

УДК 519.25

**ПОРІВНЯЛЬНИЙ АНАЛІЗ МЕТОДІВ ПРОГНОЗУВАННЯ  
СТАЦІОНАРНИХ ТА НЕСТАЦІОНАРНИХ ЧАСОВИХ РЯДІВ**

**Перерва А. С.**

**кандидат технічних наук, доцент, Ковалюк Т. В.**

Національний технічний університет України “Київський політехнічний інститут імені Ігоря Сікорського”, Україна, Київ

*Розглянуто основні методи аналізу та прогнозування часових рядів. Досліджено ефективність методу авторегресії, інтегрованої моделі авторегресії ковзаючого середнього (ARIMA), методу аналізу сингулярного спектру (SSA) та методів з використанням нейромереж на основі глибокого навчання для прогнозування часових рядів. Ефективність досліджено на прикладі даних значень функції синусоїди, мінімальної денної температури, ринкового індексу S&P 500 та денної кількості пасажирів авіалінії за критерієм найменшої середньоквадратичної похибки. Проведено порівняльний аналіз отриманих результатів та обґрунтовано доцільність використання даних методів.*

*Ключові слова: часові ряди, авторегресія, ARIMA, інтегрована модель авторегресії ковзаючого середнього, SSA, аналіз сингулярного спектру, нейромережі.*

*Перерва А. С., к.т.н., доц., Ковалюк Т. В. Сравнительный анализ методов прогнозирования стационарных и нестационарных рядов/ Национальный технический университет Украины "Киевский политехнический институт имени Игоря Сикорского", Украина, Киев*

*Рассмотрены основные методы анализа и прогнозирования временных рядов. Исследована эффективность метода*

*авторегрессии, интегрированной модели авторегрессии скользящего среднего (ARIMA), метода анализа сингулярного спектра (SSA) и методов с использованием нейронных сетей на основе глубокого обучения для прогнозирования временных рядов. Эффективность исследовано на примере данных значений функции синусоиды, минимальной дневной температуры, рыночного индекса S&P 500 и дневного количества пассажиров авиалинии по критерию наименьшей среднеквадратичной погрешности. Проведен сравнительный анализ полученных результатов и обоснована целесообразность использования данных методов.*

*Ключевые слова: временные ряды, авторегрессия, ARIMA, интегрированная модель авторегрессии скользящего среднего, SSA, анализ сингулярного спектра, нейросети.*

*Pererva A. S., Candidate of Technical Sciences, Associate Professor, Kovalyuk T. V. Comparative analysis of prediction methods of stationary and nonstationary series/ National Technical University of Ukraine "Igor Sikorsky Kiev Polytechnic Institute", Ukraine, Kyiv*

*The basic methods of analysis and forecasting of time series are considered. The effectiveness of the method of autoregression, the model of autoregressive integrated moving average (ARIMA), the method of singular spectrum analysis (SSA) and methods using neural networks based on deep learning for forecasting of time series was investigated. The effectiveness was studied on the example of the values of the function of the sinusoid, the minimum daily temperature, the market index S&P 500 and the daily number of passengers of the airline on the criterion of the lowest mean square error. A comparative analysis of the obtained results is carried out and the feasibility of using these methods is substantiated.*

*Keywords: time series, autoregression, ARIMA, integrated model of autoregressive sliding average, SSA, analysis of singular spectrum, neural networks.*

**Вступ.** На сьогодні існує велика кількість методів прогнозування часових рядів. Найпоширеніші з них [1]: ковзаюче середнє, авторегресійні методи, методи з використанням нейронних мереж та аналіз сингулярного спектру. Деякі з них добре працюють для стаціонарних часових рядів, в той час як успіх прогнозування нестаціонарних рядів значно залежить від їх предметної області.

**Мета статті.** Головною метою цієї роботи є визначення ефективності прогнозування різних видів часових рядів за допомогою математичних методів: авторегресії, ARMA (модель авторегресії ковзаючого середнього), ARIMA (інтегрована модель авторегресії ковзаючого середнього), SSA (аналіз сингулярного спектру) та методів з використанням нейромереж. Ефективність перелічених методів розглядатиметься за критерієм найменшої квадратичної помилки на прикладі таких часових рядів:

- синусоїда
- денна мінімальна температура [2]
- ринковий індекс S&P 500 [3]
- кількість пасажирів авіалінії [4]

**Виклад основного матеріалу.** Часовий ряд – це зібрані в різні моменти часу статистичні дані про значення будь-яких параметрів (в найпростішому випадку одного) досліджуваного процесу. Виявлення структури часового ряду необхідно для того, щоб побудувати математичну модель того явища, яке є джерелом часового ряду, що аналізується. Прогноз майбутніх значень часового ряду використовується для ефективного прийняття рішень та може бути

застосований у багатьох галузях, наприклад: кліматологія, океанологія, геофізика, техніка, обробка зображень, медицина.

Авторегресійна модель – це модель часових рядів, в якій значення часового ряду в даний момент лінійно залежать від попередніх значень цього ж ряду. Авторегресійний процес порядку  $p$  визначається наступним чином:

$$X_t = c + \sum_{i=1}^p a_i X_{t-i} + \varepsilon_t, \quad (1)$$

де  $a_1, \dots, a_p$  – параметри моделі (коефіцієнти авторегресії);

$c$  – константа;

$\varepsilon_t$  – білий шум.

Модель авторегресії ковзаючого середнього (ARMA) – це математична модель аналізу та прогнозування стаціонарних часових рядів, що є узагальненням моделі авторегресії та моделі ковзаючого середнього.

Моделлю  $ARMA(p, q)$ , де  $p$  і  $q$  – цілі числа, що задають порядок моделі, називається наступний процес генерації часового ряду  $X_t$ :

$$X_t = c + \varepsilon_t + \sum_{i=1}^p a_i X_{t-i} + \sum_{j=0}^q b_j \varepsilon_{t-j}, \quad (2)$$

де  $a_1, \dots, a_p$  – коефіцієнти авторегресії;

$b_0, \dots, b_q$  – коефіцієнти ковзаючого середнього;

$c$  – константа;

$\varepsilon_t$  – білий шум.

Така модель може інтерпретуватися як лінійна модель множинної регресії, в якій в якості пояснюючих змінних виступають минулі значення самої залежної змінної, а в якості регресійного залишку –

ковзаючі середні з елементів білого шуму. ARMA-процеси мають більш складну структуру в порівнянні зі схожими за поведінкою авторегресійною моделю або моделю ковзаючого середнього в чистому вигляді, але при цьому ARMA-моделі характеризуються меншою кількістю параметрів, що є однією з їхніх переваг.

Для побудови моделі ARMA по серії спостережень необхідно визначити порядок моделі (числа  $p$  і  $q$ ), а потім і самі коефіцієнти. Для визначення порядку моделі може застосовуватися дослідження таких характеристик часового ряду, як його автокореляційна функція і часткова автокореляційна функція. Для визначення коефіцієнтів авторегресії застосовуються такі методи, як метод найменших квадратів і метод максимальної правдоподібності.

Інтегрована модель авторегресії ковзаючого середнього (ARIMA) - це розширення моделі ARMA для нестационарних часових рядів, які можна зробити стаціонарними визначенням різниць деякого порядку від вихідного часового ряду (так звані інтегровані або різницево-стаціонарні часові ряди). Має наступний вигляд:

$$\Delta^d X_t = c + \varepsilon_t + \sum_{i=1}^p a_i \Delta^d X_{t-i} + \sum_{j=0}^q b_j \varepsilon_{t-j}, \quad (3)$$

де  $a_1, \dots, a_p$ - коефіцієнти авторегресії;

$b_0, \dots, b_q$ - коефіцієнти ковзаючого середнього;

$c$ - константа;

$\varepsilon_t$ - білий шум;

$\Delta^d$ - оператор різниці часового ряду порядку  $d$  (послідовне визначення  $d$  раз різниць першого порядку - спочатку від часового ряду, потім від отриманих різниць першого порядку, потім від другого порядку та інші.)

Підхід ARIMA до часових рядах полягає в тому, що в першу чергу оцінюється стаціонарність ряду. Різними тестами виявляються наявність поодиноких коренів і порядок інтегрованості часового ряду (зазвичай обмежуються першим або другим порядком). Далі при необхідності (якщо порядок інтегрованості більше нуля) ряд перетворюється визначенням різниці відповідного порядку і вже для перетвореної моделі будується деяка ARMA-модель, оскільки передбачається, що отриманий процес є стаціонарним, на відміну від вихідного нестаціонарного процесу (різницево-стаціонарного або інтегрованого процесу порядку  $d$ ).

Аналіз сингулярного спектру (SSA) - це метод аналізу часових рядів, заснований на перетворенні одновимірний часового ряду в багатовимірний ряд з подальшим застосуванням до отриманого багатомірний часового ряду методу головних компонент.

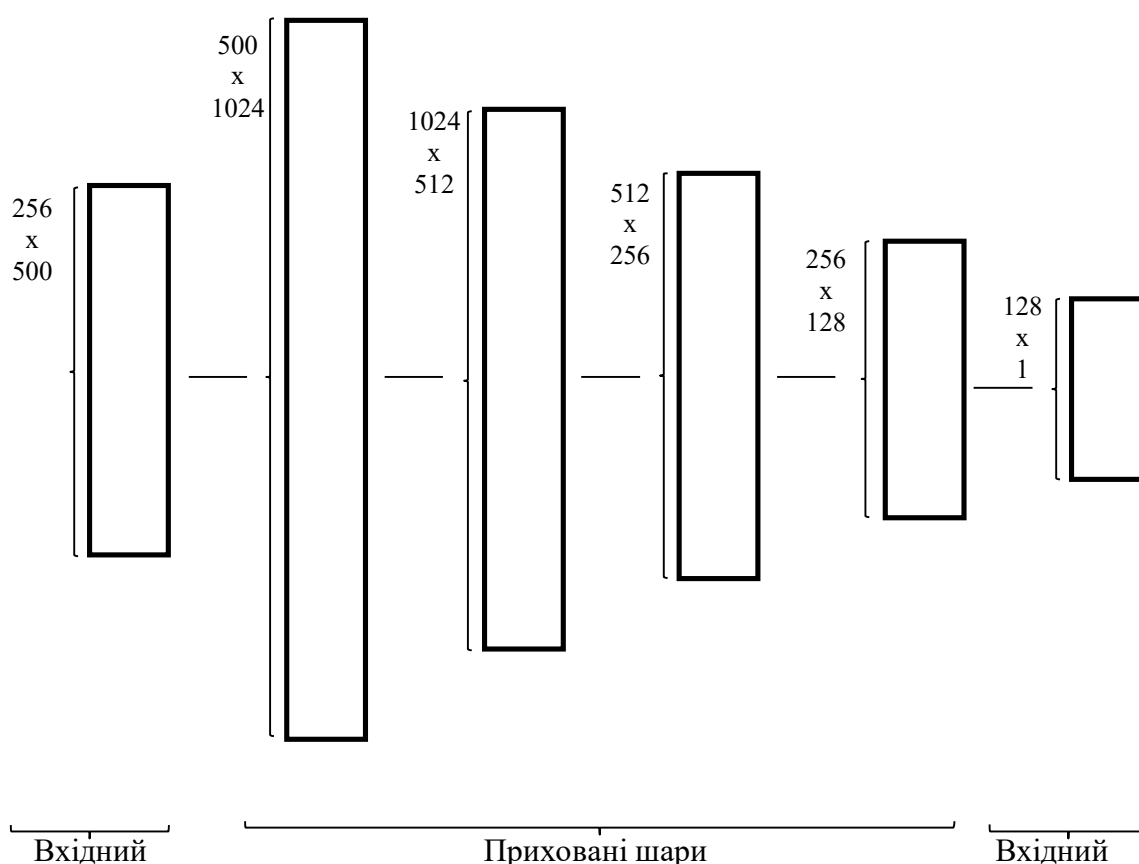
Спосіб перетворення одновимірний ряду в багатовимірний є згорткою часового ряду в матрицю, що містить фрагменти часового ряду, отримані з деяким зсувом.

SSA може бути використаний без попереднього завдання моделі ряду для аналізу довільних, в тому числі, нестаціонарних, рядів. Основна мета SSA - розкласти ряд в суму інтерпретованих компонент, таких як тренд, періодичні компоненти, шум. При цьому знання параметричної форми цих компонент не вимагається.

Нейронна мережа - це мережа простих елементів, що називаються нейронами, які отримують вхід, змінюють свій внутрішній стан (збудження) відповідно до цього входу, і виробляють вихід, залежний від входу та збудження. Мережа утворюється з'єднанням виходів певних нейронів зі входами інших нейронів з утворенням орієнтованого зваженого графу. Ваги, як і функції, що обчислюють

збудження, можуть змінюватися процесом, що називається навчанням, який керується правилом навчання.

Нейромережеві моделі можна розглядати як прості математичні моделі, що визначають функцію  $f: X \rightarrow Y$ , або розподіл над  $X$ , або над  $X$  та  $Y$ . Іноді моделі тісно пов'язують з певним правилом навчання. В зв'язку з величезною кількістю можливих архітектур нейронних мереж, будемо розглядати мережу з архітурою [5], зображеною на рис. 1.



**Рис. 1. Архітектура нейронної мережі глибокого навчання**

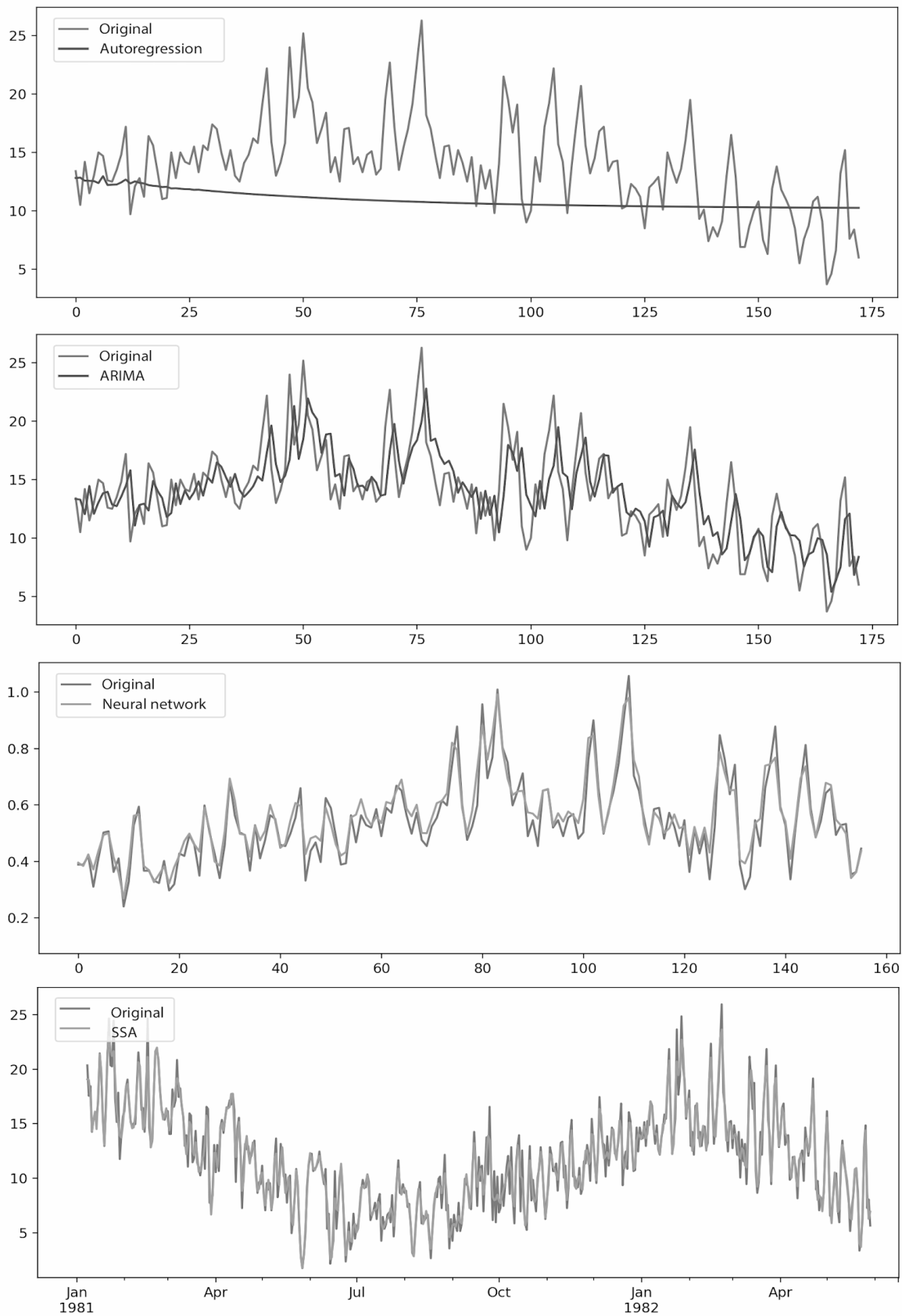
Застосувавши моделі до обраних часових рядів, отримали значення середньоквадратичних помилок, що наведені у таблиці 1.

Таблиця. 1

## Середньоквадратичні похибки прогнозів моделей

Часовий ряд	Кількість спостережень	Модель	Середньоквадратична похибка
Синусоїда	720	Ауторегресія	2.342
		ARIMA	0.021
		Аналіз сингулярного спектру	0.009
		Нейромережа	0.002
кількість пасажирів авіалінії	145	Ауторегресія	4342.589
		ARIMA	29.087
		Аналіз сингулярного спектру	7.513
		Нейромережа	0.029
денна мінімальна температура	3648	Ауторегресія	23.075
		ARIMA	8.212
		Аналіз сингулярного спектру	1.139
		Нейромережа	0.005
ринковий індекс S&P 500	197500	Ауторегресія	916.120
		ARIMA	732.21
		Аналіз сингулярного спектру	0.041
		Нейромережа	0.037





**Рис 2. Результати застосування моделей до числового ряду значень показників температури за 1981-1982 роки.**

Порівняно велика точність нейромережі пов'язана з її архітектурою, проте, це також може бути пов'язано з перенавчанням, що потребує додаткового аналізу. Найбільшу похибку має ряд ринкового індексу S&P 500.

Найбільш цікавими є нестационарні ряди, одним з прикладів яких є денна температура. Графічні результати прогнозування наведено на рис. 2. Різні показники шкал пов'язані з предобробкою (масштабуванням) вибірок.

**Результати порівняльного аналізу.** Для нестационарних рядів застосування авторегресії немає сенсу через над велику похибку. Застосування моделі ARIMA має значно меншу похибку, проте, як видно з графіку, таке передбачення повторює початковий графік з певним запізненням, і відповідно не має сенсу. Найбільш доцільним є застосування методу сингулярного спектрального аналізу та методів з використанням нейромереж. Прогнозування ринкового індексу має найбільшу похибку незалежно від методу у порівнянні з іншими типами даних, що свідчить про складність даної предметної області та необхідність подальших досліджень.

**Висновки.** В роботі розглянуто метод авторегресії, інтегрованої моделі авторегресії ковзаючого середнього (ARIMA), метод аналізу сингулярного спектру (SSA) та метод з використанням нейромереж на основі глибокого навчання для прогнозування часових рядів. В результаті аналізу результатів застосування математичних методів прогнозування різних видів часових рядів визначена їх ефективність, похибки та доцільність їх використання.

***Література:***

1. *5 Statistical Methods For Forecasting Quantitative Time Series* [Електронний ресурс] // Режим доступу: <https://www.bistasolutions.com/resources/blogs/5-statistical-methods-for-forecasting-quantitative-time-series/>
2. *Daily minimum temperatures in Melbourne, Australia, 1981-1990* [Електронний ресурс] // Режим доступу: <https://datamarket.com/data/set/2324/daily-minimum-temperatures-in-melbourne-australia-1981-1990#!ds=2324&display=line>
3. Дані ринкового індексу S&P 500 [Електронний ресурс] // Режим доступу: <http://files.statworx.com/sp500.zip>
4. *Air Passengers per month. Workshop dataset* [Електронний ресурс] // Режим доступу: <https://www.kaggle.com/rakannimer/air-passengers/data>
5. *A simple deep learning model for stock price prediction using TensorFlow* [Електронний ресурс] // Режим доступу: <https://medium.com/mlreview/a-simple-deep-learning-model-for-stock-price-prediction-using-tensorflow-30505541d877>
6. *Нові методи аналізу та прогнозування часових рядів на фінансових ринках* [Електронний ресурс] // Режим доступу: [http://www.rusnauka.com/13\\_EISN\\_2013/Economics/4\\_136384.doc.htm](http://www.rusnauka.com/13_EISN_2013/Economics/4_136384.doc.htm)
7. *Аналіз часових рядів і прогнозування* / Г. С. Кільдишев // А. А. Френкель - 1973. - 101 с.
8. *Загальна теорія статистики* / Єлісеєва І. І. // Юзбашев М. М. - 2002.- 480 с.

**References:**

1. *5 Statistical Methods For Forecasting Quantitative Time Series* // Accessibility: <https://www.bistasolutions.com/resources/blogs/5-statistical-methods-for-forecasting-quantitative-time-series/>
2. *Daily minimum temperatures in Melbourne, Australia, 1981-1990* [Electronic resource] // Accessibility:

<https://datamarket.com/data/set/2324/daily-minimum-temperatures-in-melbourne-australia-1981-1990#!Ds=2324&display=line>

3. Market index data S & P 500 [Electronic resource] // Access mode: <http://files.statworx.com/sp500.zip>

4. Air Passengers per month. Workshop dataset [Electronic resource] // Access mode: <https://www.kaggle.com/rakannimer/air-passengers/data>

5. A simple learning model for stock price prediction using TensorFlow [Electronic resource] // Access mode: <https://medium.com/mlreview/a-simple-deep-learning-model-for-stock-price-prediction-using-tensorflow-30505541d877>

6. New methods of analysis and forecasting of time series in the financial markets [Electronic resource] // Access mode: [http://www.rusnauka.com/13\\_EISN\\_2013/Economics/4\\_136384.doc.htm](http://www.rusnauka.com/13_EISN_2013/Economics/4_136384.doc.htm)

7. Analysis of time series and forecasting / G. S. Kildishev // A. A. Frenkel - 1973. - 101 p.

8. General theory of statistics / Yeliseyev I. I. // Yuzbashev M. M.-2002-480 p.