

КАПИТЕЛ 4 / CHAPTER 4⁴THEORETICAL FOUNDATIONS OF HOUR SERIES ANALYSIS AND ITS
USE FOR FORECASTING ELECTRICITY CONSUMPTION AND
ANALYZING ECONOMIC GROWTH DEPENDENCIES

DOI: 10.30890/2709-2313.2024-31-00-014

Вступ

Прогнозування споживання електроенергії та аналіз його залежності від факторів економічного зростання відіграють ключову роль у енергетичному плануванні, розподілі ресурсів та веденні політики. Точні прогнози попиту на електроенергію дозволяють ефективно розподіляти ресурси, оптимально планувати інфраструктуру та розробляти стійкі енергетичні стратегії. Більш за те, розуміння факторів, які впливають на споживання електроенергії, та їхній зв'язок з економічним зростанням може дати цінну інформацію для політиків, постачальників електроенергії та дослідників.

В дослідженні використана комбінація методів аналізу часових рядів і статистичного аналізу з використанням програмного забезпечення JMP для досягнення цілей дослідження. Обрана методологія забезпечує комплексну основу для прогнозування споживання електроенергії та аналізу його залежності від факторів економічного зростання. Нижче наведено огляд використаної методології:

1. Аналіз часових рядів. Аналіз часових рядів є фундаментальним підходом для вивчення та прогнозування часових даних. Це дозволяє досліджувати закономірності, тенденції та сезонність у наборі даних. У цьому дипломі методи аналізу часових рядів будуть застосовані для аналізу історичних даних споживання електроенергії та визначення базових закономірностей і структур. Вибрані методи аналізу часових рядів можуть включати попередню обробку даних, декомпозицію, вибір моделі та її оцінку.

2. Статистичний аналіз за допомогою JMP: JMP — це потужний програмний пакет, який пропонує широкі інструменти та можливості для аналізу даних,

⁴Authors: Skakalina Olena Viktorivna, Isai Dmytro Anatoliyovych



візуалізації та моделювання. У цьому дипломі JMP буде використано для проведення статистичного аналізу та регресійного моделювання для вивчення залежності споживання електроенергії від факторів економічного зростання. Програмне забезпечення забезпечує зручний інтерфейс, який полегшує дослідження даних, побудову моделі та інтерпретацію результатів.

Методологія ґрунтуватиметься на систематичному процесі, що включає такі кроки:

- Збір даних: перший крок передбачає збір відповідних даних про споживання електроенергії та фактори економічного зростання. Ці дані можна отримати з авторитетних джерел, таких як державні установи, органи регулювання енергетики або міжнародні організації. Вибору змінних даних буде приділено ретельну увагу, забезпечуючи їх релевантність і надійність.

- Попередня обробка даних: після збору даних вони пройдуть попередню обробку, щоб забезпечити їх якість і придатність для аналізу. Це може включати вирішення відсутніх значень, обробку викидів і перетворення даних, якщо необхідно. Мета полягає в тому, щоб отримати чистий і добре підготовлений набір даних для подальшого аналізу.

- Аналіз часових рядів: зібрані та попередньо оброблені дані будуть піддані методам аналізу часових рядів. Це може включати дослідницький аналіз даних для розуміння закономірностей і характеристик даних про споживання електроенергії, а також декомпозицію для визначення тенденції, сезонності та залишкових компонентів. Різні моделі прогнозування часових рядів, такі як ARIMA, експоненціальне згладжування або моделі простору станів, будуть застосовані для точного прогнозування майбутніх моделей споживання електроенергії.

- Статистичний аналіз за допомогою JMP: Паралельно з аналізом часових рядів, зв'язок між споживанням електроенергії та факторами економічного зростання буде досліджено за допомогою статистичного аналізу за допомогою програмного забезпечення JMP. Це передбачає імпортування набору даних у JMP, проведення дослідницького аналізу даних для розуміння розподілу та



кореляції змінних і виконання регресійного аналізу для оцінки впливу факторів економічного зростання на споживання електроенергії.

- Оцінка та інтерпретація моделі: створені моделі прогнозування часових рядів і регресійні моделі будуть оцінені за допомогою відповідних показників і критеріїв. Буде оцінено відповідність, точність і надійність моделей, щоб переконатися в їх дійсності. Результати будуть інтерпретовані та обговорені в контексті цілей дослідження та дослідницьких питань.

Поєднуючи методи аналізу часових рядів із статистичним аналізом за допомогою JMR, ця методологія забезпечує надійну основу для прогнозування споживання електроенергії та дослідження його залежності від факторів економічного зростання. Він пропонує комплексний і системний підхід для отримання цінних ідей і внеску в сферу енергетичного планування та формування політики.

4.1. Ознайомлення з даними часових рядів та їх характеристика.

Аналіз часових рядів – це потужна аналітична структура, яка зосереджена на вивченні точок даних, зібраних протягом певного часу. Він надає цінну інформацію про базові моделі, тенденції та залежності в часових даних, що робить його важливим інструментом для прогнозування споживання електроенергії. У цьому розділі розглядаються теоретичні основи аналізу часових рядів і його застосування для точного прогнозування моделей споживання електроенергії.

1. Компоненти часових рядів: Дані часових рядів зазвичай складаються з чотирьох ключових компонентів: тенденції, сезонності, циклічних коливань і випадкового шуму. Розуміння цих компонентів має вирішальне значення для декомпозиції даних і виявлення базових закономірностей. Компонент тенденції представляє довгострокову поведінку або напрямок даних, тоді як сезонність стосується повторюваних моделей, які виникають протягом коротших періодів



часу, наприклад щоденних, тижневих або річних циклів. Циклічні коливання охоплюють нерегулярні, які не є такими регулярними, як сезонність, часто під впливом економічних або ділових циклів. Нарешті, випадковий шум пояснює непередбачувані та залишкові флуктуації, присутні в даних.

2. Стаціонарність і тести на стаціонарність: Стаціонарність є важливим припущенням в аналізі часових рядів. Стаціонарний часовий ряд демонструє постійне середнє значення, дисперсію та автоковаріацію з часом, що полегшує моделювання та точне прогнозування. Різноманітні статистичні тести, такі як розширений тест Дікі-Фуллера (ADF) і тест Квятковського-Філіпса-Шмідта-Шина (KPSS), можна використовувати для оцінки стаціонарності часового ряду. Якщо ряд виявляється нестаціонарним, для досягнення стаціонарності можна застосувати відповідні перетворення або методи розрізнення.[1]

Формально часовий ряд $\{x_t\}$ вважається стаціонарним, якщо він задовольняє такі умови:

- Постійне середнє значення (μ): Середнє значення ряду залишається постійним протягом часу. $E(x_t) = \mu$, де $E()$ позначає оператор очікуваного значення.

- Постійна дисперсія (σ^2): Дисперсія ряду залишається постійною в часі. $Var(x_t) = \sigma^2$,

- Постійна автоковаріація ($\gamma(k)$): Функція автоковаріації, яка вимірює лінійну залежність між спостереженнями з різними часовими лагами, залишається постійною протягом часу. Для стаціонарного часового ряду автоковаріація залежить лише від часового лагу k а не на конкретний час t . $\gamma(k) = Cov(x_t, x_{t+k}) = \gamma(-k)$ де $Cov()$ позначає оператор коваріації.

Часовий ряд, який задовольняє ці умови, називається слабко стаціонарним. Крім того, сильніша форма стаціонарності, яка називається суворою стаціонарністю, також вимагає, щоб спільний розподіл будь-якого набору n спостережень із ряду є інваріантним до зрушень у часі.

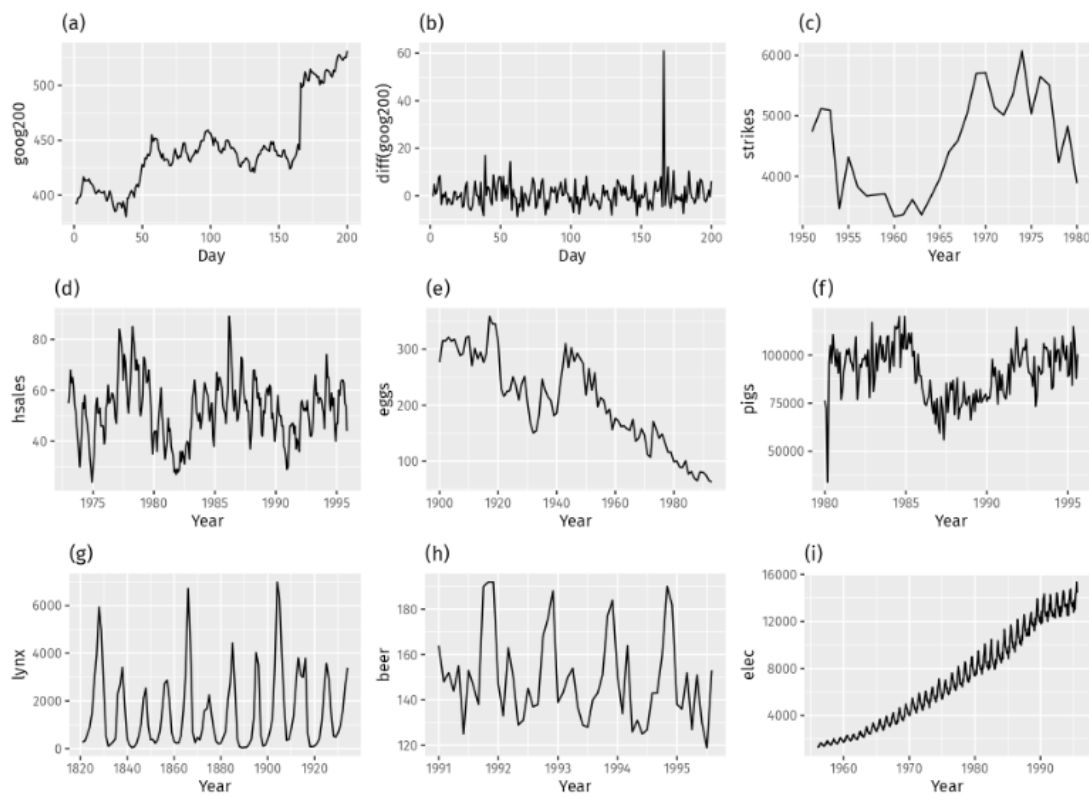


Рисунок 1 – Часові ряди (b) та (g) стаціонарні, в той час як всі інші мають сезонність або тренд

3. В аналізі часових рядів функція автокореляції (ACF) і функція часткової автокореляції (PACF) є важливими інструментами для розуміння структури залежності в межах даних часового ряду. ACF вимірює кореляцію між спостереженнями з різними часовими лагами, а PACF фіксує кореляцію між спостереженнями з різними лагами після усунення ефектів проміжних спостережень.

Функція автокореляції (ACF) із запізненням k , позначена як ρ_k , визначається як:

$$\rho_k = \frac{\gamma_k}{\gamma_0} \tag{1}$$

де γ_k представляє автоковаріацію з відставанням k і γ_0 є автоковаріацією на лагі 0. АКФ дає уявлення про стійкість ряду. Наприклад, якщо ρ_k є високим для



малого k , це свідчить про сильну автокореляцію та повільне спадання кореляції зі збільшенням затримки, що вказує на нестационарний ряд.

- Інтерпретація: висока автокореляція з відставанням k означає, що значення k періодів один від одного подібні. ACF допомагає ідентифікувати такі моделі, як сезонність і тенденції в даних. Наприклад, значний сплеск місячних даних із затримкою 12 може свідчити про річну сезонність.

- Використання: ACF необхідний для визначення порядку процесів ковзного середнього (MA) у моделях ARIMA (авторегресійне інтегроване ковзне середнє).

Часткова автокореляційна функція (PACF) із затримкою k , позначена як ϕ_{kk} , вимірює кореляцію між спостереженнями із затримкою k після усунення ефектів спостережень із проміжними затримками. Він визначається як коефіцієнт авторегресійної моделі порядку k^{th} після врахування проміжних лагів. PACF допомагає визначити порядок авторегресійного компонента в моделі ARIMA.

- Інтерпретація: PACF забезпечує більш чітке уявлення про прямі залежності між спостереженнями з різними затримками. Значна часткова автокореляція з відставанням k означає, що існує прямий зв'язок між значеннями ряду k періодів окремо, не опосередкованих значеннями в проміжних лагах.

- Використання: PACF особливо корисний для визначення порядку процесів авторегресії (AR) у моделях ARIMA. Наприклад, якщо PACF відключається після затримки p , це припускає, що AR-модель порядку p може бути доречним.

ACF, і PACF важливі для вибору моделі та діагностики в аналізі часових рядів. Досліджуючи структуру спаду ACF і визначаючи значні стрибки PACF, аналітики можуть визначити відповідний порядок членів авторегресії та ковзного середнього в моделях ARIMA, допомагаючи точному прогнозуванню та розумінню основної динаміки даних часових рядів.

4. Моделі часових рядів для прогнозування: моделі часових рядів використовуються для фіксації та прогнозування закономірностей, що



спостерігаються в даних. Серед широко використовуваних моделей:

4.1. Авторегресійне інтегроване ковзне середнє (ARIMA): моделі ARIMA поєднують компоненти авторегресії (AR), диференційного (I) і ковзного середнього (MA), щоб охопити залежності та тенденції в даних. Вони ефективні для моделювання стаціонарних часових рядів.

4.2. Моделі експоненціального згладжування: моделі експоненціального згладжування, включаючи просте експоненціальне згладжування, лінійне експоненціальне згладжування Холта та сезонне експоненційне згладжування Холта-Вінтерса, корисні для прогнозування часових рядів із компонентами тренду та сезонності.

4.3. Моделі простору станів: моделі простору станів представляють основну структуру часового ряду з використанням неспостережуваних (латентних) станів і спостережуваних вимірювань. Вони забезпечують гнучку основу для моделювання та прогнозування часових рядів зі складними закономірностями.

5. Показники оцінки прогнозу: Оцінка точності та надійності прогнозів часових рядів має вирішальне значення. Кілька показників зазвичай використовуються для оцінки продуктивності прогнозу, включаючи середньоквадратичну помилку (RMSE), середню абсолютну відсоткову помилку (MAPE) і графіки помилок прогнозу. Ці показники допомагають порівняти різні моделі прогнозування та визначити найбільш придатний підхід для прогнозування споживання електроенергії.

Теоретичні основи аналізу часових рядів забезпечують міцну основу для розуміння часової динаміки споживання електроенергії. Застосовуючи відповідні моделі та методи, ми можемо ефективно фіксувати закономірності та залежності в даних, уможливлуючи точне прогнозування майбутніх моделей споживання електроенергії. У наступних розділах ці теоретичні концепції будуть застосовані на практиці з використанням реальних даних, закладаючи основу для прогнозування споживання електроенергії та аналізу його зв'язку з факторами економічного зростання.



4.2. Огляд досліджень, що вивчають зв'язок між споживанням електроенергії та факторами економічного зростання.

Значна кількість досліджень досліджувала складний зв'язок між споживанням електроенергії та факторами економічного зростання. Ці дослідження дають цінну інформацію про динаміку та залежності між попитом на енергію та економічним розвитком. У цьому розділі представлено огляд відомих досліджень, які вивчали зв'язок між споживанням електроенергії та економічним зростанням, проливаючи світло на фактори, що впливають на моделі споживання електроенергії.

У дослідженні Apergis і Payne (2010) під назвою «Споживання енергії та економічне зростання: докази нелінійного тесту причинності Грейнджера» панельні дані з 21 країни ОЕСР були проаналізовані для вивчення довгострокового зв'язку між споживанням електроенергії та економікою. Автори знайшли докази двонаправленого причинно-наслідкового зв'язку між двома змінними, що свідчить про те, що економічне зростання стимулює споживання електроенергії, а доступність енергії сприяє економічному розвитку. [2]

Чжан і Сонг (2021) провели дослідження під назвою «Аналіз і прогноз структури енергоспоживання Китаю, Енергетична політика». Дослідження було зосереджено на короткостроковому прогнозуванні навантаження споживання електроенергії в Китаї. Застосовуючи методи глибокого навчання, зокрема рекурентні нейронні мережі (RNN), автори досягли точних прогнозів попиту на електроенергію. Дослідження підкреслює важливість передових методів моделювання для підвищення ефективності та результативності прогнозування навантаження. [3]

Дослідження, проведене Ліном і Вангом (2012) під назвою «Зв'язки між споживанням енергії, економічним зростанням, відносною ціною, прямими іноземними інвестиціями та фінансовим розвитком у Малайзії», розглядає зв'язок між економічним зростанням і споживанням електроенергії в Азіатсько-



Тихоокеанському регіоні. Використовуючи панельний аналіз даних, автори виявили позитивний і значний зв'язок між економічним зростанням і споживанням електроенергії, що вказує на те, що економічна експансія призводить до збільшення попиту на енергію.[4]

У своєму дослідженні під назвою «Споживання електроенергії та економічне зростання: дані групи з 44 країн» Нараян і Сміт (2008) дослідили довгостроковий зв'язок між споживанням електроенергії та економічним зростанням у групі з 44 країн. У дослідженні використовувалися методи коінтеграції панелей і було виявлено докази довгострокового рівноважного зв'язку між двома змінними. Результати свідчать про те, що для досягнення сталого економічного зростання слід розглянути політику енергозбереження.[4]

Башер і Садорські (2006) провели дослідження під назвою «Ризик цін на нафту та фондові ринки, що розвиваються», у якому досліджували зв'язок між споживанням електроенергії, економічним зростанням і викидами CO₂ у групі з 64 країн. Їхні висновки виявили позитивний зв'язок між споживанням електроенергії та економічним зростанням, вказуючи на те, що економічна експансія сприяє збільшенню попиту на енергію. Дослідження підкреслює важливість розуміння взаємодії між економічним зростанням, споживанням енергії та екологічною стійкістю.[6]

Ці дослідження підкреслюють важливість вивчення зв'язку між споживанням електроенергії та факторами економічного зростання. Отримані дані свідчать про двонаправлений причинно-наслідковий зв'язок, що вказує на взаємопідсилювальний характер попиту на енергію та економічного розвитку.

Крім того, передові методи моделювання, такі як глибоке навчання та аналіз панельних даних, продемонстрували свою ефективність у точному прогнозуванні споживання електроенергії та оцінці впливу економічного зростання на попит на енергію. Спираючись на ці попередні дослідницькі зусилля, це дослідження спрямовано на подальше розуміння складного зв'язку між споживанням електроенергії та факторами економічного зростання.



4.3. Методи попередньої обробки для обробки даних часових рядів.

Попередня обробка даних часових рядів є важливим кроком у їх підготовці до аналізу та моделювання. Далі описано різні методи попередньої обробки, які використовуються для ефективної обробки даних часових рядів, включаючи очищення даних, згладжування та інші необхідні перетворення.

1. Очищення даних: дані часових рядів часто містять відсутні значення, викиди або помилкові спостереження, які можуть вплинути на точність і надійність наступних аналізів. Методи очищення даних включають виявлення та усунення цих аномалій, щоб забезпечити цілісність набору даних. Відсутні значення можна обробити за допомогою методів імпутації, таких як імпутація середнього значення, пряме заповнення або зворотне заповнення. Викиди можна виявити за допомогою статистичних вимірювань, таких як z-показник або інтерквартильний діапазон (IQR), і їх можна видалити або скорегувати залежно від характеру аналізу.

2. Згладжування даних: методи згладжування використовуються для зменшення шуму та нерегулярних коливань, присутніх у даних часових рядів, що дозволяє ідентифікувати базові тенденції та закономірності. Ковзні середні та експоненціальне згладжування є поширеними методами згладжування, які застосовуються до даних часових рядів. Ковзні середні передбачають обчислення середнього підмножини суміжних точок даних, згладжуючи короткочасні коливання. Формула простого ковзного середнього(SMA):

$$SMA = \frac{A_1 + A_2 + \dots + A_n}{n}. \quad (2)$$

Експоненціальне згладжування призначає ваги минулим спостереженням, надаючи більшого значення останнім точкам даних, формула експоненціального ковзного середнього (ЕМА):

$$EMA_t = [V_t \times s] + EMA_y \times [1 - s] \quad (3)$$



Де:

EMA_t = ЕМА сьогодні,

V_t = Вартість сьогодні,

EMA_y = ЕМА вчора,

s = константа згладжування обчислюється як $s = \left(\frac{2}{d+1}\right)$, d = кількість днів.

Ці методи допомагають виявити основні закономірності та зробити дані більш придатними для моделювання та аналізу[13].

3. Перетворення даних: Методи перетворення даних використовуються для досягнення стаціонарності, особливо при роботі з нестаціонарними даними часових рядів. Загальні перетворення включають диференціювання, логарифмічне перетворення або перетворення Бокса-Кокса. Диференціювання передбачає віднімання поточного спостереження з його попереднього значення, щоб усунути тенденцію або сезонність.

$$\Delta y_t = y_t - y_{t-1} \quad (4)$$

Тут Δy_t представляє диференційований ряд, y_t це значення на час t , і y_{t-1} це значення в часі $t - 1$. Ця формула видаляє з даних будь-який лінійний тренд.

Повторне диференціювання. Якщо перше диференціювання не дає стаціонарності, можна обчислити повторно:

$$\Delta^2 y_t = \Delta y_t - \Delta y_{t-1} = (y_t - y_{t-1}) - (y_{t-1} - y_{t-2}) \quad (5)$$

Спрощуючи, отримуємо:

$$\Delta^2 y_t = y_t - 2y_{t-1} + y_{t-2} \quad (6)$$

Друге диференціювання усуває квадратичні тренди з даних.

Сезонне диференціювання використовується для видалення сезонних



моделей у даних. Для ряду з сезонним періодом s , сезонна різниця визначається як:

$$\Delta_s y_t = y_t - y_{t-s} \quad (7)$$

Ця формула віднімає значення спостереження з того самого сезону в попередньому циклі усуваючи сезонні ефекти.

Застосовуючи ці методи диференціювання, ви можете стабілізувати середнє значення ваших часових рядів, видалити тенденції та сезонність і підготувати дані для подальшого аналізу та моделювання[13].

Логарифмічне перетворення та перетворення Бокса-Кокса можна застосувати для стабілізації дисперсії та зменшення асиметрії в даних відповідно. Ці перетворення допомагають гарантувати, що дані відповідають припущенню про стаціонарність, уможливлуючи застосування відповідних моделей часових рядів. Формула:

$$d^m(t) = d^{m-1}(t) - d^{m-1}(t - 1); \quad (8)$$

4. Логарифмування даних. Може бути корисним для збільшення виражених розподілів і зменшення впливу викидів. Застосування логарифма до даних може допомогти зробити розподіл більш нормальним або стабільним. Це особливо корисно в тому випадку, коли дані зміщені або мають довгий хвіст, що ускладнює нормалізацію.

Переваги логарифмування:

- Зменшення асиметрії: якщо розподіл даних сильно спотворений або має довгий хвіст, логарифмізація може допомогти зменшити асиметрію та зробити розподіл більш симетричним.
- Стабілізація дисперсії: у наборах даних із гетероскедастичністю (різні рівні дисперсії в діапазоні даних) логарифмізація може стабілізувати дисперсію, полегшуючи моделювання взаємозв'язків і шаблонів.



- Пом'якшення викидів: логарифмізація може пом'якшити вплив викидів шляхом стискання великих значень. Це особливо корисно, коли викиди присутні, і їх вплив потрібно мінімізувати без повного видалення.

- Можливість інтерпретації: перетворення даних за допомогою логарифмування іноді може призвести до результатів, які можна краще інтерпретувати, особливо в контекстах, де зв'язок між змінними легше зрозуміти в логарифмічній шкалі (наприклад, експоненціальне зростання або спад).

Застосування в аналізі даних:

- Викривлені розподіли: Логарифмізація особливо ефективна для обробки змінних із сильно викривленими розподілами, таких як дохід, чисельність населення або економічні показники.

- Економетричний аналіз: в економетриці логарифмізація часто використовується для лінеаризації зв'язків або для того, щоб вони відповідали припущенням про нормальність і постійну дисперсію.

Логарифмізація — це цінна техніка перетворення даних, яка може покращити властивості розподілу змінних, роблячи їх більш придатними для статистичного аналізу та моделювання. Це особливо корисно під час роботи зі спотвореним розподілом даних і викидами, надаючи спосіб нормалізації та стабілізації даних для подальшого аналізу. Однак важливо враховувати конкретні характеристики даних і належним чином інтерпретувати результати після перетворення[13].

5. Нормалізація даних. Методи масштабування та нормалізації використовуються для приведення змінних до спільного масштабу та розподілу, полегшуючи порівняння та аналіз. Два найпоширеніші методи – це мінімально-максимальне масштабування та нормалізація Z-оцінки.

Основна мета нормалізації - відобразити різні дані різних одиниць вимірювання і діапазону призначень відомому поданні, порівняти їх один з одним або використовувати для розрахунку подібності об'єктів. Насправді це необхідно, наприклад, при кластеризації і деяких алгоритмах машинного навчання.



Мінімально-максимальне масштабування перетворює дані в певний діапазон, як правило, від 0 до 1, використовуючи таку формулу:

$$X_{scaled} = (X - X_{min}) / (X_{max} - X_{min}) \quad (9)$$

Де:

X_{scaled} = масштабоване значення змінної,

X = вихідне значення змінної,

X_{min} = Мінімальне значення змінної,

X_{max} = Максимальне значення змінної.

Нормалізація підтримує зв'язки між різними ознаками з точки зору їх відносних величин.

Нормалізація Z-показника стандартизує змінні шляхом віднімання середнього значення та ділення на стандартне відхилення, в результаті чого отримується розподіл із середнім значенням 0 і стандартним відхиленням 1. Формула нормалізації Z-показника:

$$X_{normalized} = (X - \mu) / \sigma \quad (10)$$

Де:

$X_{normalized}$ = Нормалізоване значення змінної,

X = вихідне значення змінної,

μ = Середнє значення змінної,

σ = стандартне відхилення змінної.

6. Повторна вибірка: Дані часових рядів часто збираються з високою частотою (наприклад, щогодини чи добові інтервали), що може бути непридатним для поточного аналізу. Техніки повторної дискретизації, такі як підвищення або зменшення дискретизації, можна використовувати для налаштування часових інтервалів даних. Підвищена дискретизація передбачає збільшення частоти даних шляхом інтерполяції відсутніх значень, тоді як



зменшення дискретизації передбачає зменшення частоти шляхом об'єднання кількох точок даних у більші часові інтервали. Повторна вибірка дозволяє краще узгодити з вимогами аналізу та зменшує складність обчислень.

Використовуючи ці методи попередньої обробки, ми можемо гарантувати, що дані часових рядів очищаються, згладжуються та трансформуються відповідним чином для подальшого аналізу. Отримані попередньо оброблені дані більше підходять для моделювання, прогнозування та дослідження зв'язку між споживанням електроенергії та факторами економічного зростання. У наступних розділах дослідження ці попередньо оброблені дані часових рядів будуть піддані передовим методам аналізу, таким як моделювання часових рядів і статистичний аналіз, використовуючи програмне забезпечення JMP, щоб отримати глибше розуміння динаміки споживання електроенергії та його зв'язку з факторами економічного зростання.

4.4. Методи декомпозиції часових рядів: методи виявлення та аналізу тенденцій, сезонності та залишкових компонентів у даних про споживання електроенергії.

Декомпозиція часових рядів є фундаментальним кроком у розумінні базових компонентів часових рядів, включаючи тренд, сезонність і залишкові коливання. У цьому розділі розглядаються різні методи, які використовуються для декомпозиції даних часових рядів, надаючи розуміння характеристик кожного компонента.

1. Компонент тенденції: Компонент тенденції представляє довгостроковий напрям або закономірність, що спостерігається в часовому ряді. Він фіксує загальне збільшення або зменшення даних протягом тривалого періоду, ігноруючи короткочасні коливання. Для виділення компонента тренду можна використовувати різні методи, такі як ковзні середні, регресійний аналіз або вдосконалені методи, такі як експоненціальне згладжування або LOESS



(локально зважене згладжування діаграми розсіювання). Ці методи допомагають визначити основну тенденцію, згладжуючи шум і нерівності в даних.

2. Сезонний компонент: сезонний компонент представляє повторювані та передбачувані моделі, які відбуваються протягом певного періоду часу, наприклад щоденні, тижневі або річні цикли. Він фіксує систематичні варіації часових рядів, пов'язані з такими факторами, як погодні умови, свята чи інші повторювані події. Для виділення сезонної складової можна використовувати декілька методів, включаючи сезонні індекси, аналіз Фур'є або сезонну декомпозицію часових рядів (STL). Ці методи допомагають визначити періодичність і амплітуду сезонних коливань.

3. Залишковий компонент: Залишковий компонент, також відомий як залишок або термін помилки, представляє незрозумілу зміну в часовому ряді після видалення тенденції та сезонних компонентів. Він складається з випадкових або непередбачуваних коливань, які не можна віднести до будь-якої конкретної моделі чи тенденції. Залишковий компонент дає уявлення про нерегулярну або ідіосинкратичну поведінку часових рядів. Його можна розрахувати шляхом віднімання розрахункової тенденції та сезонних компонентів із вихідного часового ряду.

Методи декомпозиції часових рядів дозволяють нам отримати глибше розуміння окремих компонентів, які вносять свій внесок у загальну структуру даних. Виділяючи тренд, сезонність і залишкові компоненти, аналітики можуть оцінити їхні відповідні внески в часовий ряд і вивчити їх зв'язки. Ця декомпозиція сприяє точнішому прогнозуванню, фіксуючи притаманні закономірності та динаміку даних.

4.5. Вибір відповідних моделей: визначення та оцінка методів для точного прогнозу споживання електроенергії.

Вибір правильної моделі прогнозування має вирішальне значення для



точного прогнозування майбутнього споживання електроенергії на основі історичних даних і вивчення його зв'язку з факторами економічного зростання. У цьому розділі обговорюються різні моделі прогнозування, які зазвичай використовуються в аналізі часових рядів, і наводяться фактори, які слід враховувати під час вибору найбільш прийнятної моделі.

1. Моделі ARIMA: моделі авторегресійної інтегрованої ковзної середньої (ARIMA) широко використовуються для прогнозування часових рядів. Моделі ARIMA фіксують компонент авторегресії (AR), компонент ковзного середнього (MA) і компонент диференціації (I), щоб врахувати тенденцію, сезонність і стаціонарність даних. Ці моделі є гнучкими та здатні обробляти широкий діапазон шаблонів часових рядів. Відповідний порядок моделі ARIMA, позначений як ARIMA(p, d, q), визначається шляхом аналізу автокореляційних і часткових автокореляційних функцій даних часового ряду.

2. Моделі експоненціального згладжування: Моделі експоненціального згладжування, включаючи просте експоненціальне згладжування (SES), лінійне експоненціальне згладжування Холта та потрійне експоненційне згладжування Холта-Вінтерса, популярні для прогнозування даних часових рядів. Ці моделі призначають різну вагу минулим спостереженням, надаючи більшого значення останнім даним. Моделі експоненціального згладжування особливо корисні для фіксації тенденцій і сезонності в даних, і їх можна адаптувати для роботи з різними рівнями складності.

3. Моделі сезонного прогнозування: коли часовий ряд демонструє помітні сезонні моделі, моделі сезонного прогнозування стають актуальними. Ці моделі включають сезонне ARIMA (SARIMA), сезонне експоненціальне згладжування (SES) і сезонне розкладання часових рядів (STL). Ці моделі чітко враховують сезонні коливання та пропонують покращену точність прогнозування для даних із повторюваними моделями.

4. Моделі на основі машинного навчання: В останні роки методи машинного навчання, такі як нейронні мережі, опорна векторна регресія та випадкові ліси, набули популярності у прогнозуванні часових рядів. Ці моделі мають здатність



фіксувати складні нелінійні зв'язки та обробляти великомасштабні набори даних. Моделі, засновані на машинному навчанні, часто вимагають більше обчислювальних ресурсів і детального налаштування параметрів, але можуть забезпечити точні прогнози, особливо коли дані демонструють складні закономірності або нелінійну динаміку.

При виборі відповідної моделі прогнозування слід враховувати кілька факторів, включаючи характеристики часових рядів даних, наявність тенденції та сезонності, наявність історичних даних і бажаний горизонт прогнозування. Важливо також оцінити припущення моделі, оцінити її ефективність на основі даних перевірки та порівняти її з альтернативними моделями, використовуючи відповідні показники, такі як середня квадратична помилка (MSE), середня абсолютна помилка (MAE) або показники точності прогнозу, такі як середній абсолютний відсоток. Помилка (MAPE).

В подальшому різні моделі прогнозування, включаючи ARIMA, експоненціальне згладжування та потенційно моделі на основі машинного навчання, будуть застосовані до попередньо оброблених даних часових рядів споживання електроенергії. Ефективність і придатність кожної моделі буде оцінено за допомогою статистичних показників і методів перевірки. Цей комплексний аналіз дасть змогу зрозуміти найбільш ефективний підхід до прогнозування споживання електроенергії та дозволить дослідити його взаємозв'язок із факторами економічного зростання.

4.6. Оцінка моделі та критерії вибору: оцінка ефективності моделі прогнозування за допомогою YITS, BITS, RMSE, MAPE та інших показників.

Точна оцінка та вибір моделей прогнозування мають важливе значення для забезпечення надійних і надійних прогнозів споживання електроенергії та визначення факторів, що впливають на його зв'язок з економічним зростанням.



Розглянемо критерії оцінки, які зазвичай використовуються для оцінки ефективності моделей прогнозування, і допомога в їх виборі.

1. Інформаційний критерій Акаїке (AIC): AIC є широко використовуваним критерієм вибору моделі, який вимірює компроміс між складністю моделі та придатністю. Він враховує кількість параметрів у моделі та функцію ймовірності, штрафуючи моделі з більшою кількістю параметрів. Чим нижче значення AIC, тим кращий баланс моделі між придатністю та складністю.

2. Байєсівський інформаційний критерій (BIC): Подібно до AIC, BIC є ще одним критерієм, який використовується для вибору моделі. Він містить сильніший штраф за складність моделі, додаючи більшу вагу кількості параметрів. BIC, як правило, вибирає простіші моделі порівняно з AIC. Як і у випадку з AIC, нижче значення BIC вказує на ефективнішу модель.

3. Середньоквадратична помилка (RMSE): RMSE – це широко використовувана міра точності прогнозу, яка обчислює середню різницю між спостережуваними значеннями та прогнозованими значеннями, зведеними в квадрат, а потім виведеними з кореня. Він забезпечує загальний показник того, наскільки добре модель передбачає майбутні значення часового ряду. Нижче значення RMSE вказує на кращу точність.

4. Середня абсолютна відсоткова похибка (MAPE): MAPE вимірює різницю у відсотках між спостережуваними значеннями та прогнозованими значеннями, усередненими за прогнозований період. Він забезпечує відносну міру точності прогнозу, що особливо корисно під час порівняння моделей у різних часових рядах або горизонтах прогнозування. Нижче значення MAPE означає кращу точність.

5. Декомпозиція дисперсії помилки прогнозу (FEVD): FEVD використовується для аналізу внеску кожної змінної або фактора в пояснення дисперсії помилки прогнозу. Це допомагає визначити відносну важливість різних факторів, що впливають на точність прогнозу. Розкладаючи дисперсію помилки прогнозу, ми можемо отримати уявлення про те, які фактори мають сильніший вплив на прогноз споживання електроенергії.



6. Надійність і стабільність моделі: в додаток до конкретних критеріїв оцінки, згаданих вище, надійність і стабільність моделі є ключовими міркуваннями. Надійна модель має стабільно добре працювати в різних підмножинах даних або в різних періодах часу. Стабільність означає здатність моделі створювати послідовні прогнози при застосуванні до подібних наборів даних або за наявності незначних змін у даних.

Використовуючи ці критерії оцінки, є можливість оцінити ефективність моделей прогнозування та прийняти обґрунтовані рішення щодо вибору моделі. Важливо враховувати кілька критеріїв у комплексній оцінці, щоб забезпечити збалансовану оцінку точності, складності та можливості узагальнення моделі. Вибрані моделі слугуватимуть надійним інструментом в цьому дослідженні.

4.7. Повний огляд програмного забезпечення JMP та його розширених можливостей для статистичного аналізу даних.

JMP — це потужний програмний інструмент, призначений для дослідницького аналізу даних, статистичного моделювання та візуалізації. Розроблений SAS Institute, JMP надає дослідникам і аналітикам зручний інтерфейс і широкий спектр інструментів для ефективного аналізу та інтерпретації складних наборів даних. У цьому розділі представлено огляд JMP і висвітлено його можливості для аналізу даних.

Зручний інтерфейс: JMP пропонує інтуїтивно зрозумілий інтерактивний інтерфейс, який дозволяє користувачам легко переміщатися між своїми даними та виконувати різні аналізи. Програмне забезпечення наголошує на візуальному дослідженні, що дозволяє користувачам динамічно взаємодіяти зі своїми даними та швидко отримувати розуміння. Функція перетягування та операції «наведіть і клацніть» роблять його доступним як для новачків, так і для досвідчених користувачів.

Імпорт і підготовка даних: JMP підтримує імпорт даних із різних форматів



файлів, включаючи електронні таблиці, бази даних і текстові файли. Він надає ефективні інструменти підготовки даних для обробки відсутніх значень, викидів та інших завдань очищення даних. Можливості обробки даних JMP дозволяють користувачам перетворювати, об'єднувати та фільтрувати набори даних, гарантуючи, що дані мають відповідний формат для аналізу.

Дослідницький аналіз даних (EDA): JMP чудово підходить для дослідницького аналізу даних, надаючи широкий спектр статистичних і графічних методів для виявлення закономірностей, взаємозв'язків і викидів у даних. Користувачі можуть генерувати описову статистику, створювати підсумкові таблиці та діаграми та виконувати інтерактивну візуалізацію, щоб отримати повне розуміння розподілу даних, мінливості та тенденцій.

Статистичне моделювання: JMP пропонує різноманітні методи статистичного моделювання для дослідження зв'язків, побудови прогнозних моделей і перевірки гіпотез. Він підтримує регресійний аналіз, дисперсійний аналіз (ANOVA), узагальнені лінійні моделі (GLM) і багато інших передових статистичних методів. Програмне забезпечення забезпечує інтерактивне середовище побудови моделі, що дозволяє користувачам оцінювати відповідність моделі, інтерпретувати результати та приймати обґрунтовані рішення.

Аналіз часових рядів: JMP містить інструменти, спеціально розроблені для аналізу часових рядів, що дозволяє аналізувати часові дані, виявляти тенденції та сезонність, а також створювати моделі прогнозування. За допомогою JMP користувачі можуть застосовувати такі методи, як декомпозиція часових рядів, автокореляційний аналіз і моделювання ARIMA, щоб отримати уявлення про динаміку залежних від часу даних.

Візуалізація даних: JMP акцентує увагу на візуальному дослідженні та пропонує широкий набір можливостей динамічної візуалізації даних. Користувачі можуть створювати інтерактивні графіки, діаграми розсіювання, гістограми, коробкові діаграми, теплові карти та інші візуалізації, щоб розкривати закономірності, ідентифікувати викиди та ефективно повідомляти



результати. Візуалізації в JMP можна налаштувати, що дозволяє користувачам налаштовувати кольори, мітки та інші елементи дизайну для покращення візуального представлення своїх даних.

Звітування та співпраця: JMP надає функції для створення звітів і обміну аналізами з іншими. Користувачі можуть створювати індивідуальні звіти, які включають аналізи, графіки та статистичні дані, що полегшує документування та передачу результатів. JMP також підтримує експорт результатів у різні формати, включаючи PDF, Excel і PowerPoint, полегшуючи співпрацю та презентацію результатів.

Можливості JMP роблять його цінним інструментом для аналізу даних у цій дипломній роботі. Його зручний інтерфейс, комплексні статистичні інструменти та акцент на візуальному дослідженні дають змогу ефективно аналізувати дані часових рядів, моделювати зв'язки та отримувати цінну інформацію про динаміку споживання електроенергії та її зв'язок із факторами економічного зростання.

4.8. Регіональні відмінності у взаємозв'язку між ВВП та енергоспоживанням

Аналіз виявив інтригуючі регіональні відмінності у взаємозв'язку між валовим внутрішнім продуктом (ВВП) та споживанням енергії. Зокрема, у розвинутих країнах, які славляться своєю міцною економікою, наприклад у США, Німеччині та Південній Кореї, з'явилася суперечлива тенденція: незважаючи на рекордно високий рівень ВВП, споживання енергії в цих країнах за останні роки знизилася. Навпаки, в інших регіонах спостерігалася протилежна картина, коли вищий ВВП корелював зі збільшенням енергоспоживання.

Це несподіване відкриття спонукає до глибшого вивчення факторів, що лежать в основі цих розбіжних тенденцій. Кілька гіпотез можуть пояснити це явище:

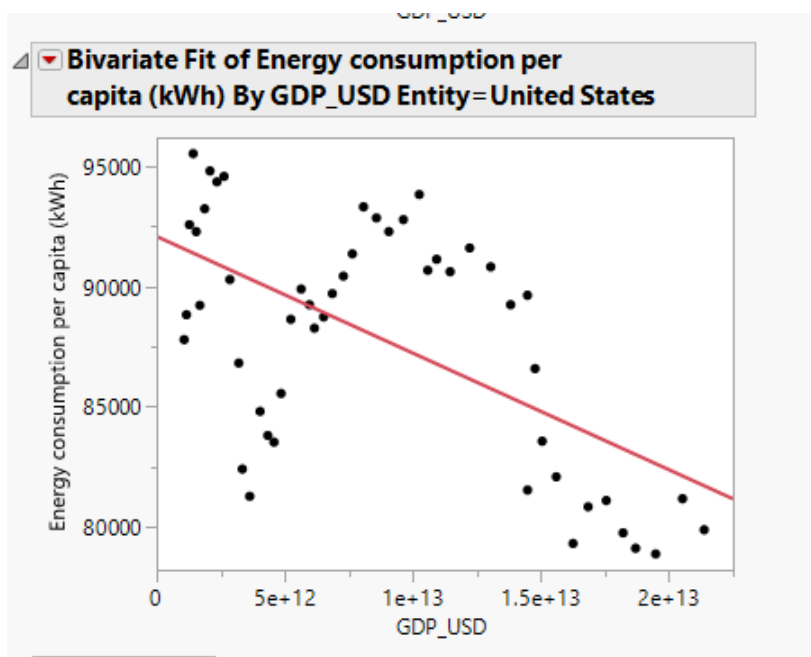


Рисунок 2 – Відношення енергоспоживання до ВВП в США

- Технологічний прогрес: розвинені країни часто використовують передові технології та методи енергоефективності для підвищення продуктивності та зменшення енергоємності. Отже, незважаючи на економічне зростання, їхні галузі можуть досягти вищих рівнів виробництва при зменшенні споживання енергії, що призведе до зниження загального споживання енергії.

- Структурні зміни в економіці. Структура розвинутих економік зазнає безперервної еволюції зі зрушенням у бік секторів, орієнтованих на послуги, які є менш енергоємними порівняно з традиційними виробничими галузями. У міру переходу економіки до галузей, що базуються на знаннях, моделі споживання енергії можуть відповідно змінюватися.

- Енергетична політика та екологічна обізнаність: розвинені країни зазвичай впроваджують сувору енергетичну політику та віддають пріоритет ініціативам сталого розвитку для пом'якшення впливу на навколишнє середовище. Ця увага до енергоефективності та впровадження відновлюваних джерел енергії може сприяти спостережуваному зниженню споживання енергії, незважаючи на економічне зростання.

- Аутсорсинг і глобалізація: розвинені країни часто передають енергоємні



виробничі процеси в країни з економікою, що розвивається, з нижчими витратами на робочу силу. У результаті їхнє внутрішнє споживання енергії може зменшитися, тоді як споживання енергії, пов'язане з імпортними товарами, може бути віднесено до інших регіонів.

- Інвестиції в інфраструктуру: розвинені країни часто інвестують кошти в модернізацію енергетичної інфраструктури та просування енергозберігаючих технологій. Ці інвестиції можуть призвести до підвищення ефективності та зниження загального попиту на енергію, навіть якщо ВВП продовжує зростати.

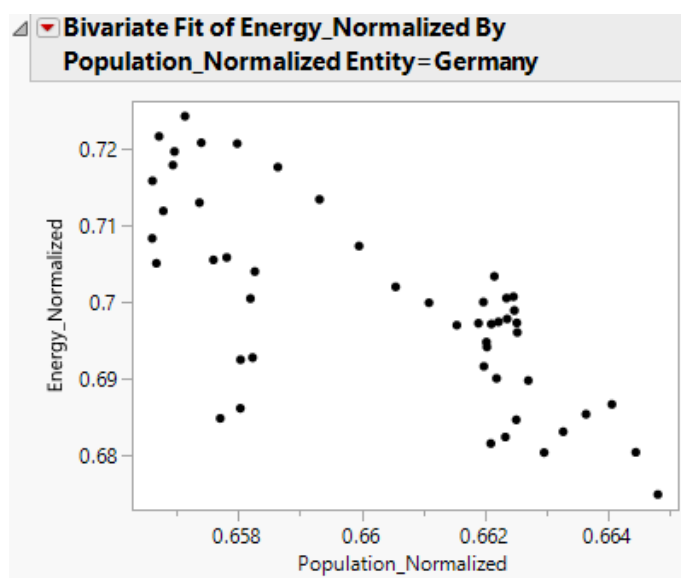


Рисунок 3 – Відношення енергоспоживання до кількості населення в Німеччині.

Підсумовуючи, контрастні тенденції, що спостерігаються у взаємозв'язку між ВВП та споживанням енергії, підкреслюють багатогранний характер економічного розвитку та динаміки енергетики. Подальші дослідження та поглиблений аналіз виправдані для вивчення складної взаємодії факторів, що формують моделі споживання енергії в різних регіонах, і для інформування про формування стратегічної політики для майбутнього сталого енергетичного розвитку.

Висновки

Мета цього дослідження полягала в тому, щоб передбачити споживання



електроенергії за допомогою аналізу часових рядів і вивчити залежність споживання електроенергії від факторів економічного зростання. Використовуючи можливості програмного забезпечення JMP, ми застосували надійні методики, такі як збір даних, попередня обробка, декомпозиція часових рядів, вибір моделі та оцінка для розширеного статистичного аналізу. Наш аналіз включає набір даних про глобальне споживання енергії на душу населення та дані про ВВП/ВВП на душу населення з таких джерел, як Kaggle і Our World.

Методичний внесок цього дослідження :

1. Збір і попередня обробка даних: Дані ретельно збиралися з надійних джерел, таких як Kaggle і Our World in Data, за період у 49 років. Етапи попередньої обробки включали обробку відсутніх даних, виявлення та лікування викидів, нормалізацію даних і об'єднання наборів даних щодо споживання енергії та ВВП.

2. Дослідницький аналіз даних (EDA): детальний EDA проводився з використанням надійних інструментів візуалізації JMP, щоб зрозуміти базові закономірності, тенденції та зв'язки в даних.

3. Аналіз часових рядів і розбіжності: дані часових рядів аналізували за допомогою таких методів, як декомпозиція, щоб визначити тенденцію, сезонність і залишкові компоненти. Для досягнення стаціонарності було застосовано розрізнення, що є вирішальним кроком для точного моделювання часових рядів.

4. Моделювання ARIMA: для прогнозування споживання електроенергії було вибрано та впроваджено модель ARIMA (авторегресійна інтегрована ковзна середня). Параметри моделі були ретельно підібрані на основі AIC, BIC та інших критеріїв відбору, щоб забезпечити найкращу відповідність.

5. Оцінка та перевірка моделі. Ефективність моделі ARIMA оцінювалася за допомогою таких показників, як RMSE, MAE і MAPE. Для перевірки припущень моделі та перевірки на наявність будь-яких аномалій була проведена діагностика залишків.