ПРИКЛАДНА МАТЕМАТИКА, ІНФОРМАТИКА ТА ОБЧИСЛЮВАЛЬНА ТЕХНІКА

УДК 51-77 https://doi.org/10.31713/vt1202119

Кінда В. В., аспірант (Рівненський державний гуманітарний університет, м. Рівне)

РОБОТА ІЗ ЧАСОВИМИ РЯДАМИ. КОМПОНЕНТИ РЯДІВ ТА СТАТИСТИЧНІ ХАРАКТЕРИСТИКИ

Розглянуто часові ряди та його складові компоненти такі як тренд, сезонність та циклічність. Проведено опис та класифікацію основних видів тренду та сезонності. Досліджено типи моделей, компоненти яких мають адитивну та мультиплікативну поведінки. Розглянуто найпоширеніший варіант для аналізу тенденції ряду та його тренду, сезонності у вигляді застосування моделей із експоненційного згладжування. Визначено автокореляційну функцію та описано зв'язок автокореляційної функції часового ряду із наявністю тренду у відповідному часовому ряді. Визначено строгу стаціонарність та її необхідність для точного прогнозу часового ряду. Розглянуто методику зведення нестаціонарних рядів до стаціонарних шляхом застосування маніпуляцій та перетворень вхідного ряду. Описано один із критеріїв перевірки гіпотези на стаціонарність часового ряду Дікі-Фуллера.

Ключові слова: **часовий ряд; стаціонарний процес; тренд; сезонність; прогноз.**

З розвитком інтернету речей (internet of things) дедалі актуальнішими постають задачі збереження, структуризації та маніпуляції з даними. Більшість дослідників зійшлися на тому, що обсяг цифрової інформації подвоюється кожні 1.5 року. В своїй більшості (близько 95%) – це неструктуровані дані і близько 5 відсотків – структуровані дані, які складаються з різного роду баз даних чи іншої структурованої інформації.

Прикладом структурованої інформації можна назвати часові ряди. Часовий ряд – це ознака, значення якої вимірюється через постійно однакові часові проміжки. Проміжок часу може бути різний: секунда, день, рік і т.д. Ключовою особливістю є те, що вимірювання проходять через однакові часові інтервали, інакше отримаємо не часовий ряд, а випадковий процес який необхідно обробляти іншими методами.

Часові ряди(1) можуть бути багатомірними та одномірними [1].

Рис. 1. Простий приклад часового ряду. Промисловий індекс DJI

$$y_1, \ldots, y_T, \ldots, ; y_t \in R(R^n). \qquad (1)$$

До найпростіших прикладів економетричних часових рядів можна віднести: ринкові ціни валют/акцій (рис. 1); обсяги та ціни продажу товарів або послуг, тощо.

При прогнозуванні часових рядів ставиться умова, що значення ряду в минулому містять інформацію про його поводження в найбутньому, на відміну від класичних задач аналізу даних, де передбачається незалежність спостережень.

Загальна задача прогнозування має наступний вигляд:

Необхідно знайти f_T : $y_{T+h} \approx f_T(y_T, \ldots, y_1, h) \equiv \hat{y}_{T+h|T}$,

де $h \in \{1,2,\ldots,H\}, H$ – горизонт прогнозування.

Наступними компонентами часових рядів є тренд, сезонність та циклічність.

Під трендом розуміємо плавну довгострокову зміну рівня ряду. В категорії тренду можна розділити на наступні групи:

[1] відсутність тренду;
[2] лінійний тренд;
[3] експоненційний тренд;
[4] степеневий тренд.

Аналіз тенденції часового ряду досить часто відбувається за допомогою згладжування. Найбільш поширеним методом вирівнювання є просте експоненційне згладжування (Simple Exponential Smoothing).

$$S_t = \alpha y_t + (1-\alpha)S_{t-1},$$

де $\alpha \in (0,1)$.

У випадку застосування формули рекурсивно, кожне нове згладжене значення (яке є також прогнозом) обчислюється як зважене середнє поточного спостереження і згладженого ряду. Очевидно, результат згладжування залежить від параметра α. Якщо він рівний 1, то попередні спостереження повністю ігноруються. Якщо ж його значення рівне 0, то ігноруються поточні спостереження. Проблема SES полягає у тому, що не враховується тренд та сезонність та вона працює лише на малий горизонт прогнозу. Для виявлення тренду застосовуються більш складні моделі із класу експоненційного згладжування – модель Хольта може врахувати лише лінійний тренд, модель Хольта-Вінтерса може врахувати мультиплікативний експоненційний тренд та сезонність [2].

Під сезонністю розуміємо циклічні зміни рівня ряду із постійним періодом.

В термінах сезонності можна виділити:

[5] відсутність сезонності;

[6] адитивну сезонність;

[7] мультиплікативну сезонність.

Сезонні компоненти, за своєю природою, можуть бути адитивними або мультиплікативними. Наприклад, протягом грудня продажі певного виду товару збільшуються на 100 умовних одиниць щороку. Для того щоб врахувати сезонне коливання, можна додати в прогноз на кожен грудень 100 умовних одиниць (понад відповідного річного середнього). В цьому випадку сезонність – адитивна. Інший варіант – нехай в грудні продажі збільшилися на 40%, тобто в 1.4 рази. Тоді, якщо загальні продажі малі, то абсолютне збільшення продажів в грудні теж відносно мале. Якщо ж в цілому продажі великі, то абсолютне збільшення продажів буде пропорційно більшим. В цьому випадку сезонність буде мультиплікативною. Відмінність між двома видами сезонності полягає в тому, що в адитивної моделі сезонні флуктуації не залежать від значень ряду, тоді як в мультиплікативної моделі величина сезонних флуктуацій залежить від значень часового ряду (рис. 2) [4].

Адитивна модель:

$$Pr_t = S_t + I_{t-p}\,. \tag{2}$$

Мультиплікативна модель:

$$Pr_t = S_t * I_{t-p}, \tag{3}$$

де S_t позначає (просте) експоненціно згладжене значення ряду в момент t, I_{t-p} позначає згладжений сезонний фактор в момент $t-p$, p – довжина сезону. Таким чином, в порівнянні з простим експоненційним згладжуванням, прогноз уточнюється додаванням або множенням сезонної компоненти. Ця компонента оцінюється незалежно за допомогою простого експоненційного згладжування.

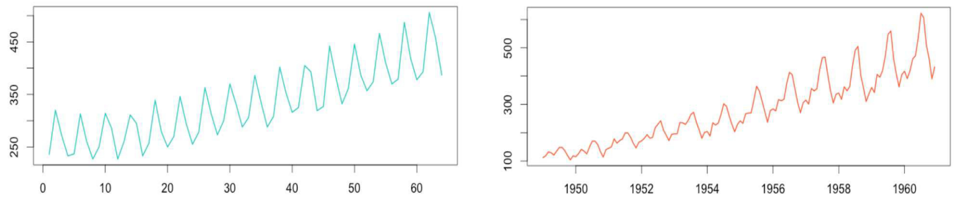

Рис. 2. а) об'єм виробництва води в Австралії (адитивна сезонність);
б) кількість пасажирських авіаперевезень в Австралії
(мультиплікативна сезонність)

Одним із перших необхідних кроків дослідження часового ряду є побудова автокореляційної функції – графіку автокореляції при різних лагах(4)

$$r_\tau = \frac{\mathbb{E}(y_t - \mathbb{E}_y)(y_{t+\tau} - \mathbb{E}y)}{\mathbb{D}y} \tag{4}$$

$r_\tau \in [-1,1]$, τ – лаг автокореляції.

Якщо кореляція при малих лагах монотонно спадає та коливається біля нуля синусоїдально при більших лагах — це є ознакою часового ряду із ярко вираженим трендом (рис. 3).

Оцінити значимість автокореляції для визначеного лагу можна за допомого статистичного критерію Стюдента, Q-критерію Люнга-Бокса. Більше можна дізнатись в літературі [2].

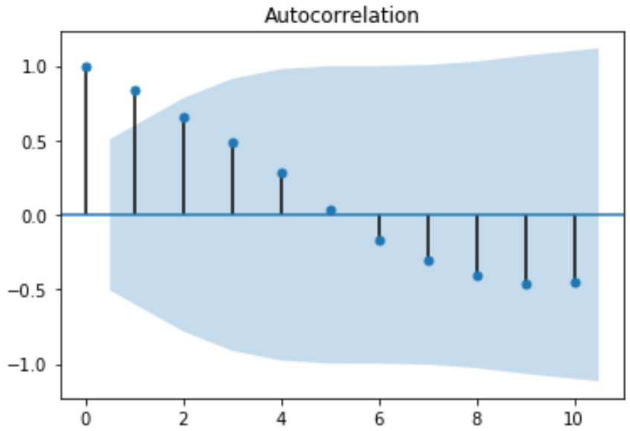

Рис. 3. Приклад автокореляційної функції побудованої в середовищі Python із використанням модуля statsmodels

Наступним пунктом при роботі із часовим рядом являється перевірка його на стаціонарність. $y_1, ..., y_T$ являється строго стаціонарним якщо для $\forall s$ розподіл $y_t, ..., y_{t+s}$ не залежить від t. Це означає, що характеристики ряду, а саме математичне сподівання та дисперсія — не змінюються в часі, а автоковаріаційна функція залежить лише від різниці часу $(t_1 - t_2)$.

Стаціонарність процесу є важливою, оскільки при ній легко будувати прогноз, так як ми вважаємо, що майбутні статистичні характеристики ряду не будуть відрізнятися від спостережуваних поточних. Більшість моделей часових рядів так чи інакше моделюють і прогнозують ці характеристики (наприклад, математичне сподівання або дисперсію), тому в разі нестаціонарності вихідного ряду прогноз виявиться неточним. На жаль, більшість часових рядів, з якими доводиться стикатися не є стаціонарними, але з цим можна та потрібно боротися.

Наявність тренду та сезонності в часовому ряді одразу свідчить про не стаціонарність ряду, проте можна спробувати звести його до стаціонарного. Є різні методики – логарифмування (монотонно змінюючу дисперсію переводить в константу). Логарифмування є частковим випадком однопараметричного перетворення Бокса-Кокса. Більш детально можна прочитати в авторів [3]. Після проведених маніпуляцій необхідно перевірити вихідний ряд на стаціонарність (можна використовувати критерій Дікі-Фуллера, або

KPSS).

Припустимо, є певний набір даних із значенням часового ряду. Постає завдання за наявними спостереженнями визначити його стаціонарність. Необхідно провести стандартну процедуру тестування гіпотези:

а) $H_0 : \phi = 1$ – процес не стаціонарний;

б) $H_1 : |\phi| < 1$ – альтернативна гіпотеза (тобто процес стаціонарний).

З тестуванням гіпотези не все так просто, тому що якщо справжнє значення $\phi = 1$, t – статистика не розподілена за законом Стьюдента і її розподіл не прямує до стандартного нормального при збільшенні кількості спостережень. В такому випадку ми не можемо просто взяти таблицю критичних значень Стьюдента і перевірити гіпотезу по ній[3].

Під t – статистикою тут розуміється відношення відхилення оцінки параметру моделі від його істинного значення до стандартної помилки оцінки коефіцієнту:

$$t = \frac{\hat{\phi} - \phi}{s_{\hat{\phi}}},$$

де $\hat{\phi}$ – оцінка параметра авторегресії першого порядку, $s_{\hat{\phi}}$ – стандартна похибка оцінки $\hat{\phi}$. Оцінка коефіцієнта $\hat{\phi}$ в альтернативній моделі може будуватися за допомогою звичайного методу найменших квадратів (МНК).

Висновки. В даній роботі проведено опис основних пунктів, які потрібно врахувати при роботі із часовими рядами. Наведо класифікацію основних характеристик ряду. Розглянуто адитивні та мультиплікативні моделі. Визначено строгу стаціонарність та методи маніпуляцій для зведежння нестаціонарних часових рядів до стаціонарних. Розглянуто критерій перевірки часового ряду на стаціонарність.

1. Box G. E. P. and Jenkins G. M. Time Series Analysis, Forecasting and Control, rev. Ed., San Francisco : Holden-Day, 1976 **2.** Statistical Analysis System. URL: sas.com/ru_ua/home.html (дата звернення: 12.03.2021). **3.** Stationarity in time series. URL: https://towardsdatascience.com/stationarity-in-time-series-analysis-90c94f27322 (дата звернення: 12.03.2021). **4.** Additive and Multiplicative models. URL: http://www-ist.massey.ac.nz/dstirlin/CAST/CAST/Hmultiplicative/multiplicative1.html (дата звернення: 12.03.2021).

REFERENCES:

1. Box G. E. P. and Jenkins G. M. Time Series Analysis, Forecasting and Control, rev. Ed., San Francisco : Holden-Day, 1976 **2.** Statistical Analysis System. URL: sas.com/ru_ua/home.html (data zvernennia: 12.03.2021). **3.** Stationarity in time series. URL: https://towardsdatascience.com/stationarity-in-time-series-analysis-90c94f27322 (data zvernennia: 12.03.2021). **4.** Additive and Multiplicative models. URL: http://www-ist.massey.ac.nz/dstirlin/CAST/CAST/Hmultiplicative/multiplicative1.html (data zvernennia: 12.03.2021).

Kinda V. V., Post-graduate Student (Rivne State University for the Humanities, Rivne)

WORK WITH TIME SERIES. COMPONENTS OF TIME SERIES AND STATISTICAL CHARACTERISTICS

Time series and its components such as trend, seasonality and cyclicity are considered. The description and classification of the main types of trend and seasonality are carried out. The components of time series which have additive and multiplicative behavior, are studied. The most widespread variant for the analysis of a tendency of time series and its trend, seasonality in the form of application of models on exponential smoothing is considered. The autocorrelation function is defined and the connection of the autocorrelation function of the time series with the presence of a trend in the corresponding time series is described. Strict stationarity and its necessity for the exact forecast of a time series are defined. The method of reduction of nonstationary series to stationary ones by manipulations and transformations of the input series is considered. One of the criteria for testing the hypothesis on the stationarity of the Dickey-Fuller time series is described.

***Keywords*: time series; stationary process; trend; seasonality; forecast.**

Кинда В. В., аспирант (Ровенский государственный гуманитарный университет, г. Ровно)

РАБОТА С ВРЕМЕННЫМИ РЯДАМИ. КОМПОНЕНТЫ РЯДОВ И СТАТИСТИЧЕСКИЕ ХАРАКТЕРИСТИКИ

Рассмотрены временные ряды и его компоненты такие как тренд, сезонность и цикличность. Проведено описание и классификацию основных видов тренда и сезонности. Исследованы типы моделей, компоненты которых имеют адитивное и мультипликативное поведения. Рассмотрены наиболее распространенный вариант для анализа тенденции ряда и его тренда, сезонности в виде применения моделей экспоненциального сглаживания. Определены автокорреляционной функции и описано связь автокорреляционной функции временного ряда с наличием тренда в соответствующем временном ряду. Определены строгую стационарность и ее необходимость для точного прогноза временного ряда. Рассмотрена методика возведения нестационарных рядов в стационарных путем применения манипуляций и преобразований входного ряда. Описан один из критериев проверки гипотезы на стационарность временного ряда Дики-Фуллера.

***Ключевые слова*: временной ряд; стационарный процесс; тренд; сезонность; прогноз.**
