

УДК 004.67

Воробець І. – ст. гр. СНм-61

*Тернопільський національний технічний університет імені Івана Пулюя*

## **ВИКОРИСТАННЯ МОДЕЛЕЙ ARIMA ДЛЯ ПРОГНОЗУВАННЯ ЧАСОВИХ РЯДІВ ІЗ ВЛАСТИВІСТЮ ЦИКЛІЧНОСТІ**

Науковий керівник: к.т.н., доц. Фриз М.Є.

Vorobets I.

*Ternopil Ivan Puluj National Technical University*

## **USING ARIMA MODELS FOR FORECASTING TIME SERIES WITH CYCLICITY FEATURE**

Supervisor: Ph.D., Assoc. Prof. Fryz M.Ye.

Ключові слова: прогнозування часових рядів, циклічність, моделі ARIMA.

Keywords: time series forecasting, cyclicity, ARIMA models.

Часовий ряд є упорядкованою відповідно до часу послідовністю значень певної змінної. По суті, часовий ряд варто розглядати як випадковий процес з дискретним часом.

Прогнозування часових рядів полягає у знаходженні майбутніх значень часового ряду з якомога найбільшою точністю на основі минулих значень часового ряду та інформації про його поведінку у минулому. Сфери застосування прогнозування часових рядів обмежуються лише використанням самих рядів. Так, прогнозування рядів має широкий спектр різних сфер застосування: від аналізу фондового ринку та оцінювання росту економіки окремих регіонів до моделювання поширення епідемій і прогнозу погоди.

Шаблон поведінки часового ряду визначається трьома його властивостями: трендом, циклічністю та сезонністю [1]. Тренд показує тенденцію зростання або спадання значень ряду протягом певного періоду часу. На великому часовому проміжку тренд може описувати зміну значень від зростання до спадання або навпаки. Властивість циклічності проявляється тоді, коли відбуваються коливання значень часового ряду без фіксованого періоду. Сезонність описує вплив на поведінку часового ряду різних сезонних факторів, наприклад, день тижня, місяць, квартал року. Відмінність сезонності та циклічності полягає в тому, що сезонність має фіксований і визначений період. Циклічність, натомість, передбачає довшу середню тривалість коливань і більш мінливу їх величину.

Кожен часовий ряд має зазвичай декілька таких властивостей. Саме на основі властивостей тренду, циклічності та сезонності й базуються методи та моделі прогнозування часових рядів. Одними з найбільш часто вживаних моделей прогнозування є моделі ARIMA (англ. autoregressive integrated moving average, укр. авторегресійне інтегроване ковзне середнє). Фактично моделі ARIMA є поєднанням авторегресійних моделей (англ. autoregressive model, AR) і моделей ковзного середнього (англ. moving average model, MA).

Авторегресійні моделі прогнозують значення часового ряду на основі його минулих значень. Моделі ковзного середнього визначає прогноз часового ряду на основі похибок минулих прогнозів. Основним обмеженням авторегресійних моделей і моделей ковзного середнього є стаціонарність часового ряду. Таке саме обмеження

мають і моделі ARIMA. При роботі з нестационарними даними моделі ARIMA використовують різниці часових рядів для досягнення їх стаціонарності.

Загалом модель ARIMA описується за формулою:

$$\Delta^d y_t = c + \phi_1 \Delta^d y_{t-1} + \dots + \phi_p \Delta^d y_{t-p} + \theta_1 \varepsilon_{t-1} + \dots + \theta_q \varepsilon_{t-q} + \varepsilon_t,$$

де  $y_t$  – значення часового ряду на момент часу  $t$ ,  $\Delta^d y_t$  – різниця ряду ступеня  $d$ , різниця першого ступеня записується як  $\Delta y_t = y_t - y_{t-1}$ ,  $c$  – константа,  $\varepsilon_t$  – значення білого шуму,  $\phi_1, \dots, \phi_p$  – коефіцієнти авторегресійної моделі,  $\theta_1, \dots, \theta_q$  – коефіцієнти моделі ковзного середнього.

Визначення моделі ARIMA для певного часового ряду полягає у виборі значень параметрів  $p$ ,  $d$  і  $q$ , що є досить складним завданням. Параметр  $p$  позначає порядок авторегресійної частини авторегресійної,  $d$  – ступінь різниці ряду,  $q$  – порядок частини ковзного середнього [1].

Для прогнозування часових рядів із циклічністю добре підходять моделі ARIMA. Важливе значення при цьому має параметр  $p$ . Так, для створення прогнозів циклічних рядів необхідною умовою є  $p > 1$  [2].

У ході використання моделі прогнозування отримуються точкові прогнози на визначений горизонт прогнозування. Оцінити отримані прогнози, можна за допомогою двох основних похибок: середньої абсолютної похибки (англ. mean absolute error, MAE) та середньої абсолютної відсоткової похибки (англ. mean absolute percentage error, MAPE) [1]. Похибка MAE обчислюється як середнє абсолютних значень похибок прогнозу та використовує ті самі одиниці, що й значення часового ряду. Помилка MAPE обчислюється як середнє абсолютних значень відсоткових похибок, які подають значення у відсотках і не залежать від одиниць значень ряду.

Для побудови моделей прогнозування часових рядів існує багато різних інструментів. У даній роботі прогнозування здійснюється за допомогою бібліотеки Statsmodels. Statsmodels – це відкрита бібліотека мови програмування Python, яка надає функції для проведення описової статистики, побудови й оцінки статистичних моделей, виконання статистичних тестів [3]. Зокрема, Statsmodels дозволяє працювати з моделями прогнозування часових рядів, у тому числі і з моделями ARIMA.

Завданням даної роботи є використання засобів бібліотеки Statsmodels для прогнозування часового ряду із властивістю циклічності. У якості такого часового ряду вибрано набір даних активності сонячних плям Національного центру геофізичних даних США [4]. Цей набір даних включає щорічні дані кількості сонячних плям, зібрані протягом 1700-2000 років.

### Література:

1. Hyndman R.J., Athanasopoulos G. Forecasting: principles and practice. 3rd ed. Melbourne, Australia : OTexts, 2021. 442 p.
2. Hyndman R.J. Cyclic and seasonal time series. URL: <https://robjhyndman.com/hyndsight/cyclistics>.
3. Introduction – statsmodels. URL: <https://www.statsmodels.org/stable/index.html>.
4. Solar Indices | NCEI. URL: <https://www.ngdc.noaa.gov/stp/solar/solar-indices.html>.