR = \frac{L(A^\wedge)}{L(A^\wedge)} > 1, \tag{3.6}$$

оскільки із визначення максимуму функції правдоподібності $L(A^\wedge) > L(A^\wedge)$.
Необхідно визначити, якою може бути величина $L(A^\wedge) - L(A^\wedge)$, щоб можна було прийняти гіпотезу H^0, тобто чи є обмеження, включені до H^0 суттєвими. Відповідна статистика така:

$$LRT = 2(logL(A^\wedge) - logL(A^\wedge)) \approx \chi^2_m. \tag{3.7}$$

Тому, щоб перевірити гіпотезу, значення LRT необідно обчислити і порівняти з χ^2_m — статистикою, де кількість ступенів свободи визначається кількістю обмежень у гіпотезі H^0. Якщо $LRT > \chi^2_m$, то гіпотеза H^0 відхиляється.

3.3 Використання моделі ARIMA для прогнозування часових рядів

У моделі ARIMA при прогнозуванні змінної в певний момент у майбутньому, лагові значення цієї змінної, які є пояснюючими змінними (регресорами) моделі, можна вважати фіксованим на вибірковому значенні або випадковим. Перша можливість призводить до умовного прогнозу, таких як множинна регресія, а друга можливість призводить до безумовних передбачень. Тому при використанні для прогнозування такої моделі, як ARIMA, враховуються як умовні, так і безумовні прогнози. Як ми всі знаємо, умовна дисперсія випадкових величин не буде перевищувати її безумовну дисперсі, тому точність умовного прогнозу завжди вище.

Якщо модель задана правильно, може бути два джерела помилок прогнозу: невизначеність майбутнього значення випадкових величин ε та відсутність точних значень коефіцієнтів моделі (тільки їх розрахункових значень).

При використанні моделі ARIMA для прогнозування оцінене значення коефіцієнта моделі та значення регресора залежать від доступних вибірок, тому через спостереження часового ряду важко проаналізувати умовну дисперсію, що виражає помилку прогнозу. . Зазвичай вони обмежуються припущенням, що коефіцієнти точно відомі. Очевидно, що таке припущення зменшує дисперсію помилок передбачення, тим самим покращуючи гіпотетичну точність умовних і безумовних передбачень.

Щоб досягти мінімальної середньоквадратичної помилки (MSE), візьмемо умовне математичне сподівання: $M\{y_{t+\tau} / y_1, \dots y_t\}$.

Прогноз моделі $MA(q)$: $y_t = \theta + \varepsilon_t + b_1\varepsilon_{t-1} + b_2\varepsilon_{t-2} + \dots + b_q\varepsilon_{t-q}$.

Якщо коефіцієнти моделі точно відомі, і є значення y_t для $t \in [1, n]$, то безумовним прогнозом точки в будь-який момент часу буде математичне сподівання процесу, тобто θ. Умовним прогнозом в моменту часу $t + 1$ буде умовне математичне сподівання:

$$\hat{y}_t(1) = M\{\theta + \varepsilon_{t+1} + b_1\varepsilon_t + \dots + b_q\varepsilon_{t-q+1} | y_1, \dots, y_t\}. \qquad (3.8)$$

Випадкова величина ε зліва пов'язана із спостереженнями. Оскільки спостережуване значення складається з «модельного значення» та помилки, то умовне математичне сподівання всіх компонентів, окрім ε_{t+1}, не дорівнює нулю.

Наприклад, $M\{\varepsilon_t / y_1, \dots y_t\}$ це залишок між спостереженням моделі та розрахунком (прогнозуванням), тобто $e_t = y_t - \widehat{y_t}$. Тому умовне математичне сподівання всіх минулих значень від випадкової складової слід замінити на відповідні залишки. Прогноз також побудований так, щоб просуватися на 2 або більше кроків. Усі майбутні замінюються нулями, а минулі — обчислюваними залишками. Тому для моделі $MA(q)$ прогноз залежить від похибки на попередньому кроці. Починаючи з кроку $(q+1)$, умовним прогнозом є математичне сподівання θ, тобто умовний прогноз збігається з безумовним прогнозом.

Умовна дисперсія помилки прогнозу на 1 крок випередження становить:

$$D(y_{t+1} - \hat{y}_t(1) | y_1, \dots, y_t) =$$
$$M\{((\theta + \varepsilon_{t+1} + b_1\varepsilon_t + \dots + b_q\varepsilon_{t-q+1} - \theta - b_1 e_t - \dots - b_q e_{t-q+1}) | y_1, \dots, y_t)^2\} =$$
$$\sigma_E^2 \tag{3.9}$$

Аналогічно, дисперсія, передбачена, на 2 кроки вперед, дорівнює:

$$D(y_{t+2} - \hat{y}_{t+2} | y_1, \dots, y_t) = (1 + b_1^2)\sigma_\varepsilon^2 \tag{3.10}$$

а дисперсія на τ кроків дорівнює

$$(1 + b_1^2 + \dots + b_\tau^2)\sigma_\varepsilon^2 \text{ для } \tau < q.$$

Якщо $\tau \geq q$, дисперсія помилки умовного прогнозу така ж, як і безумовного прогнозу, що дорівнює дисперсії випадкового процесу y_t .

Прогноз моделі $AR(p)$: $y_t = \theta + a_1 y_{t-1} + a_2 y_{t-2} + \dots + a_p y_{t-p} + \varepsilon_t$.

Щоб спрогнозувати на крок вперед, можна написати:

$$\hat{y}_t(1) = M\{y_{t+1} | y_1, \dots, y_t\} = M\{\theta + a_1 y_t + \dots + a_p y_{t-p+1} + \varepsilon_{t+1} | y_1, \dots, y_t\} =$$
$$\theta + a_1 y_t + \dots + a_p y_{t-p+1} \tag{3.11}$$

Тобто у рівняння моделі підставляють p минулих значень реалізації часового ряду. Для прогнозу на два кроки вперед отримують:

$$\hat{y}_t(2) = M\{y_{t+2}|y_1, \dots, y_t\} =$$
$$M\{\theta + a_1 y_{t+1} + \dots + a_p y_{t-p+2} + \varepsilon_{t+2}|y_1, \dots, y_t\} \qquad (3.12)$$

Математичне сподівання від випадкової похибки ε знову дає 0, умовне математичне сподівання від $y_1, y_2, \dots, y_t$ дорівнює такому ж значенню, але цей вираз містить умовне математичне сподівання y_{t+1}, отримане на попередньому кроці. Через значення реалізації ви можете замінити його вираз і отримати розширену формулу. Насправді зручніше розглядати рекурентне співвідношення, яке пов'язує послідовні значення прогнозу. Цей зв'язок є лінійним різницевим рівнянням p –го порядку, і його розв'язок прагне, якщо збільшується t , до величини $\frac{\theta}{1-a_1-\dots-a_p}$, тобто знов таки до безумовного прогнозу.

Розрахунок умовної дисперсії помилки прогнозу подібний до випадку моделі ковзного середнього, але навіть для моделей малого порядку докази стають досить великими. Наприклад, для моделі *AR*(2) без вільного члена прогноз на крок вперед становить: $\hat{y}_t(1) = a_1 y_t + a_2 y_{t-1}$.

Очевидно, що дисперсія помилки прогнозу на 1 крок дорівнює:

$$D(y_{t+1} - \hat{y}_t(1)|y_1, \dots, y_t) =$$
$$\{((\theta + \varepsilon_{t+1} + b_1\varepsilon_t + \dots + b_q\varepsilon_{t-q+1} - \theta - b_1 e_t - \dots - b_q e_{t-q+1})|y_1, \dots, y_t)^2\} = \sigma_E^2$$
$$(3.13)$$

Для прогнозу на 2 кроки відповідно отримуємо:

$$\hat{y}_t(2) = a_1\hat{y}_t(1) + a_2 y_t = a_1(a_1 y_t + a_2 y_{t-1}) + a_2 y_t$$
$$y_{t+2} = a_1 y_{t+1} + a_2 y_t + \varepsilon_{t+2} = a_1(a_1 y_t + a_2 y_{t-1} + \varepsilon_{t+1}) + a_2 y_t + \varepsilon_{t+2}$$

Дисперсія помилки прогнозу на 2 кроки вперед дорівнює:

$$(1 + b_1^2)\sigma_\varepsilon^2. \qquad (3.14)$$

Для прогнозу на 3 кроки отримуємо:

$$\hat{y}_t(3) = a_1\hat{y}_t(2) + a_2\hat{y}_t(1) = a_1(a_1(a_1 y_t + a_2 y_{t-1}) + a_2 y_t) + a_2(a_1 y_t + a_2 y_{t-1})$$
$$y_{t+3} = a_1 y_{t+2} + a_2 y_{t+1} + \varepsilon_{t+3} = a_1(a_1(a_1 y_t + a_2 y_{t-1} + \varepsilon_{t+1}) + a_2 y_t + \varepsilon_{t+2}) +$$
$$a_2(a_1 y_t + a_2 y_{t-1} + \varepsilon_{t+1}) + \varepsilon_{t+3}.$$

Дисперсія помилки прогнозу на 3 кроки дорівнює:

$$(1 + b_1^2 + b_2^2 + 2b_1^2 b_2 + b_1^4)\sigma_\varepsilon^2. \tag{3.15}$$

Очевидно, що дисперсія помилки прогнозу буде збільшуватись з кожним кроком.

Значно легше отримати вираз для дисперсії помилки прогнозу, якщо перейти від $AR(p)$ представлення до еквівалентного MA представлення: $y_t = \theta + \varepsilon_t + \psi_1 \varepsilon_{t-1} + \dots + \psi_q \varepsilon_{t-q} + \cdots$ має нескінченну кількість компонентів. Тоді дисперсію помилки прогнозу на τ кроків можна виразити формулою:

$$\sigma_\varepsilon^2 \sum_{j=0}^{\tau-1} \psi_j^2 \ (\psi_0 = 1) \tag{3.16}$$

Для загальної моделі $ARMA(p,q)$ необхідно об’єднати весь вищезгаданий вміст. Відповідно до моделі, підставляючи туди для часу $1, t$ спостереження y_t та розраховані значення залишків, обчислюють прогнозовані значення y_t. Для майбутніх моментів замінюємо залишок на нуль і замість y_t підставляють їхні прогнозовані значення. Дисперсію похибки передбачення розраховують за формулою (3.16).

Наприклад, для моделі $ARMA(1,1)$:

$$y_t = (1 - aL)^{-1}(1 + bL)\varepsilon_t = (1 + aL + a^2L^2 + \cdots)(1 + bL)\varepsilon_t$$
$$= \varepsilon_t + \eta_1 \varepsilon_{t-1} + \eta_2 \varepsilon_{t-2} + \cdots,$$

де, $\eta_1 = a + b$, $\eta_2 = a(a + b), \ldots, \eta_k = a^{k-1}(a + b)$, починаючи з другого коефіцієнти спадають за геометричною прогресією. Звідси можна легко обчислити дисперсію помилки прогнозу на τ кроків:

$$\sigma_\varepsilon^2 \sum_{j=0}^{\tau-1} \eta_j^2 = \sigma_\varepsilon^2 (1 + \sum_{j=0}^{\tau-1} a^{2j} (a+b)^2) = \sigma_\varepsilon^2 (1 + \frac{(a+b)(1-a^{2\tau})}{1-a^2}). \tag{3.17}$$

Для цієї моделі дисперсія похибки прогнозу асимптотично дорівнює дисперсії часового ряду.

У всіх розглянутих випадках прогноз умовної точки асимптотично наближається до математичного сподівання послідовності, а дисперсія помилки прогнозу наближається до дисперсії послідовності. Це означає, що для стаціонарного процесу вплив наявної інформації на прогноз і його точність поступово падає до нуля. Крім того, у міру збільшення діапазону прогнозу дисперсія помилки не буде перевищувати дисперсію часового ряду. На жаль, цей висновок є результатом нереалістичних припущень, що коефіцієнти моделі відомі точно [40, с.39].

3.4 Огляд програмних пакетів аналізу та прогнозування часових рядів

У наш час існує досить велика кількість статистичних програм, які використовують для прогнозування та аналізу часових рядів. Їх умовно поділяють на дві групи: універсальні системи обробки та аналізу статистичної інформації, що реалізують широкий спектр модулів для розв'язання різних задач у сфері data mining, у тому числі задач прогнозування рядів динаміки, та пакети, спеціалізовані саме на аналізі та прогнозуванні часових рядів.

До першої групи віднесять такі програмні продукти: R, Python, MATLAB, STATISTICA, Stata, SAS, SPSS, KXEN, Weka, NCSS, GMDH Shell та UNISTAT. До другої комплекси Sales-Forecast, Forecast Pro та Zaitun Time Series. Нижче наведено опис кожного з вище перерахованих програних пакетів.

1) R – одночасно мова програмування та вільне програмне середовище для статистичної обробки даних і роботи з графікою. Для роботи з часовими рядами R містить бібліотеки forecast, hybridForecast, AnalyzeTS, тощо. Реалізовані в бібліотеках функції дозволяють проводити декомпозицію часового ряду на складові, згладжування, пошук аномалій, прогнозування за допомогою нейронних мереж, моделей ARIMA, нечіткої логіки, тощо.

2) Python – високорівнева мова програмування загального призначення. Бібліотеки statsmodels.tsa (Time Series Analysis), statsmodels.tsa.statespace, tsa.vector_ar, містять моделі та класи функцій, корисні для аналізу часових рядів. Такі інструменти включають моделі прогнозування AR, MA, ARMA, SARIMAX, VAR, VARMA, VARMAX, процедури згладжування ковзким середнім та експоненційне згладжування, модель Хольта-Вінтерса, побудова графіків автокореляційної функції та періодограми, декомпозиція ряду, перевірка ряду на стаціонарність та інші статистичні тести [15-17].

3) MATLAB – пакет прикладних програм для розв'язання задач технічних розрахунків і назва мови програмування, що використовується в цьому пакеті. Для роботи з часовими рядами у середовище Matlab вбудований пакет System Identification Toolbox. Він включає реалізацію моделей лінійної та нелінійної регресії, моделі AR, ARMAX, ARIMA, ARX та інші [4].

4) STATISTICA – це інтегрована система аналізу та управління даними, а також інструмент розробки додатків для бізнесу, економіки, фінансів, промисловості, медицині, страхуванні та інших областей. Включає модуль для обробки та аналізу часових рядів, який дозволяє проводити декомпозицію рядів динаміки, згладжування, кореляційний та спектральний аналіз, обчислення статистик часового ряду, а також прогнозування за допомогою моделі ARIMA та нейронних мереж [18,39].

5) Stata – це пакет статистичних прикладних програм, який включає модуль для обробки та аналізу часових рядів. Щодо аналізу часових рядів, Stata дозволяє будувати моделі прогнозування ARIMA, ARCH/GARCH, містить реалізацію процедур Кохрейна-Оркутта, Прейс-Вінстена, оцінки Ньюї-Веста, корелограми та періодограми [14].

6) Пакет SAS (Statistical Analysis System) – це професійний статистичний пакет, розроблений компанією SAS Institute Inc. Програма SAS включає реалізацію моделі Бокса-Дженкінса (ARIMA), регресійних та авторегресійних моделей прогнозування, процедуру експоненціального згладжування, спектральний та фазовий аналіз часових рядів [12,34].

7) SPSS (Statistical Package for the Social Sciences) – комп'ютерна програма статистичної обробки даних для проведення прикладних досліджень в соціальних науках. Програма SPSS включає реалізацію моделей ARIMA та експоненційного згладжування, засобів ідентифікації аномальних значень у часових рядах [3].

8) KXEN – інструмент моделювання, в основі якого лежить підхід Data Mining. Пакет KXEN включає модуль для аналізу часових рядів KXEN Time Series, або KTS. Він дозволяє виділяти значущі шаблони і тренди в часових

рядах, а також періодичну, сезонну складову та залишки. Дозволяє будувати короткі та середньострокові прогнози [11,32].

9) Weka (Waikato Environment for Knowledge Analysis) – безкоштовне програмне забезпечення для аналізу даних. Версія Weka 3.7.3 та подальші містять розширений модуль для аналізу та прогнозування часових рядів, який включає виділення складових рядів, побудову прогнозної моделі та оцінку її якості [19].

10) NCSS – це комп’ютерна програма для аналізу статистичних даних. NCSS включає в себе більше 230 задокументованих статистичних процедур для аналізу даних та прогнозування. Серед можна виділити такі: моделі ARIMA та ARMA, методи декомпозиції часового ряду на складові, експоненційне згладжування, гармонічний аналіз, лагова трансформація, рівняння Юла-Уолкера, аналіз сезонної та трендової складової, спектральний аналіз, обчислення автокореляції залишків моделі ряду [6,7,5].

11) GMDH Shell – це програмний інструмент для інтелектуального аналізу даних та прогнозування на основі МГУА (методу групового урахування аргументів). Програмний пакет дає можливість переглядати формулу побудованої моделі часового ряду та отримані прогнози у зручному вигляді таблиць і графіків. Дозволяє працювати з різними типами електронних таблиць та баз даних [2].

12) UNISTAT – програмний пакет для статистичного аналізу даних. Програмний комплекс містить реалізацію таких засобів роботи з часовими рядами: моделі ARIMA та експоненційного згладжування, тест тренду Ньюманна, побудова регресії Кокса, аналіз Каплана-Мейер та Вейбулла, діаграми Паретто, перетворення Фур’є, тощо [21].

13) Sales-Forecast – це програмна система, яка дозволяє в автоматичному режимі будувати прогнози різного ступеня деталізації для великого числа товарів і товарних груп багатономенклатурної продукції, враховуючи при цьому календар подій та індивідуальні особливості часових рядів. У Sales-Forecast втілені відомі засоби прогнозування та аналізу часових

рядів: моделі ARIMA, моделі експоненціального згладжування, нейронні мережі, методи сезонної декомпозиції та спектрального аналізу [13].

14) Forecast Pro – програмне забезпечення для побудови прогнозів, орієнтоване на бізнес-застосування. Forecast Pro забезпечує автоматичний підбір прогнозної моделі з наступного набору: моделі експоненційного згладжування, авторегресійні моделі Бокса-Дженкінса, моделі переривчастого попиту Кростона; модель Census X11; моделі дискретного попиту; регресійні моделі [10].

15) Zaitun Time Series представляє собою практичний інструмент для статичного аналізу, моделювання та прогнозування часових рядів. Пакет дозволяє обчислювати статистичні характеристики ряду, будувати нейронні мережі, містить графічні інструменти, які полегшують аналіз динамічних процесів. Програма дозволяє виділяти та аналізувати присутні у ряді тенденції, проводити його декомпозицію та згладжування, будувати корелограму [23]. Окрім цього, Zaitun Time Series має можливість роботи з даними ринку онлайн.

Більшість з представлених вище пакетів (наприклад, Sales Forecast, Forecast Pro, UNISTAT, GMDH Shell, SPSS) характеризуються високою вартістю ліцензій, що суттєво обмежує їхню доступність для широкого кола користувачів. Інші, наприклад R, є вільними у розповсюдженні, але вимагають від користувача д рівня кваліфікації в області програмування та статистики. Існують також безкоштовні пакети, які є простими в застосуванні, наприклад Zaitun Time Series. Однак зазвичай такі програмні системи містять дуже обмежену кількість методів прогнозування часових рядів. Також, незважаючи на зростаючий інтерес до ансамблевих моделей прогнозування, кількість програмних пакетів, які містять їхню реалізацію, є обмеженою. Тому виникає потреба у створенні програмного продукту, який містить реалізацію ансамблевих моделей, надає можливість порівняння результатів прогнозування, отриманих різними методами, є простим в освоєнні як для фахівців, так і для непідготовлених користувачів.

Практичне досліджування

Я використовувала пакет STATISTICA для аналізу різних моделей часових рядів, який призначений для обробки статистичних даних різними методами. Він включає сім модулів, один з яких підходить для моєї роботи. Цей модуль називається: Time Series / Forecasting – аналіз часових рядів і прогнозування. Наприклад, якщо вам потрібно вивчити будь-який фізичний процес і оцінити спектр потужності, або якщо ви спостерігаєте економічний ряд даних, наприклад, ціни на акції, і хочете передбачити це на кілька днів наперед, то модуль Time Series / Forecasting – може допомогти виконати ці завдання.

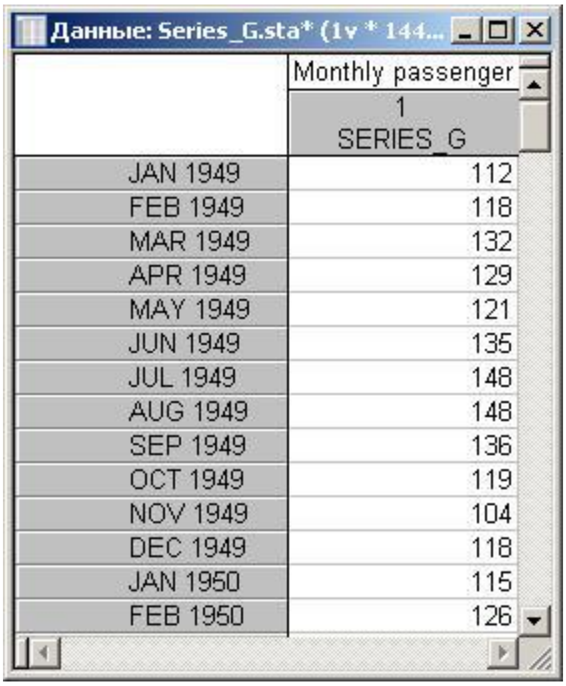
Головна панель цього модуля, дозволяє викликати такі модулі аналізу часових рядів:

- ARIMA та автокореляційна функція – модель авторегресії та інтегрованого ковзкого середнього (метод Бокса-Дженкінса).
- Interrupted time series analysis – аналіз часових рядів з інтервенцією.
- Exponential smoothing and forecasting – експоненціальне згладжування та прогнозування.
- Seasonal decomposition – сезонна декомпозиція.
- X11/У2k-monthly – аналіз часових рядів з урахуванням місяців і кварталів.
- Distributed lags analysis – аналіз часових рядів з розподіленими лагами (регресійна модель для двох часових рядів).
- Spectral (Fourier) analysis – спектральний аналіз Фур’є [45, с.11].

На основі статистичних даних наведемо приклад побудови деяких моделей часових рядів.

Приклад 1. Використаємо класичний приклад автора Бокса і Дженкінса ряд G, у нас є данні, які представляють собою міжнародні місячні авіаперевезення за 12 послідовних років з 1949 по 1960 роки. Присвячений цей приклад, моделі авторегресії та інтегрованого ковзкого середнього ARIMA.

Це типові дані про перевезення. Такі дані виникають під час аналізу залізничного транспорту, автомобільного транспорту, водного транспорту тощо. До них застосовуються методи, що описуються нами.



Данные: Series_G.sta* (1v * 144...

	Monthly passenger 1 SERIES_G
JAN 1949	112
FEB 1949	118
MAR 1949	132
APR 1949	129
MAY 1949	121
JUN 1949	135
JUL 1949	148
AUG 1949	148
SEP 1949	136
OCT 1949	119
NOV 1949	104
DEC 1949	118
JAN 1950	115
FEB 1950	126

Рисунок 4.1 Дані авіаперевезень

Ряд має чітко зростаючий тренд, а також сезонну складову (наприклад, у березні перевезення зазвичай вище, ніж у лютому та квітні).

1) У програму завантажуємо файл з даними series_g.sta
2) Відкриваємо вкладку Statistics обираємо Advanced Linear /Nonlinear Models, і натискаємо на вкладку часові ряди та прогнозування (Time Series / Forecasting).

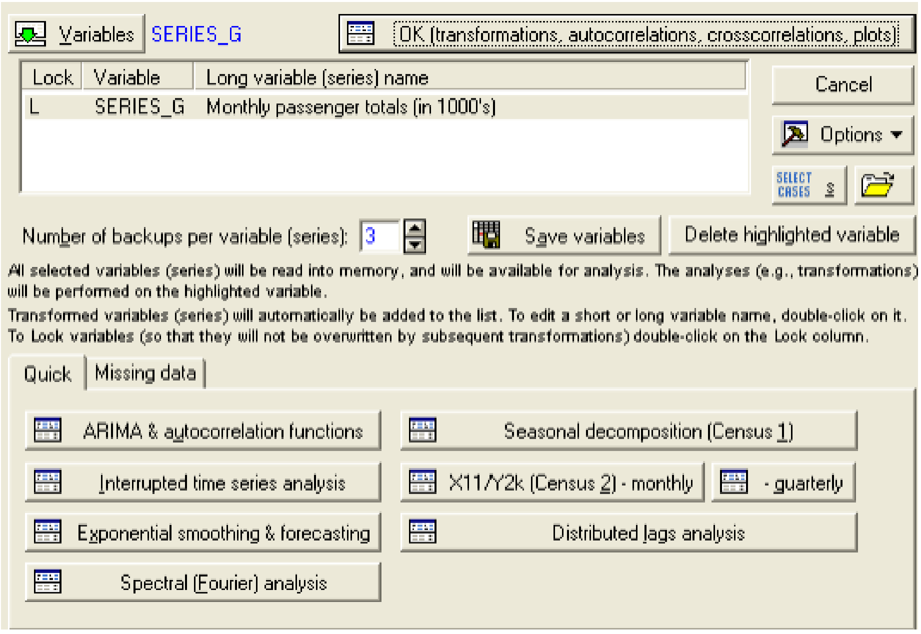


Рисунок 4.2 Панель аналізу часових рядів та прогнозування

3) Розберемо часовий ряд на складові, натиснувши кнопку Сезонна декомпозиція.

Переходимо на вкладку додатково (Advanced). Виберемо адитивну сезонну модель (Additive seasonal model).

Встановлюємо кількість резервних копій (Number of backups per variable) для кожного ряду на 10.

Відмічаємо всі параметри для розкладання ряду в області «On OK append components to active work area».

Натискаємо кнопку підсумок (Summary: Seasonal decomposition), щоб отримати таблицю, яка розбиває ряд на компоненти.

Отримуємо відповідні графіки:

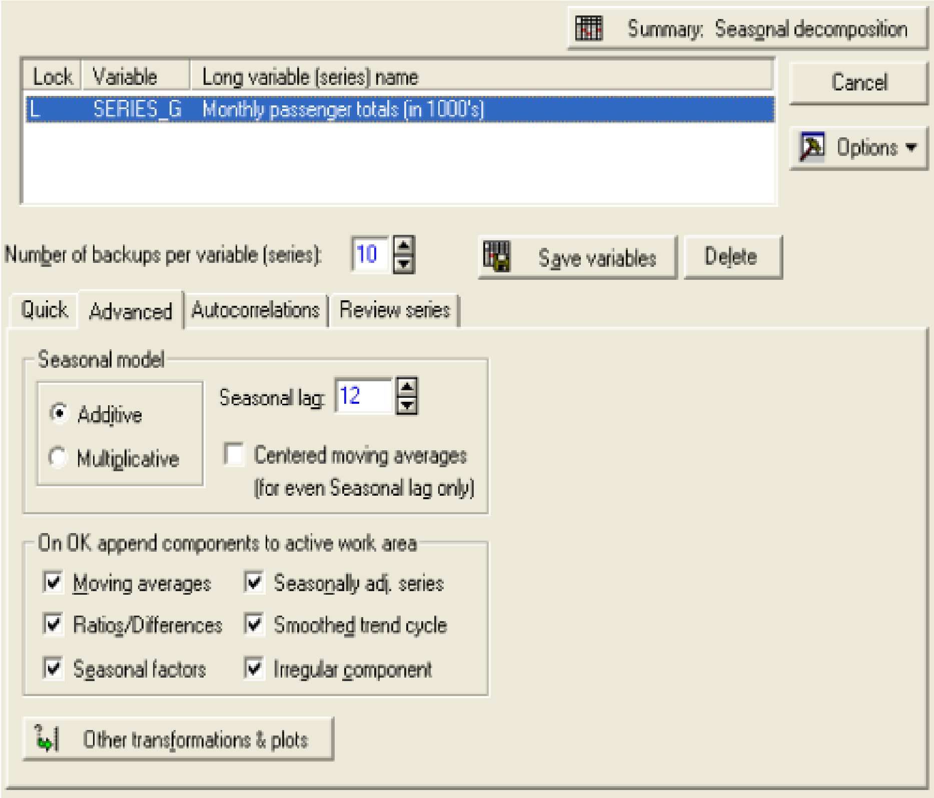
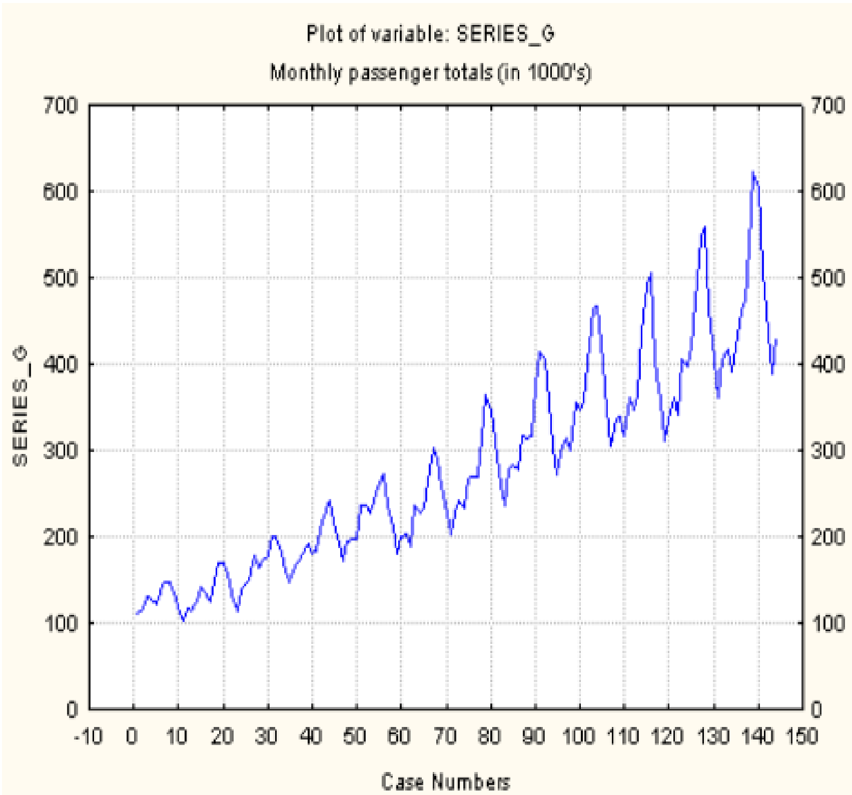
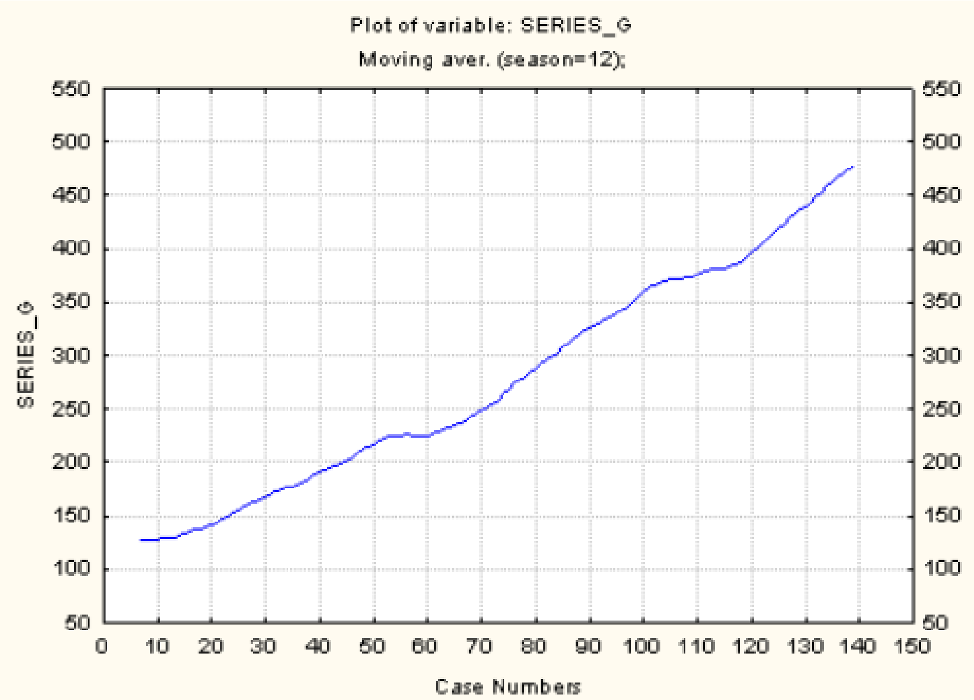


Рисунок 4.3 Модуль декомпозиції часових рядів на компоненти

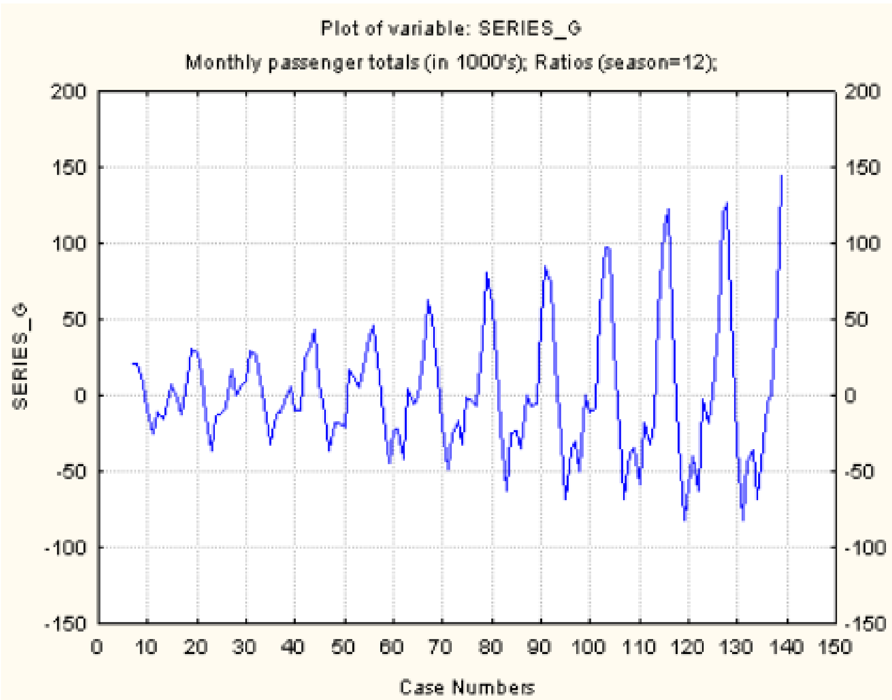
SERIES_G – вихідний часовий ряд.



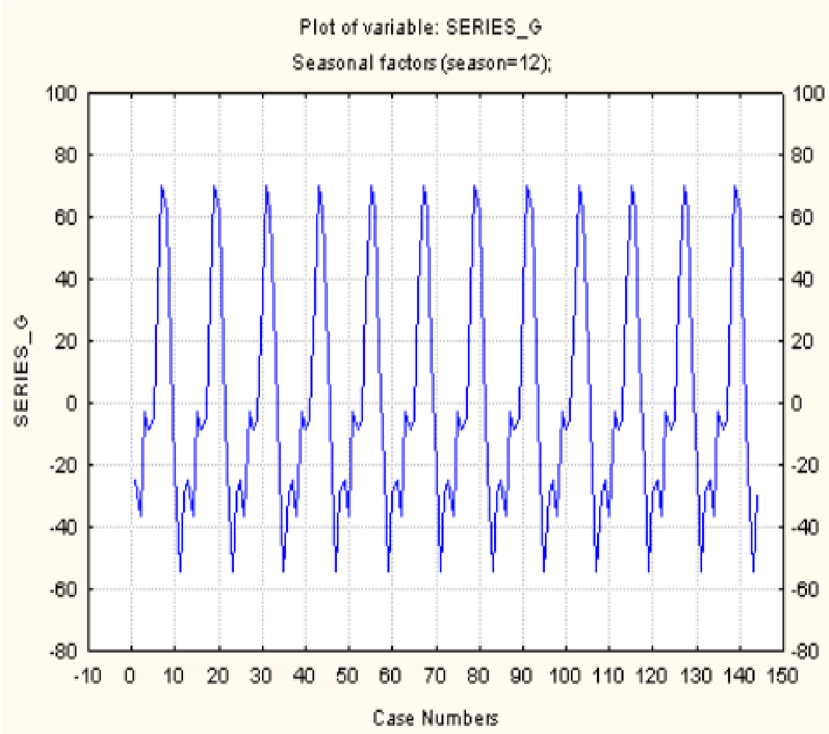
Ковзне середнє (Moving averages) – результат вихідного ряду, перетворений ковзким середнім з лагом 12 (тренд).



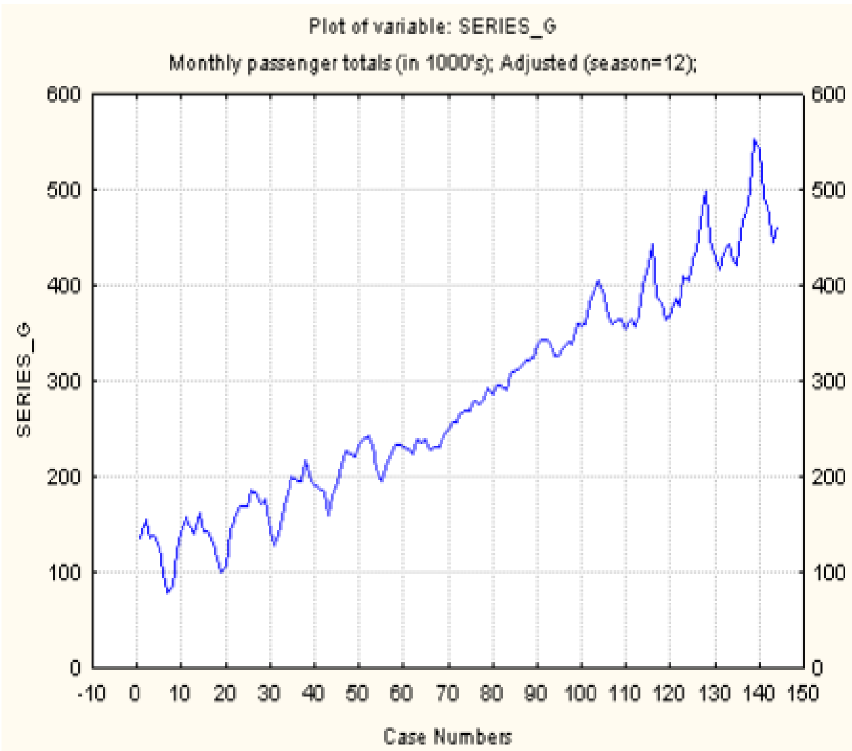
Сезонна складова (Diffrnces) – різниця між вихідним рядом і ковзким середнім.



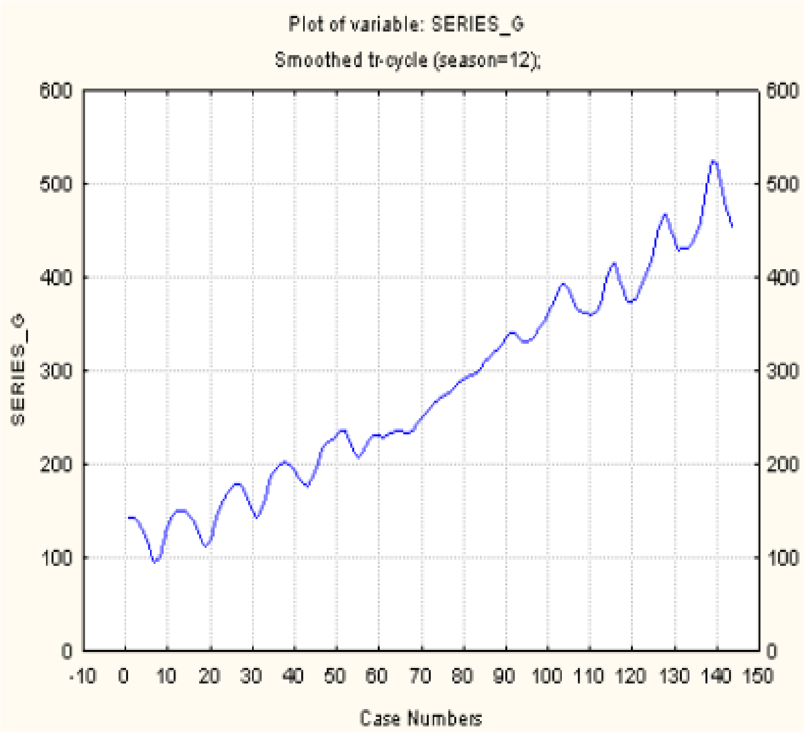
Сезонний фактор (Seasonal Factors) – середнє значення сезонної складової.



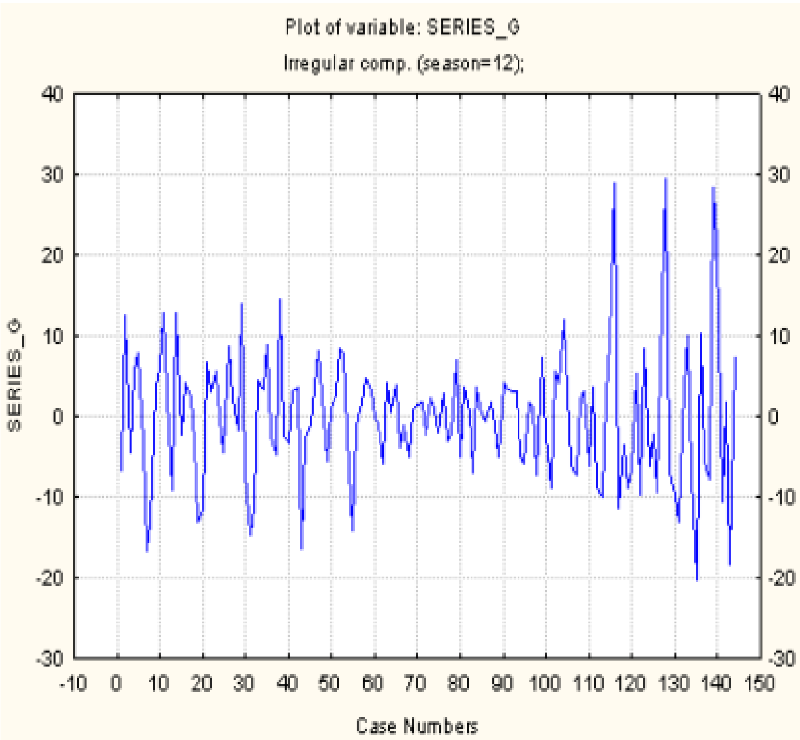
Коригований ряд (Adjusted series) – отриманий шляхом віднімання сезонних компонентів із вихідного ряду.



Згладжуюча послідовність, скоригована за тенденцією (Smoothed Trend-c – Adjusted Series) - отримана експоненційним згладжуванням.



Помилка (Irreg. Compon) – випадковий компонент, часового ряду



4) Побудова автокореляційної функції

На останньому малюнку показана випадкова складова часового ряду. Якщо детермінована складова змодельована правильно, вона повинна бути стаціонарною, включаючи тренд та сезонні компоненти. Перевіряємо стаціонарність шляхом зведення автокореляційної функції до нуля. Для цього перейдемо на вкладку Автокореляція і виберемо ряд Irregular comp. Встановимо число лагів на 50.

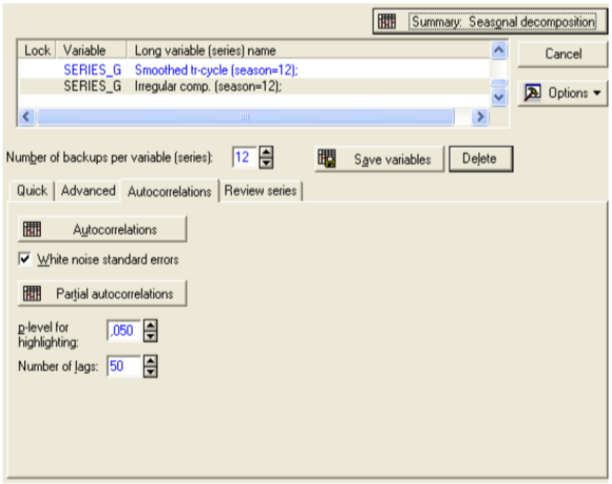


Рисунок 4.4 Панель функцій автокореляції

Натиснувши кнопку автокореляції, ми отримаємо графік:

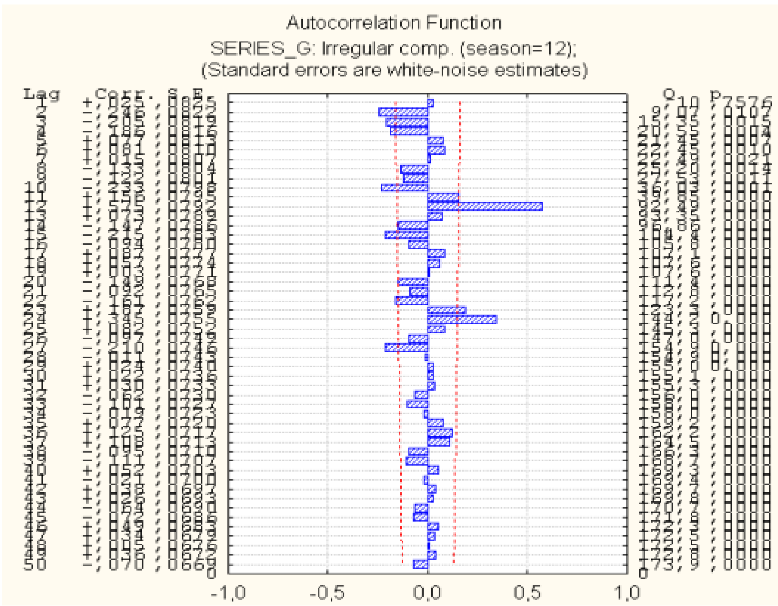


Рисунок 4.5 Автокореляційна функція випадкової складової

Автокореляційна функція випадкової складової прагне до нуля, тому послідовність розкладання ряду проведено вірно.

Використання моделі ARIMA для прогнозування часових рядів

5) На основі аналізу вихідних даних можна зробити висновок: вимірюване значення проводиться раз на місяць із сезонним інтервалом 12 місяців. Отримаємо графік вихідних даних. Для цього натискаємо на кнопку побудова графіка - Plot. Графік вихідних даних наведено на рис. 4.6.

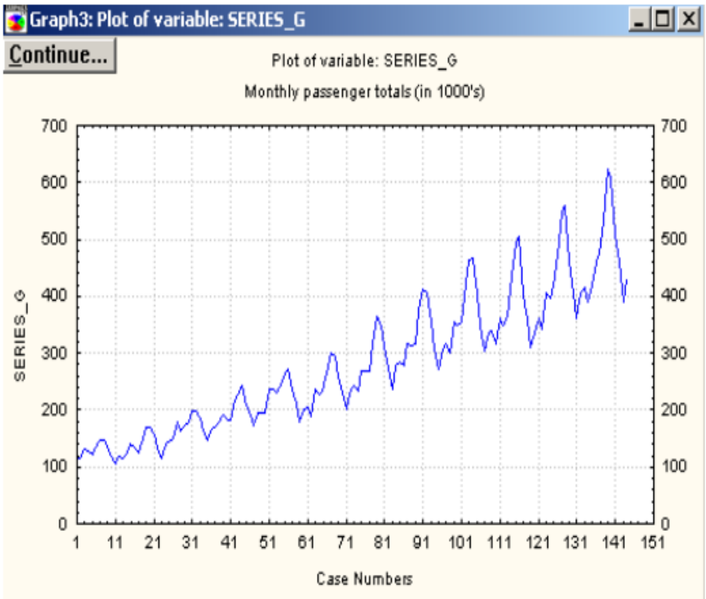


Рисунок 4.6 Вихідні дані

З візуальної точки зору вихідні дані мають зростаючий тренд, дисперсія не постійна, оскільки амплітуда коливань зростає. З рисунка видно, що закон розподілу 4.7 не є нормальним, оскільки він рухається вліво. Часовий ряд не є стаціонарним. Аналіз автокореляційної функції підтвердив це (рис.4.8). Перш ніж застосувати модель ARIMA до часового ряду, його слід зробити стаціонарним. Використаємо інші кнопки перетворення вихідних даних Other transformations & plots.

Спочатку змінимо масштаб на осі X, щоб він дорівнював 12, оскільки закономірності в ряду повторюється через 12 місяців. Обираємо вкладку Properties (min, Edit step), за допомогою правої кнопки миші.

Переходимо на вкладку $x = f(x)$. У вікні, що з’явиться, можна обрати потрібне перетворення ряду.

Можливі наступні перетворення ряду:

Додати константу - Add a constant;

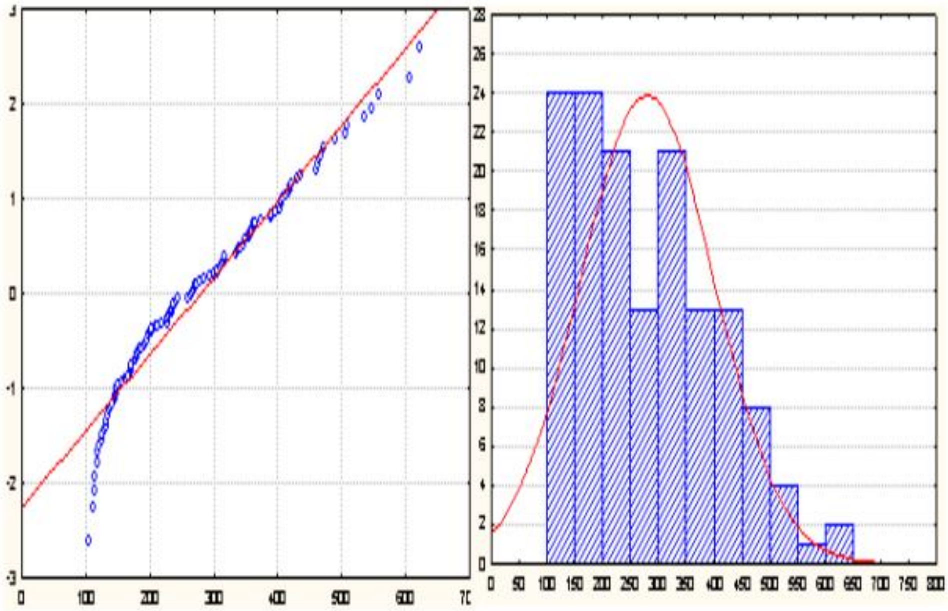


Рисунок 4.7 Графік для перевірки нормальності закону розподілу

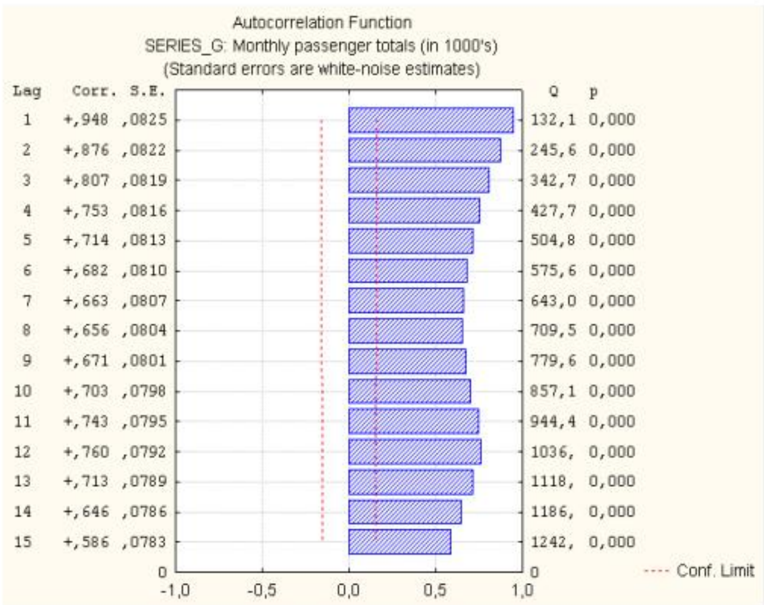


Рисунок 4.8 Графік функції автокореляції вихідного ряду

На вкладці Difference – Integrate можна використати відповідну формулу для обчислення нового значення ряду X, де значення зсуву lag (затримка) вказується в полі lag. Звертаємо увагу, що лаг в даному випадку є додатнім значенням.

6) Щоб усунути тенденцію у часовому ряді, беремо різницю в один лаг. Для цього натискаємо кнопку Other transformations & plots у головному вікні програми, а потім відкриваємо вкладку Difference-Integrate.

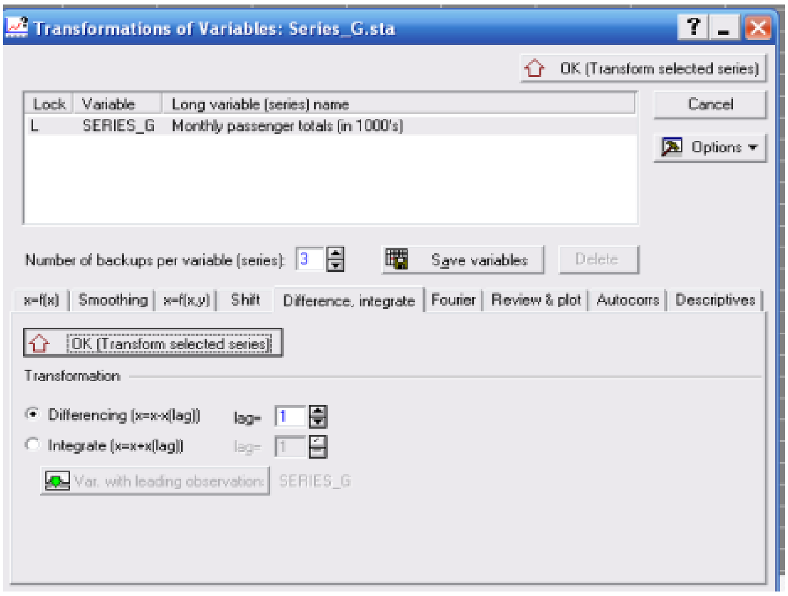


Рисунок 4.9 Взяття різниці в 1 лаг

7) Позбудумося сезонності способом взяття різниць з лагом 12.

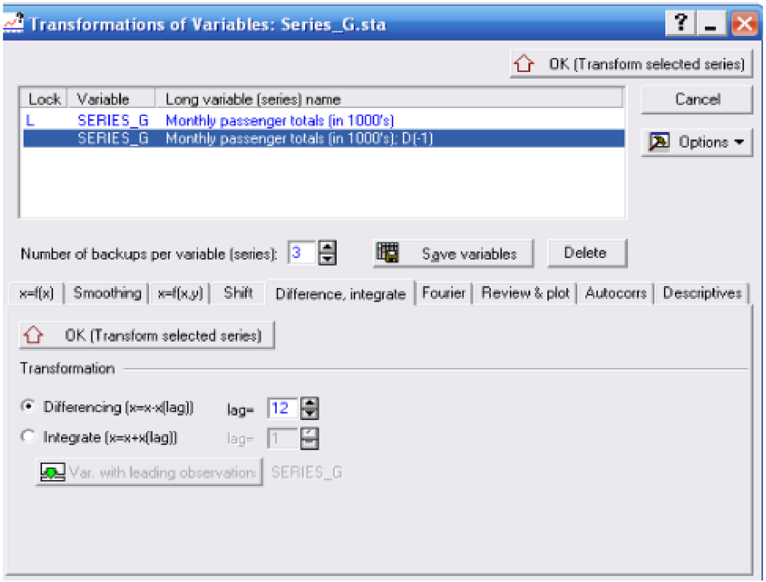


Рисунок 4.10 Взяття різниці в 12 лаг

8) Щоб усунути дисперсію, ми виконуємо логарифмічні операції над часовим рядом:

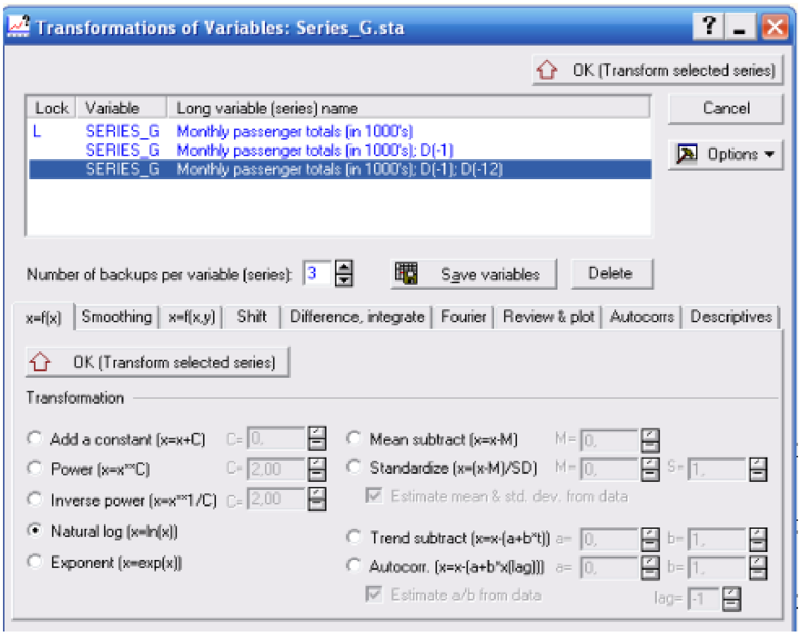


Рисунок 4.11 Логарифмування ряду

Отриманий часовий ряд має такий вигляд (рис. 4.12):

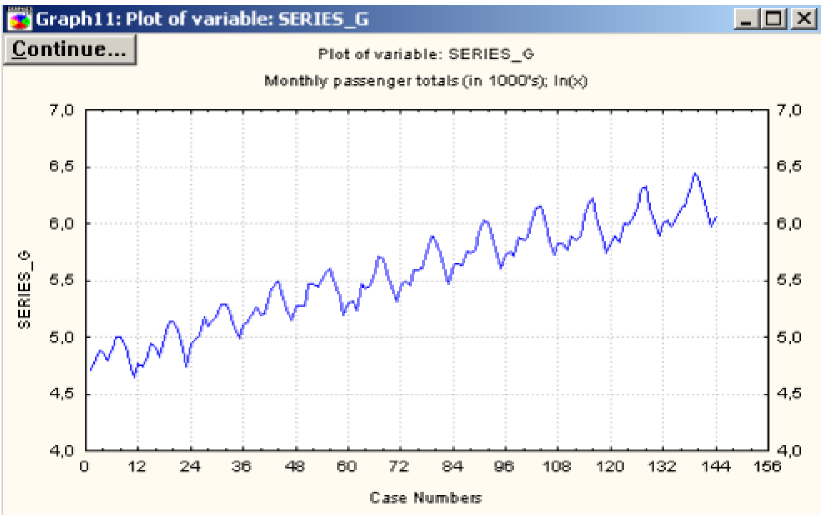


Рисунок 4.12 Прологарифмований часовий ряд

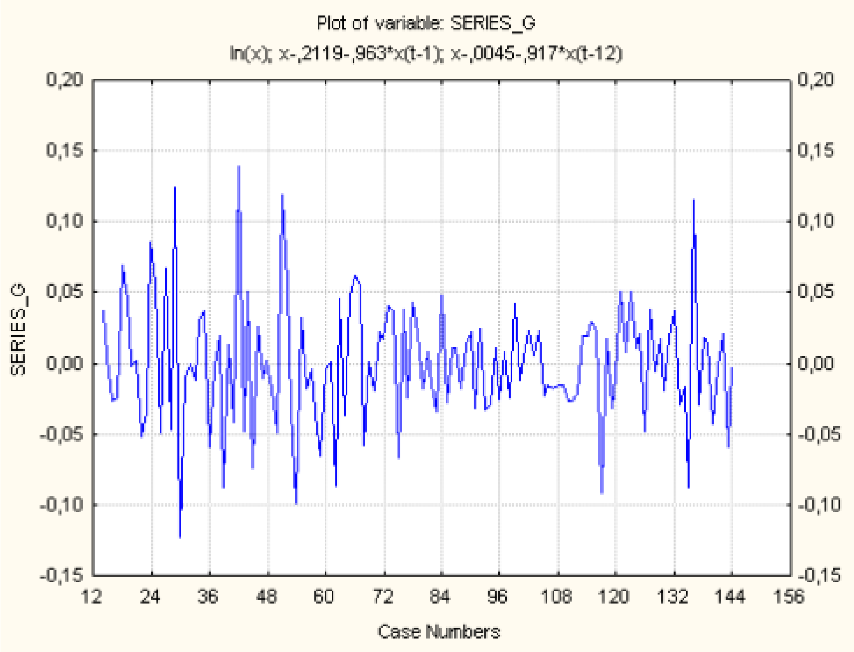


Рисунок 4.13 Послідовність, отримана шляхом віднімання трендів та сезонних компонентів

Ідентифікація моделі ARIMA.

Виберемо структуру моделі ARIMA на основі графіків автокореляційної функції та часткової автокореляційної функції.

На основі ACF та $PACF$ можна зробити висновок, що слід використовувати модель із сезонною складовою ARIMA (1,1,1)(1,1,1).

Введемо заданий порядок моделі у відповідні комірки програми. На панелі змінної трансформації (Transform variable) вказуємо виконані раніше операції – логарифмування і різниці з лагом 1 та 12.

Виберемо метод оцінки в нижньому лівому куті вікна під назвою: Estimation of Maximum Likelihood – оцінювання методом максимальної правдоподібності.

Для реалізації цього методу в системі передбачено дві розрахункові програми: Approximate – наближена, Exact – точна.

Наприклад, виберемо метод Exact – точний. У вікні Variables, що відкриється не забуваємо виділити перший рядок – вихідний часовий ряд.

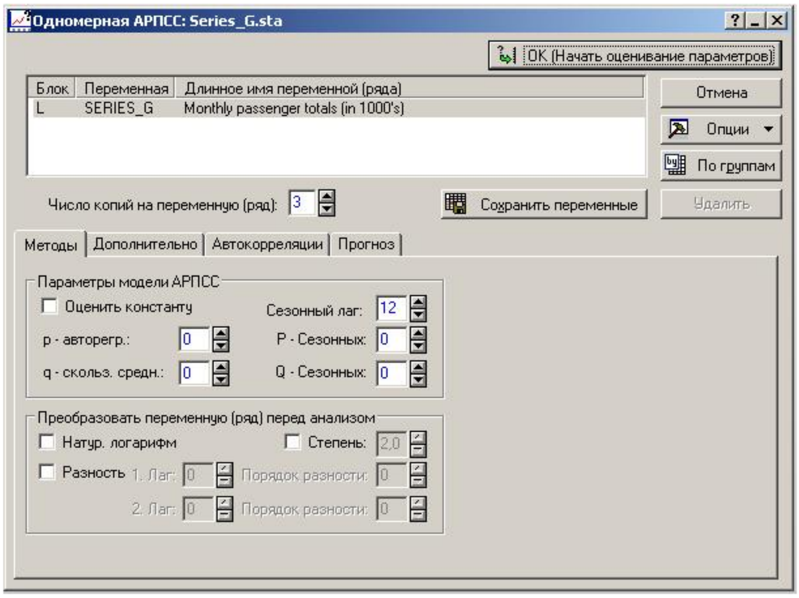


Рисунок 4.14 Вікно розпізнавання моделі ARIMA

- p – Autoregressive – авторегресивні параметри (регулярний);
- P – Seasonal – сезонний параметр авторегресії;
- q – Moving average – параметр ковзкого середнього (регулярний);
- Q – Seasonal – сезонний параметр ковзкого середнього.

9) Знаходження параметрів моделі ARIMA.

Натискаємо кнопку OK (Begin parameter estimation), щоб отримати модель параметра. Вона має вигляд:

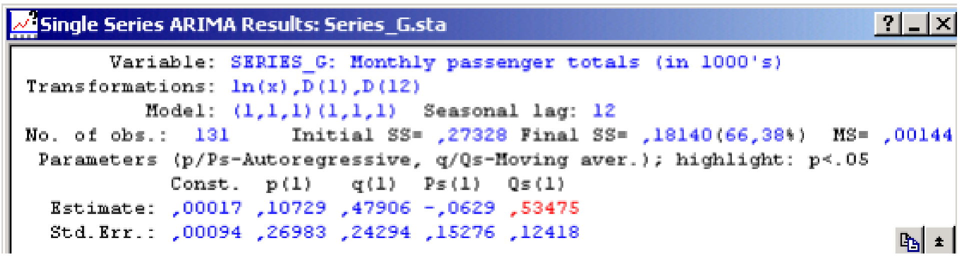


Рисунок 4.15 Параметри моделі ARIMA

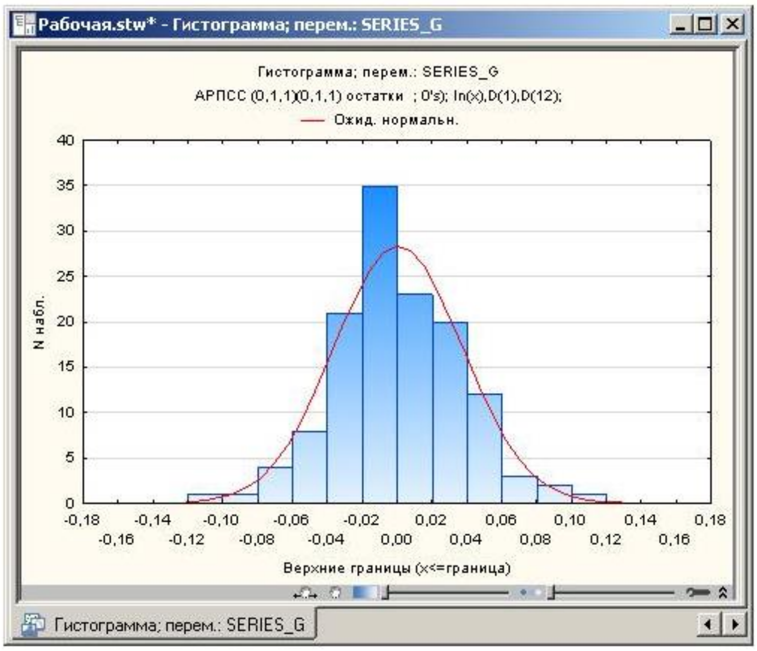
Запишемо рівняння моделі:

$$Y_t = 0{,}00017 + 0{,}10729Y_{t-1} - 0{,}47906e_{t-1} - 0{,}0629Y_{t-12} - 0{,}53475e_{t-12} + e_t$$

Розглянемо чисельні оцінки, натиснувши кнопку Parameter estimates – оцінки параметрів. Перша колонка таблиці містить оцінки параметрів, друга колонка – асимптотична стандартна помилка, третя колонка – значення t-критерію, четверта колонка – рівень значущості, п’ята і шоста колонка – відповідно 95% верхній і нижній межі відповідають довірчим інтревалам параметрів моделі.

10) Перевіримо модель на адекватність.

Адекватність моделі оцінюється на основі статистичного аналізу залишків. Якщо залишки є білим шумом, то вони мають нормальний закон розподілу, математичне очікування дорівнює 0, і немає автокореляції. Розглянемо графік автокореляційної функції та гістограму нормального розподілу.



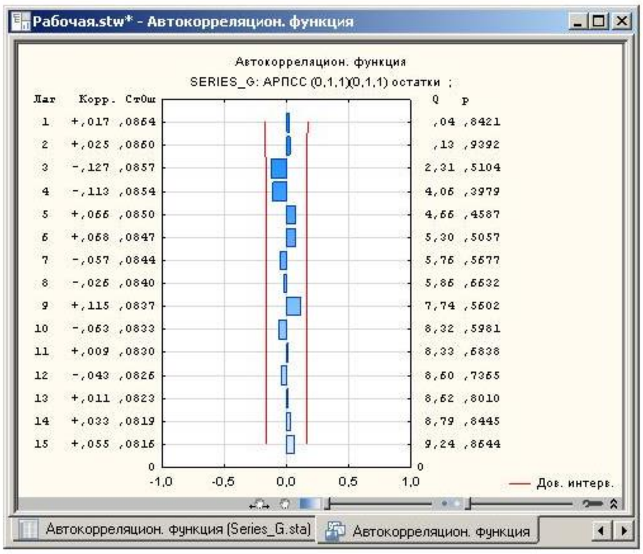


Рисунок 4.16 Графік автокореляційної функції та графік розподілу випадкових компонентів

Оскільки в ряду залишків немає автокореляції, а розподіл значень залишків приблизно повторює криву, задану нормальним законом, то отриману модель можна вважати адекватною.

11) Перейдемо на вкладку Додатково (Advanced), можна знайти прогнозоване значення часових рядів для наступних кількох періодів..

Кількість лагів прогнозу – Number of cases.

Номер випадку, з якого будемо починати прогнозувати – Start at case.

Confidence level – рівень довіри. Тут можна встановити рівень достовірності, щоб виміряти надійність прогнозованих результатів. Зазвичай, це 0,9 або 0,95.

Натискаємо кнопку Plot series & forecasts, і отримаємо графічне представлення прогнозованого ряду.

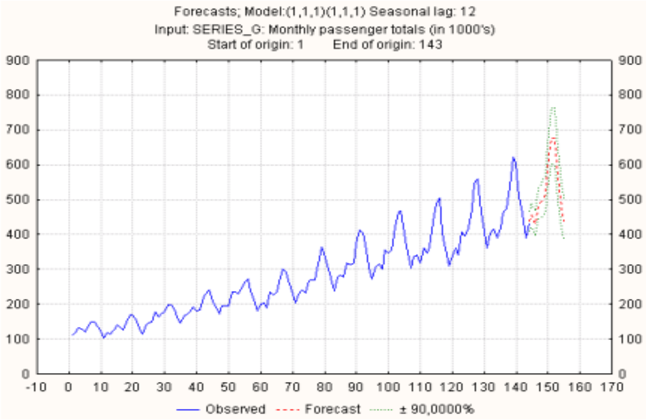


Рисунок 4.17 Графічне представлення прогнозних значень ряду

Приклад 2. Приклад проводився за допомогою статистичної програми EViews, яка використовується та орієнтовна в основному на часові ряди економетричного аналізу. На рисунку 4.18 наведено дані прибутків деякої компанії з виготовлення товарів для дому з деревини. Дані взяті за два роки помісячно. Для того щоб знайти модель часового ряду, в першу чергу перевіримо чи стаціонарний наш ряд.

Jan20	1	109
Feb 20	2	150
Mar 20	3	200
Apr 20	4	150
May 20	5	119
Jun 20	6	168
Jul 20	7	250
Aug 20	8	170
Sep 20	9	147
Oct 20	10	198
Nov 20	11	260
Dec 20	12	200

Jan 21	13	147
Feb 21	14	226
Mar 21	15	270
Apr 21	16	225
May 21	17	170
Jun 21	18	250
Jul 21	19	349
Aug 21	20	235
Sep 21	21	155
Oct 21	22	247
Nov 21	23	350
Dec 21	24	248

Рисунок 4.18 Дані прибутків компанії з виготовлення товарів для дому з деревини

1. Побудуємо графік наших початкових даних.

Для цього у вікні Time Series Analysis вибираємо ARIMA & autocorrelation function.

У ділоговому вікні обираємо Review series (рисунок 4.19) ⇒ Review and plot variables ⇒ Review highlighted variable ⇒ Plot.

На рисунку 4.20 побудовано графік початкових даних ряду для попередньої візуальної оцінки даних.

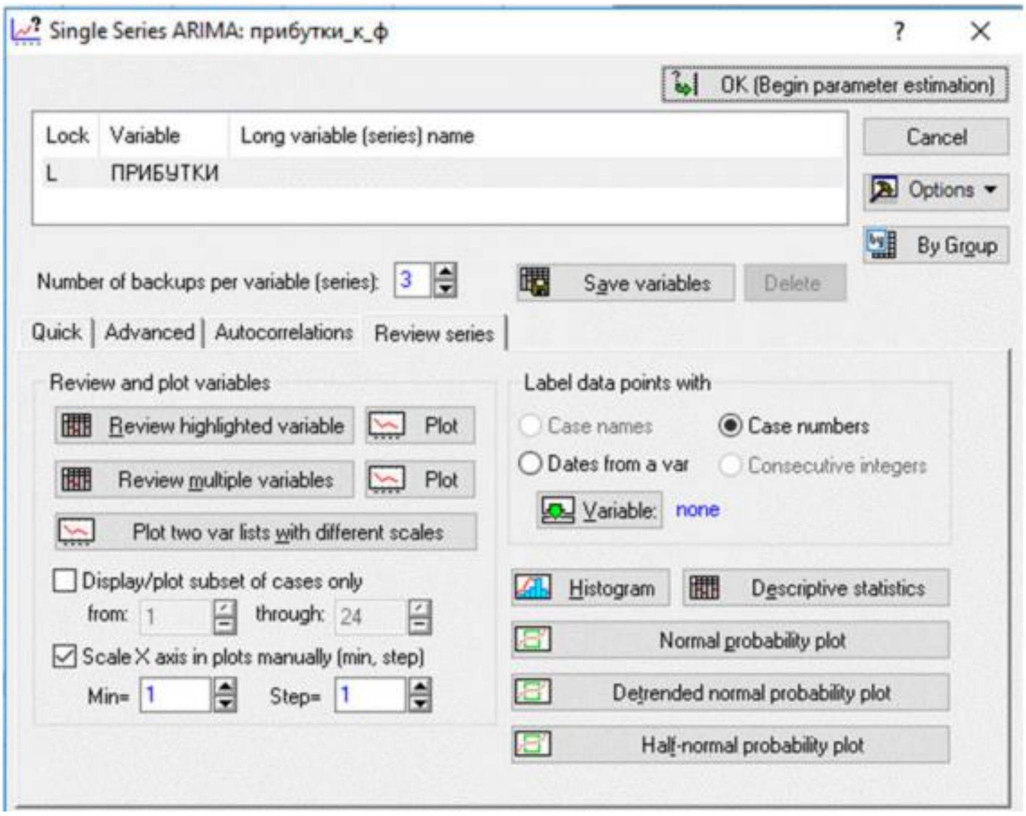


Рисунок 4.19 Вікно Single Series ARIMA

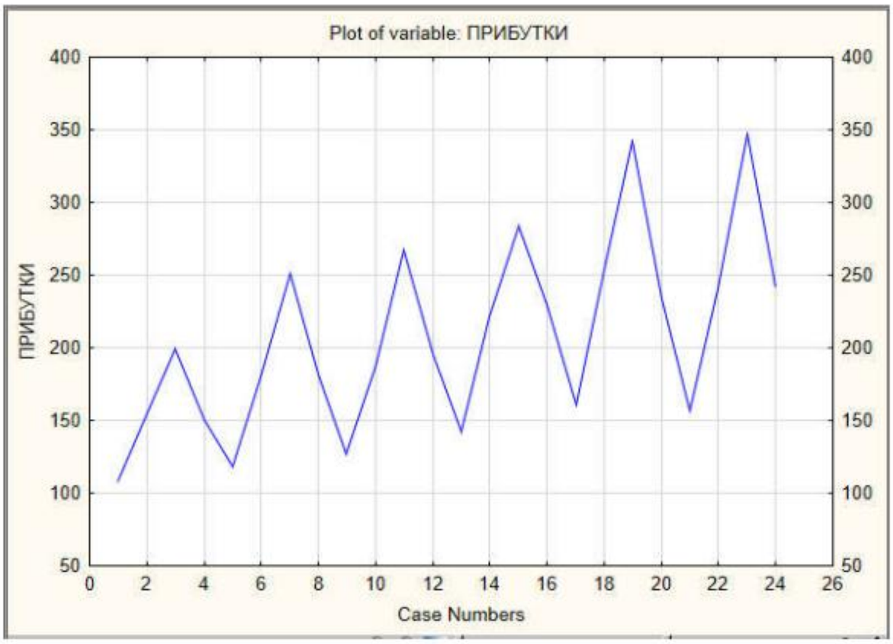


Рисунок 4.20 Графік вихідних даних часового ряду

Як бачимо, даний часовий ряд не є стаціонарним. На рисунку бачимо плавне збільшення значень часового ряду, тобто спостерігається тренд, деяка сезонність.

Щоб зробити ряд даних (рисунок 4.18) стаціонарним, а потім підібрати до нього авторегресійну модель, спочатку потрібно послідовно застосувати кілька перетворень до часового ряду прибутків компанії з виготовлення товарів для дому з деревини за місяцями.

2. Застосуємо логарифмічне перетворення Natural log на вкладці x = f (x) (рисунок 4.21). Графік прологарифмованого ряду (рисунок 4.22) бачимо має меншу дисперсію, а його коливання зменшилися.

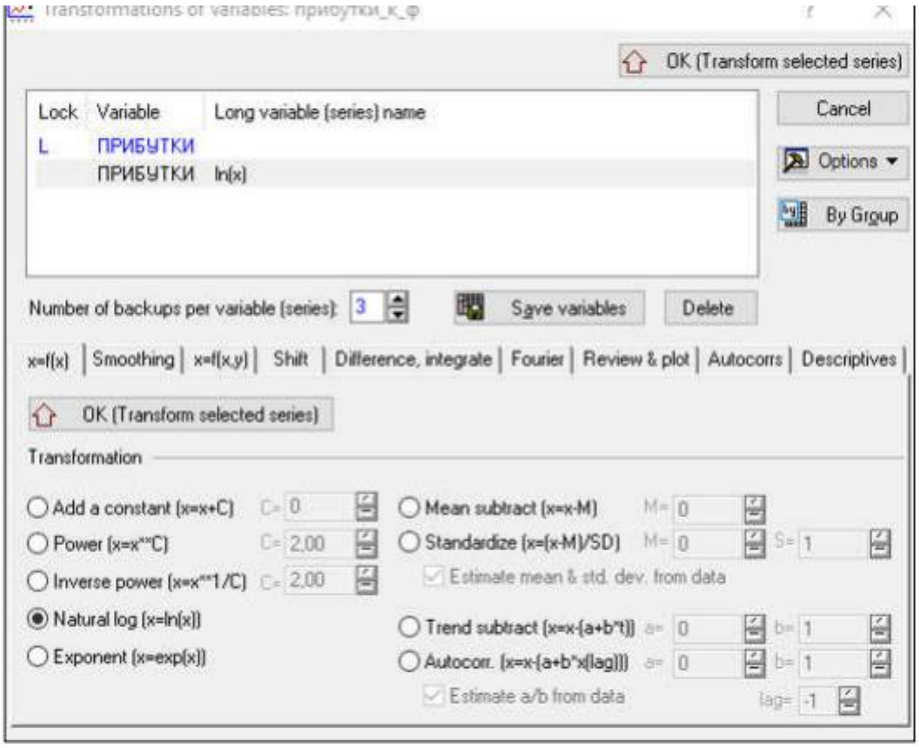


Рисунок 4.21 Логарифмічне перетворення ряду

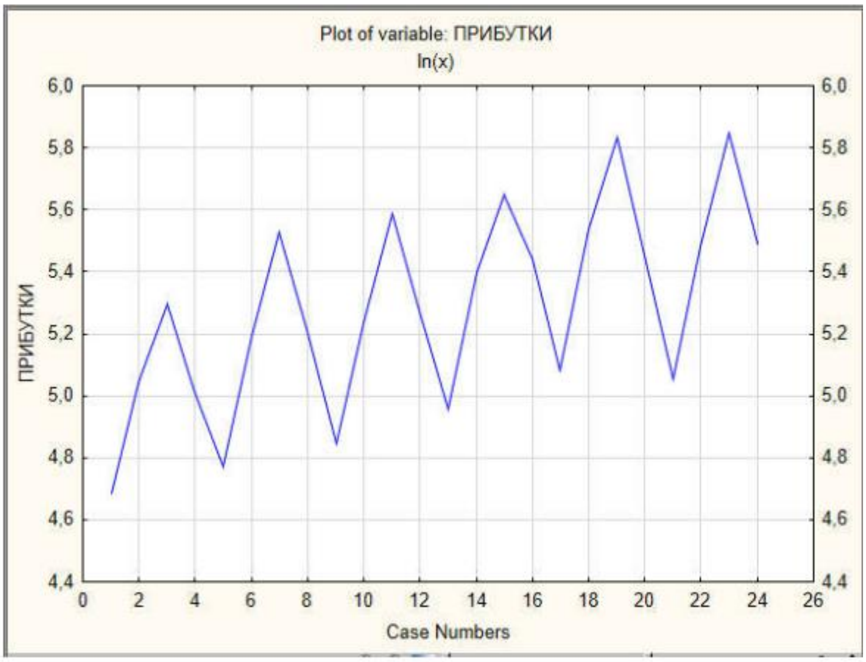


Рисунок 4.22 Графік вихідних даних прологарифмованого ряду

Після зменшення розкиду ідентифікуємо параметри моделі для подальшого їх оцінювання.

З рисунку 4.22 видно, що тренд ряду виражений і вибіркова автокореляційна функція (рисунок 4.23) (вкладка Autocorrs вікна

Transformation of Variables), не має тенденції до згасання, то це вказує на нестаціонарність ряду.

На наступному етапі застосуємо до часового ряду різницеве перетворення (вкладка Differencing, Integrate вікна Transformation of Variables, lag=1).

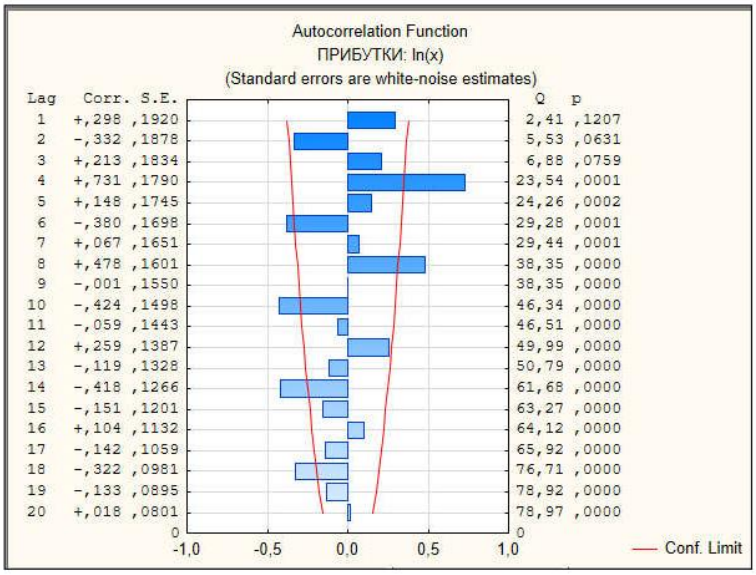


Рисунок 4.23 Графік автокореляційної функції

Графік перетвореного ряду показаний на рисунку 4.24. Бачимо що після змін ряд став стаціонарний.

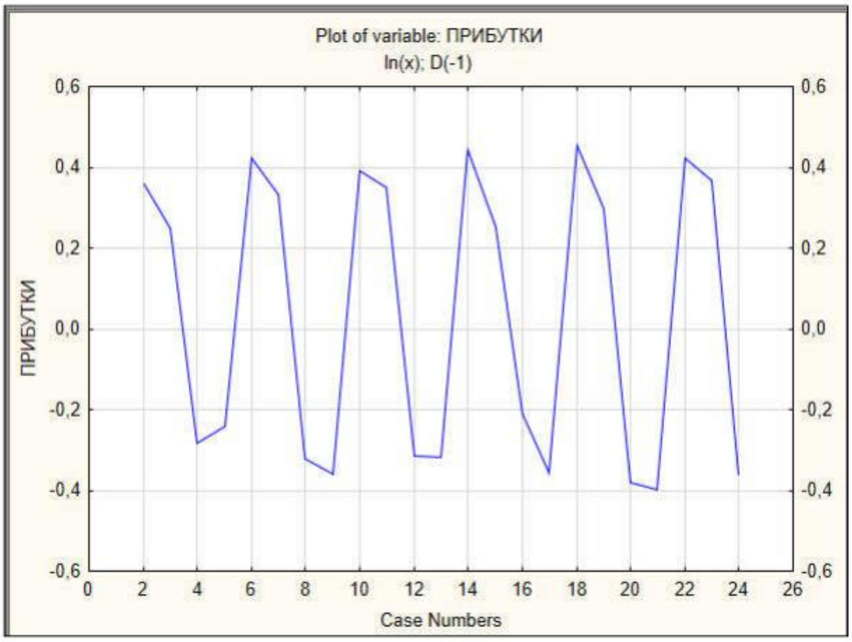


Рисунок 4.24 Графік прибутків після застосування двох перетворень: Ln, Differencing з лагом 1

3. Діалогове вікно, призначене для оцінювання моделі приведено на рисунку 4.25.

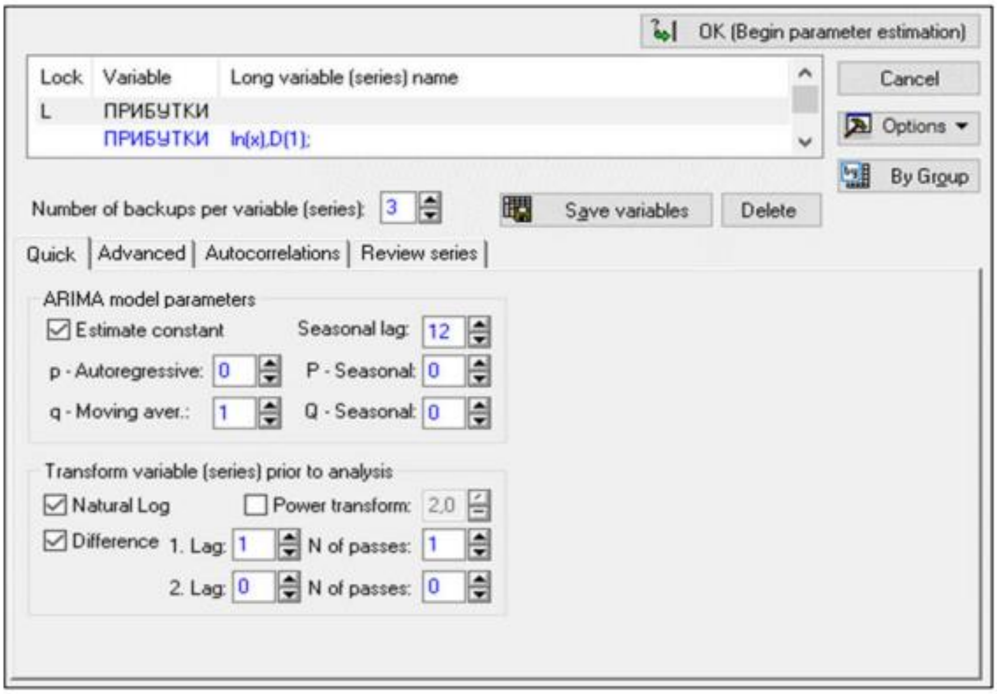


Рисунок 4.25 Діалогове вікно для оцінювання моделі

На рисунках 4.26 – 4.32 приведено аналіз моделі ARIMA(0,1,1).

На рисунку 4.26 – вікно з результатами обчислень. Бачимо, у вікні те що позначено червоним кольором, вказує на те,що коефіцієнти рівняння моделі стійкі, і значущі. Коефіцієнти моделі більше ніж в два рази перевищують свої стандартні помилки. Задамо параметри прогнозу за допомогою вкладки Advanced в полі Forecasting.

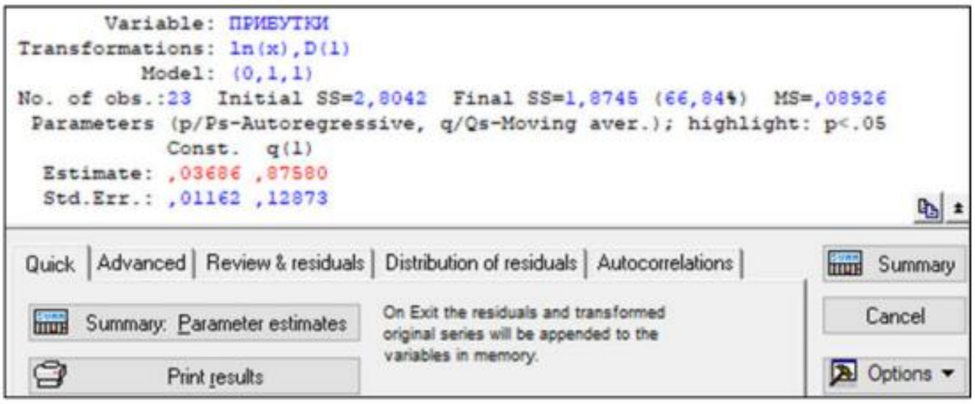


Рисунок 4.26 Вікно з результатами обчислень

Далі проводимо перевірку адекватності моделі.

На рисунку 4.27 приведено значення оцінок, їх стандартні помилки, t-статистики, Р-значення, довірчі інтервали (95%) оцінок. Оцінки коефіцієнтів моделі є статистично значущими, вони всі позначені червоним кольором.

Paramet.	Input: ПРИБУТКИ (прибутки_к_ф) Transformations: ln(x),D(1) Model:(0,1,1) MS Residual= ,08926					
	Param.	Asympt. Std.Err.	Asympt. t(21)	p	Lower 95% Conf	Upper 95% Conf
Constant	0,036860	0,011624	3,171164	0,004601	0,012688	0,061033
q(1)	0,875802	0,128731	6,803371	0,000001	0,608092	1,143512

Рисунок 4.27 Аналіз оцінок параметрів моделі ARIMA(0,1,1)

4. Наступним етапом є перевірка, чи є помилки моделі білим шумом. Для цього перевіримо, чи залишки регресії, подібні на білий шум, тобто чи залишки мають нульову автокореляцію. На рисунку 4.28 зображена корелограма залишків.

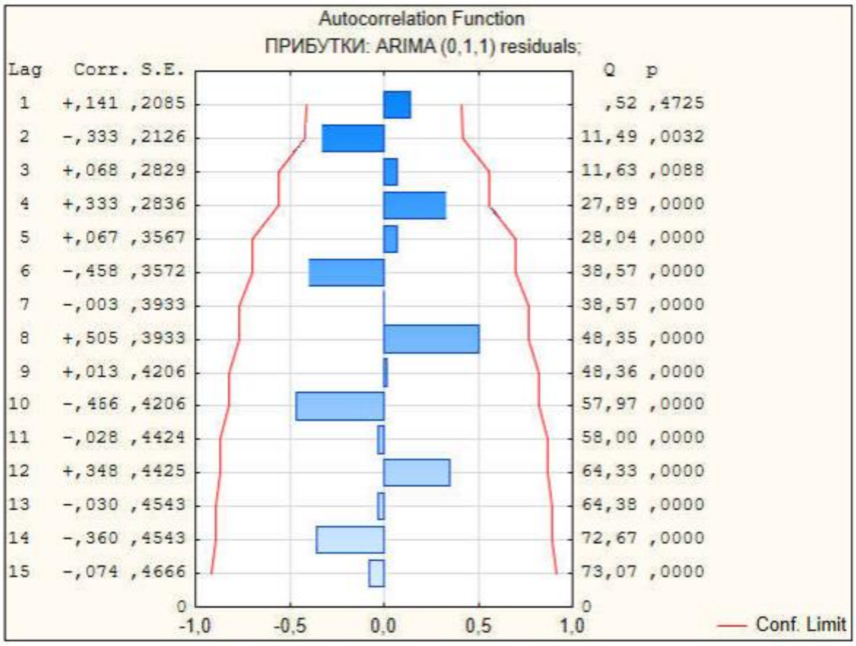


Рисунок 4.28 Автокореляційна функція залишків моделі ARIMA(0,1,1)

З рисунку 4.28 бачимо, що коефіцієнти автокореляції незначущі (не виходять за межі довірчого коридору) і розташовані хаотично, тобто залишки незалежні, подібні на білий шум.

Для перевірки нормальності розподілу залишків моделі використаємо гістограму залишків (рисунок 4.29) і графік нормального розподілу залишків (рисунок 4.30). По результату бачимо, що гістограма майже знаходиться в межах нормальної кривої, тому розподіл близький до нормального. Це підтверджується й з рисунку 4.30 – всі точки групуються близько прямої лінії на графіку.

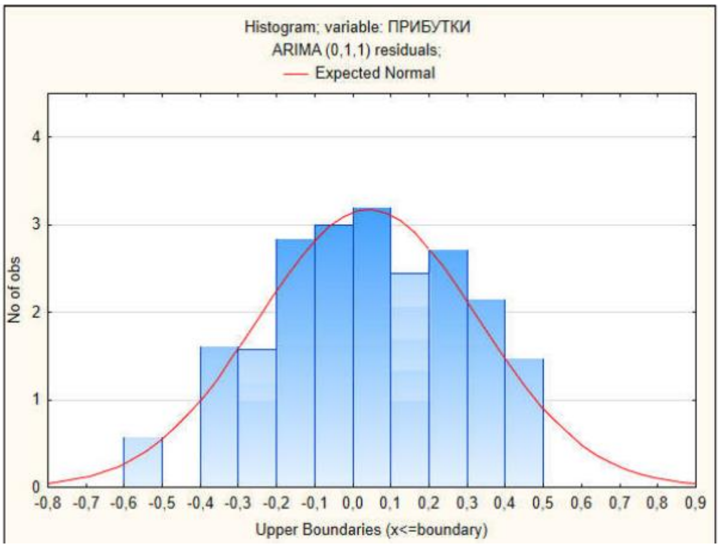


Рисунок 4.28 Гістограма залишків моделі ARIMA(0,1,1)

Графік ряду з доданими прогнозними значеннями прибутків і довірчими інтервалами для них на три місяці приведено на рисунку 4.31, числові значення їх на рисунку 4.32.

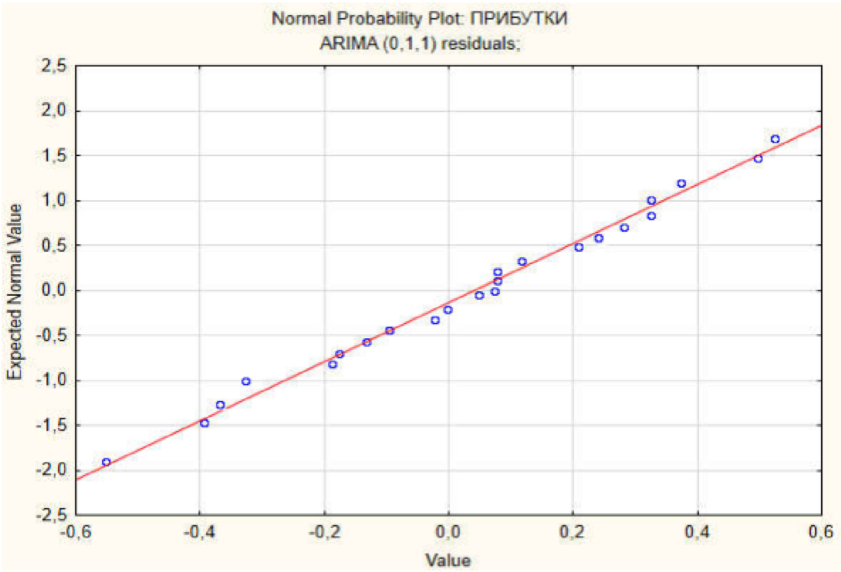


Рисунок 4.30 Графік нормального розподілу залишків моделі ARIMA(0,1,1)

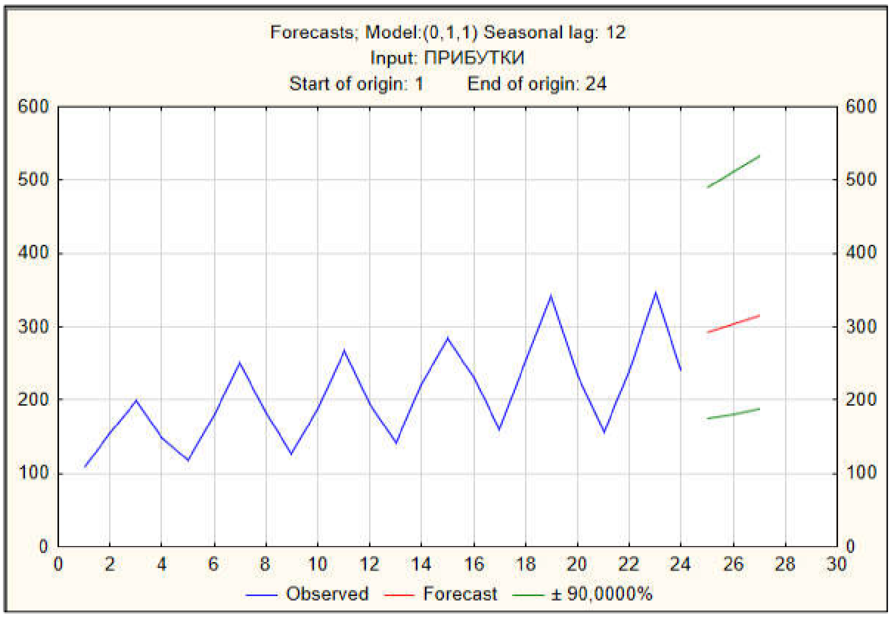


Рисунок 4.31Графік ряду з прогнозними значеннями і довірчими інтервалами

CaseNo.	Forecast	Lower 90,0000%	Upper 90,0000%
	Forecasts; Model:(0,1,1) Seasonal lag: 12 (прибутки_к_ф) Input: ПРИБУТКИ Start of origin: 1 End of origin: 24		
25	293,3820	175,4544	490,5719
26	304,3979	181,3248	511,0064
27	315,8275	187,3971	532,2761

Рисунок 4.32 Прогнозні значення і довірчі інтервали

Отже, в даному розділі і на розглянутих прикладах на основі запропонованих алгоритмів розроблено модуль ідентифікації моделі, модуль оцінювання моделі і перевірки її адекватності, модуль прогнозування за моделлю, що дало можливість покращити прогнозування прибутків для компанії, що нами розглядалась.

Часові ряди в EXCEL

Якщо фіксувати значення якогось процесу через певні проміжки часу, вийдуть елементи часового ряду. Їх мінливість намагаються поділити на закономірну та випадкову складові. Закономірні зміни членів низки, як правило, передбачувані.

Зробимо аналіз часових рядів у Excel.

Приклад: торговельна мережа аналізує дані про продаж товарів магазинами, що у містах із населенням менше 50 000 людей. Період – 2017-2020 рр.

Завдання – виявити основну тенденцію розвитку.

Внесемо дані про реалізацію до таблиці Excel:

Рік	Квартал	Продаж,млн грн.
2017	1	165
	2	253
	3	316
	4	287
2018	1	257
	2	308
	3	376
	4	351
2019	1	410
	2	431
	3	443
	4	389
2020	1	436
	2	459
	3	492
	4	470

На вкладці Дані натискаємо кнопку Аналіз даних. В запропонованому переліку обираємо інструмент Експоненційне згладжування. Експоненційне згладжування – це метод вирівнювання, який зазвичай використовується для динамічного ряду, значення якого сильно коливаються.

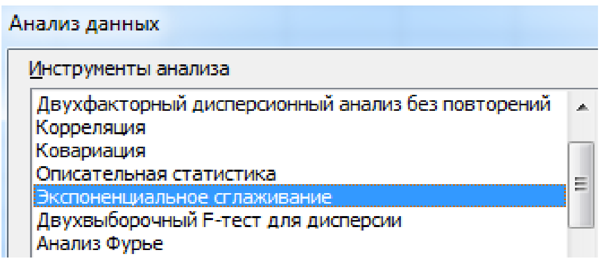


Рисунок 4.33 Інструменти аналізу

Заповнюємо діалогове вікно. Вхідний інтервал – діапазон зі значення продажів. Фактор згасання – коефіцієнт експонентного згладжування (за замовчуванням – 0,3). Вихідний інтервал – посилання на верхню ліву комірку вихідного діапазону. Сюди програма помістить згладжені рівні та розмір визначить самостійно. Ставимо галочки на вкладці Висновок графіка та Стандартні похибки.

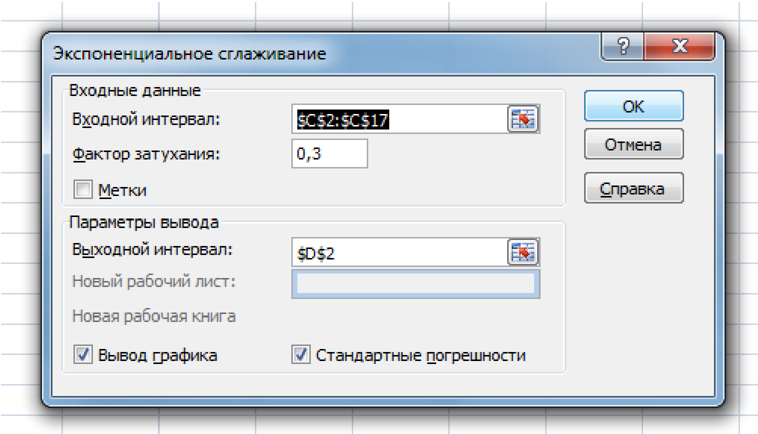


Рисунок 4.34 Експонентне згладжування

Закриваємо діалогове вікно натисканням ОК. Результати аналізу:

	A	B	C	D	E
1	Рік	Квартал	Продаж,млн грн.		
2	2017	1	165	#Н/Д	#Н/Д
3		2	253	165	#Н/Д
4		3	316	226,6	#Н/Д
5		4	287	289,18	#Н/Д
6	2018	1	257	287,654	72,43644
7		2	308	266,1962	54,57954
8		3	376	295,4589	29,9554
9		4	351	351,8377	55,29949
10	2019	1	410	351,2513	52,39318
11		2	431	392,3754	57,55863
12		3	443	419,4126	40,59545
13		4	389	435,9238	42,81602
14	2020	1	436	403,0771	37,63893
15		2	459	426,1231	35,78697
16		3	492	449,1369	38,15165
17		4	470	479,1411	36,52417

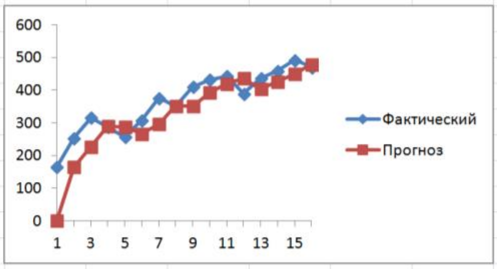


Рисунок 4.20 Результати аналізу

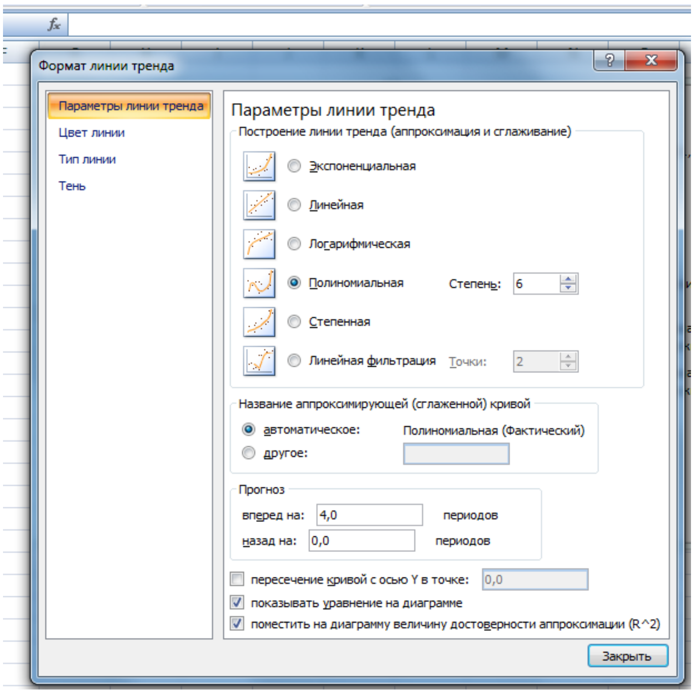
Для розрахунку стандартних похибок Excel використаємо формулу: =КОРІНЬ(СУМКВРАЗН('діапазон фактичних значень'; 'діапазон прогнозних значень')/'розмір вікна згладжування').

Наприклад: =КОРІНЬ(СУМКВРАЗН(C3:C5; D3:D5)/3) = 72,43644.

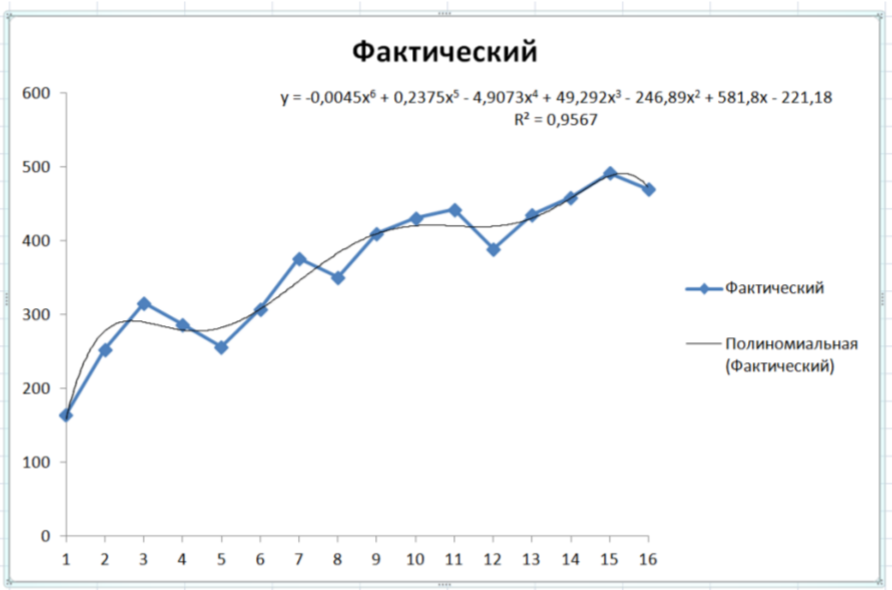
Прогнозування часового ряду в EXCEL

Складемо прогноз продажів, використовуючи дані з попереднього прикладу. На графіку, де відображаються фактичні обсяги реалізації продукції, зобразимо лінію тренду (права кнопка за графіком - Додати лінію тренду).

Налаштовуємо параметри лінії тренду:

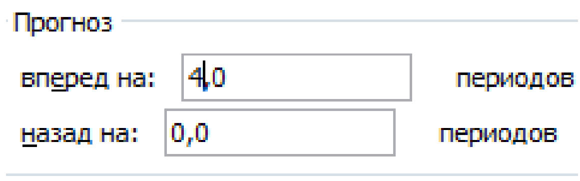


Вибираємо поліноміальний тренд, що максимально скоротитиме помилку прогнозної моделі.

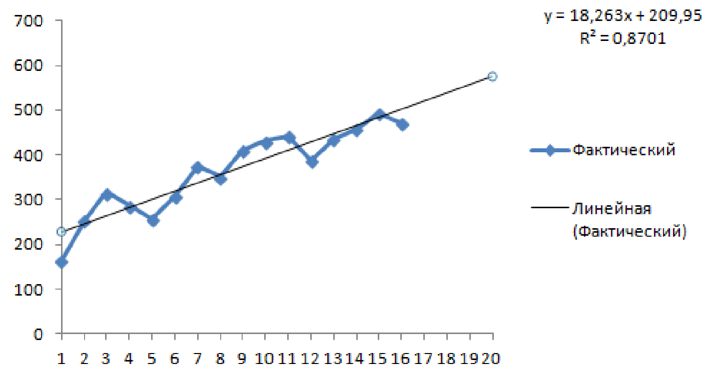


Отримали результат: $R^2 = -0{,}9567$. Це відношення пояснює 95,67% змін обсягів продажів з часом.

Рівняння тренду – це модель формули, яка використовується для обчислення прогнозованого значення. Для прогнозування продажів рекомендується використовувати лінійну лінію тренду. Для перегляду прогнозу на графіку необхідно задати кількість періодів у параметрах.



Отримуємо досить оптимістичний результат:



У нашому прикладі таки експоненційна залежність. Тому при побудові лінійного тренду більше помилок та неточностей.

Висновки

Таким чином, використання моделей часових рядів в наші часи є надзвичайно важливими.

У даній роботі потрібно було проаналізувати та дати характеристики деяким моделям часових рядів, що наддадуть можливість спрогнозувати тенденцію розвитку в будь-якій сфері діяльності.

Використання часових рядів дає змогу передбачити майбутні події на основі даних минулого чи теперішнього часу. Що дуже важливо в наш час, так як в будь-якій сфері торгівлі, бізнесу, навіть наприклад в економіці України, це дає можливість динамічно реагувати на зміни.

У цій роботі описуються різні моделі сезонних коливань, враховуючи моделювання функцій та кореляційно пов'язаних систем показників, а також система показників стійкості рівнів та тренду.

Часові ряди займають велике місце у закономірності їх руху протягом тривалого періоду часу.

Статистика повинна описувати зміни статистичних показників з часом. Як змінюються валовий національний продукт і національний дохід країни з року в рік? Як зростали чи зменшувалися рівень безробіття та рівень заробітної плати? Чи є великі коливання виробництва зерна, чи є тенденція до зростання? На всі поставленні запитання може відповісти лише система статистичних методів, призначена розробці досліджень, зміні часу або дослідженням динаміки.

В першому розділі мною було розглянуто основні поняття, методи часових рядів. Детально розглянула класифікацію та елементи часових рядів.

Також було розглянуто приклади на застосуваня різних методів часових рядів.

Зауважимо, що задачі можна вирішувати або частково автоматизувати за допомогою різних пакетів програмного забезпечення для аналізу, таких як R, Python, MATLAB, STATISTICA, Stata, SAS, SPSS, KXEN, Weka, NCSS, GMDH Shell та UNISTAT.

Для розв’язання прикладів я використовувала програму Statistica та Excel. Программа Excel відноситься до програмного забезпечення загального призначення, тобто для її використання не потрібно мати від студентів спеціальні знання.

У другому розділі розглянули характеристики динаміки. Коригування рівнів часового ряду. А також метод Ірвіна, який використовується для виявлення аномальних значень рівнів часового ряду. Під аномальним рівнем розуміють одиничне значення рівня часового ряду, що відповідає потенційним можливостям досліджуваної економічної системи, і при збереженні рівня ряду суттєво впливає на значення основних характеристик ряду.

Причини, які можуть утворювати аномальні явища, можуть бути помилки пов’язані з технічним порядком, або помилки першого роду, вони підлягають виявлення та усунення.

Третій розділ був присвячений стаціонарним часовим рядам. А саме найбільшу увагу було приділено прогнозуванню часових рядів за допомогою ARIMA моделей. Розглянуто в яких випадках слід використовувати цю модель, також розглянула приклад на застосування цієї моделі, стаціонарність ряду перевіряла за допомогою збіжністю до нуля автокореляційної функції.

Можна зробити висновок, що важливим є оцінка адекватності моделі часових рядів, яка проводиться за допомогою статистичного аналізу залишків. Якщо залишки, що утворилися є білим шумом, то вони мають нормальний закон розподілу, математичне сподівання, яке дорівнює нулю та відсутність автокореляції. Якщо автокореляція в ряду залишків відсутня, а розподіл значень залишків приблизно повторює криву, задану нормальним законом, то отриману модель можна вважати адекватною.

Список використаних джерел

1. Funkhouser H.G. A note on a 10th century graph. Osiris, 1, Bruges, 1936
2. GMDH [The Best Artificial Neural Network Solution in 2021] (n.d.). Raise Forecast Accuracy with Powerful Time Series Analysis Software. Retrieved February 3, 2021, from https://www.gmdhshell.com/time-series-analysis-software
3. IBM Knowledge Center (n.d.). Forecasting with the Time Series Node. Retrieved February 3, 2021, from https://www.ibm.com/support/knowledgecenter/SS3RA7_18. 2.2/modeler_tutorial_ddita/clementine/example_broadband_forecast.html.
4. MathWorks (n.d.). System identification toolbox. Retrieved February 3, 2021, from https://www.mathworks.com/products/sysid.html.
5. NCSS (Forecasting and Time Series Analysis) [modern blog] (n.d.). Retrieved February 3, 2021, from http://engturki.blogspot.com/2007/06/ncss-forecasting-andtime-series.html.
6. NCSS [Statistical software]. (n.d.). NCSS 2021 Data Analysis & Graphics: Overview. Retrieved February 3, 2021, from https://www.ncss.com/software/ncss/.
7. NCSS [Statistical software]. (n.d.). NCSS 2021 Data Analysis & Graphics: Time Series and Forecasting Methods in NCSS. Retrieved February 3, 2021, from https://www.ncss.com/software/ncss/time-series-and-forecasting-in-ncss/.
8. Paul S.P. Cowpertwait Introductory Time Series with R, January 2009. – 86 pages
9. Playfair W. A Letter on our Agricultural Distresses, London, 1821
10. Pro-spo.ru [Про вільне програмне забезпечення та нові інформаційні технології] (n.d.). Forecast Pro – програмне забезпечення для бізнес-прогнозування. Retrieved February 3, 2021, from http://pro-spo.ru/winmat/2285-forecast-pro-.
11. SAP [SAP Help Portal] (n.d.). Kxen.TimeSeries. Retrieved February 3, 2021, from

https://help.sap.com/viewer/e298bfb935ae49999de48dc9c269b90b/3.2/enUS/7d23501473721014833e3e699b0e91070.html.

12.SAS [Analytics Software & Solutions] (n.d.). Прогнозування та аналіз часових рядів. Лекція 1. SAS. (n.d.). Retrieved February 3, 2021, from https://www.sas.com/content/dam/SAS/ru_ru/doc/academic/VMK_MGU/2015/le c7/EM2015_7.pdf.

13.Softline (n.d.). Ефективні інструменти прогнозування продажів: система Sales-Forecast. Retrieved February 3, 2021, from http://softline.ru/news/22529.

14.STATA [Stata is statistical software for data science] (n.d.). Retrieved February 3, 2021, from https://www.stata.com/

15.Statmodels [statistical models, hypothesis tests, and data exploration] (n.d.). Time Series analysis tsa. Retrieved February 3, 2021, from https://www.statsmodels.org/dev/tsa.html.

16.Statmodels [statistical models, hypothesis tests, and data exploration] (n.d.). Time Series Analysis by State Space Methods statespace. Retrieved February 3, 2021, from https://www.statsmodels.org/dev/statespace.html.

17.Statmodels [statistical models, hypothesis tests, and data exploration] (n.d.). Vector Autoregressions tsa.vector_ar. Retrieved February 3, 2021, from https://www.statsmodels.org/dev/vector_ar.html.

18.StatSoft (n.d.). Методи прогнозування. Retrieved February 3, 2021, from http://statsoft.ru/solutions/tasks/forecast/.

19.The University of WAIKATO (n.d.). WEKA. The workbench for machine learning. Retrieved February 3, 2021, from https://www.cs.waikato.ac.nz/ml/weka/.

20. Time Series Analysis James D.H.Hamilton, 1994, – 799 с.

21.UNISTAT Statistical Software (n.d.). Forecasting and Smoothing. Retrieved February 3, 2021, from https://www.unistat.com/guide/time-series-analysisforecasting-and-smoothing/.

22. Yule G. Udny Statistical Papers: Selected by Alan Stuard and M.G. Kendall. Charles Graffin & Co., London, 1971, – 447 pages
23. Zaitun Time Series [Free Time Series Analysis and Forecasting Software] (n.d.). Zaitun Time Series Online Documentation. Retrieved February 3, 2021, from https://www.zaitunsoftware.com/?q=content/zaitun-time-series-online-documentation.
24. Анализ временных рядов и прогнозирование, Афанасьев В.Н., Юзбашев М.М., 2001, – 228 с.
25. Анализ временных рядов и пронозирование Н.Н.Валеев, А.В.Аксянова, Г.А.Гадельшина, 2010, –160 с.
26. Анализ временных рядов, 2 издание, О.А.Подкорытова,М.В.Соколов, 2017, – 267 с.
27. Бокс Дж., Дженкинс Г. Анализ временных рядов. Прогноз и управление. – М.: Мир, 1974, – 406 с.
28. Временные ряды М.Кэндэл, 1981, –199с.
29. Джевонс С. Основи науки. Трактат про логіку та науковий метод, 1881, – 713с.
30. Економетричні моделі динаміки, тема 5, – 16с.
31. Економічне прогнозування: вступ, К.Холден, Д.А.Піл, Дж.Л.Томпсон, Київ – 1996, – 216 с.
32. ІНТУІТ [Національний відкритий університет]. (n.d.). Лекція 27: Інструмент KXEN. Retrieved February 3, 2021, from https://www.intuit.ru/studies/courses/6/6/lecture/210?page=3.
33. Лук'яненко І.Г., Городніченко Ю.О. Сучасні економетричні методи у фінансах. Навчальний посібник. – К.: Літера ЛТД, 2002. – 352с.
34. Матеріали студенту (n.d.). Аналіз часових рядів засобами пакета SAS. Retrieved February 3, 2021, from http://studik.net/analiz-vremennyx-ryadovsredstvami-paketa-sas/.
35. Оліскевич М.О. Основи економетрії часових рядів. Навчальний посібник. – Видавничий центр ЛНУ ім. І. Франка, 2008. – 321с.

36. Присенко Г.В., Равікович Є.І. Прогнозування соціально-економічних процесів. – К.: КНЕУ, 2005. – 362 с.
37. Прогнозування та аналіз часових рядів. Методичні вказівки до практичних занять та самостійної роботи студентів спеціальності 051 «Економіка» освітня програма «Економічна кібернетика», «Економічна аналітика» / Укл.: Юрченко М. Є. – Чернігів: ЧНТУ, 2018. – 88 с.
38. Розробка програмного комплексу аналізу та ансамблевого прогнозування рядів динаміки, Долгіх Анастасія Олегівна, Дніпро –2021, – 197 с.
39. Учебник по программе STATISTICA (n.d.). ГЛАВА 1. Коротка екскурсія по системі STATISTICA. Retrieved from http://www.hr-portal.ru/statistica/gl1/gl1.php.
40. Часові ряди Ставицький А.В., 2003 Електронний підручник
41. Черняк О. І., Комашко О. В., Ставицький А. В., Баженова О. В. Економетрика, 2009. – 395 с.
42. Марчук В. И., Токарева С. В. Способы обнаружения аномальных значений при анализе нестационарных случайных процессов: монография. Шахты: Изд. ЮРГУЭС, 2009. – 209 с.
43. Irvin J. O. On a criterion for the rejection of outlying observation // Biometrika. 1925. V. 17. P. 238–250.
44. Лук'яненко І. Г., Жук В. М. Аналіз часових рядів. Частина перша : Побудова ARIMA, ARCH/GARCH моделей з використанням пакета E.Views 6.0. Практичний посібник для роботи в комп'ютерному класі / І. Г. Лук'яненко, В. М. Жук. – К. : НаУКМА ; Аграр Медіа Груп, 2013. – 187 с.
45. Методичні вказівки щодо виконання лабораторних робіт з навчальної дисципліни «Інтелектуальний аналіз даних» для студентів денної форми навчання зі спеціальності 122 – «Комп'ютерні науки» за освітньо-професійною програмою «Комп'ютерні науки» освітнього ступеня «Бакалавр» (частина 2), проф. І.В. Шевченко; старш. Викл. Т. В. Горлова, 2020. – 59 с.

46. Арженовський С. В., Молчанов І. Н. Статистичні методи прогнозування. – Ростов, – 2001. – 74 с.