http://doi.org/10.35784/iapgos.3652

received: 15.05.2023 | revised: 31.08.2023 | accepted: 24.09.2023 | available online: 30.09.2023

COMPARISON OF THE EFFECTIVENESS OF TIME SERIES ANALYSIS METHODS: SMA, WMA, EMA, EWMA, AND KALMAN FILTER FOR DATA ANALYSIS

Volodymyr Lotysh, Larysa Gumeniuk, Pavlo Humeniuk

Lutsk National Technical University, Faculty of Computer and Information Technologies, Department of Automation and Computer-Integrated Technologies, Lutsk, Ukraine

Abstract. *In time series analysis, signal processing, and financial analysis, simple moving average (SMA), weighted moving average (WMA), exponential moving average (EMA), exponential weighted moving average (EWMA), and Kalman filter are widely used methods. Each method has its own strengths and weaknesses, and the choice of method depends on the specific application and data characteristics. It is important for researchers and practitioners to understand the properties and limitations of these methods in order to make informed decisions when analyzing time series data. This study investigates the effectiveness of time series analysis methods using data modeled with a known exponential function with overlaid random noise. This approach allows for control of the underlying trend in the data while introducing the variability characteristic of real-world data. The relationships were written using scripts for the construction of dependencies, and graphical interpretation of the results is provided.*

Keywords: data analysis, modeling, moving average, Kalman filter

PORÓWNANIE SKUTECZNOŚCI METOD ANALIZY SZEREGÓW CZASOWYCH: SMA, WMA, EMA, EWMA I FILTR KALMANA DO ANALIZY DANYCH

Streszczenie. *W analizie szeregów czasowych, przetwarzaniu sygnałów i analizie finansowej szeroko stosowane są: prosta średnia ruchoma (SMA), ważona średnia ruchoma (WMA), wykładnicza średnia ruchoma (EMA), wykładniczo-ważona średnia ruchoma (EWMA) i filtr Kalmana. Każda z metod ma swoje mocne i słabe strony, a wybór metody zależy od konkretnego zastosowania i charakterystyki danych. Dla badaczy i praktyków ważne jest zrozumienie właściwości i ograniczeń tych metod w celu podejmowania świadomych decyzji podczas analizy danych szeregów czasowych. W niniejszej pracy zbadano skuteczność metod analizy szeregów czasowych z wykorzystaniem danych modelowanych znaną funkcją wykładniczą z nałożonym szumem losowym. Takie podejście pozwala na kontrolowanie głównego trendu w danych przy jednoczesnym wprowadzeniu zmienności typowej dla danych rzeczywistych. Do budowy zależności zostały napisane skrypty. Podana jest graficzna interpretacja wyników.*

Słowa kluczowe: analiza danych, modelowanie, średnia ruchoma, filtr Kalmana

Introduction

Time series analysis is an important tool in many scientific fields, from finance to biology. Accurately modeling and forecasting time series data can be a challenging task, and to solve this problem, numerous statistical methods have been developed. Among the most common methods are Simple Moving Average (SMA), Weighted Moving Average (WMA), Exponential Moving Average (EMA), Exponential Weighted Moving Average (EWMA), and Kalman filter.

SMA is the basic method of smoothing time series data by calculating the average of a fixed number of previous values.

WMA extends the basic idea of SMA by assigning weights to previous values, with the most weight usually given to the most recent values.

EMA is a more complex method that exponentially weighs previous values, with the most weight given to the most recent values.

EWMA is a variant of EMA that allows for different smoothing parameters for different time periods.

The Kalman filter is a powerful tool that can be used to estimate the state of a system based on noisy measurements. It can be applied to time series data by treating the underlying system as a state space model and using the filter to estimate unobserved state variables.

Each of these methods has its strengths and weaknesses, and the choice of method depends on the specific problem facing the researcher. In this article, we will compare these methods by investigating their effectiveness on a given dataset.

The Simple Moving Average (SMA) is discussed in [5]. The authors note that SMA is a simple form of local polynomial regression. SMA is often used to smooth short-term fluctuations in time series data and to identify long-term trends. However, the authors note that SMA has limitations and may not be suitable for all time series data. Additionally, SMA does not take into account data variability, which can lead to erroneous results. Overall, although SMA is a simple and widely used method of smoothing time series data, it is important to consider its limitations and choose the appropriate method based on the characteristics of the data.

The Weighted Moving Average (WMA) is discussed in [3]. The authors present several methods for smoothing time series data, including Simple Moving Average (SMA), Exponential Smoothing, and Weighted Moving Average (WMA). Hyndman and Athanasopoulos provide examples of calculating weighted moving average using the R programming language and discuss the advantages and limitations of this smoothing method. They also compare WMA with other smoothing methods and provide recommendations for choosing an appropriate method based on the characteristics of the time series data and forecasting goals.

Exponential Moving Average (EMA) is also discussed in [3]. The authors describe EMA as a popular method for smoothing time series data and making short-term forecasts. They also describe the formula for calculating EMA, which includes specifying a smoothing parameter, often denoted as alpha, which determines the weight given to the most recent observation. They explain how to choose an appropriate alpha value based on the characteristics of the analyzed data and provide examples of applying the EMA method to different types of time series data. Overall, [3] is a comprehensive introduction to the EMA method and its application for time series forecasting.

The entire article [2] is dedicated to exponential smoothing, which is a method of smoothing time series data. The article discusses the basic idea of exponential smoothing and its various extensions, including the Exponentially Weighted Moving Average (EWMA).

In particular, the article explains that EWMA is a method of exponential weighting of past observations in computing the smoothed value of a time series. The weight assigned to each observation is based on the smoothing parameter, which determines the rate at which the weight decreases with increasing observation time. The article also compares the effectiveness of EWMA with other methods of smoothing time series data and discusses some practical issues related to implementing the method.

The Kalman filter is discussed in several sections of [1]. The authors present the Kalman filter as a method for estimating the state of a linear dynamic system based on noisy measurements. They also discuss the assumptions and limitations of the Kalman filter and provide several examples of its use in time series analysis and forecasting. It is noted that the Kalman filter is a key component of state space models, as it allows for the estimation of latent variables in the presence of noisy observations. The authors discuss how to set up state space models for time series analysis and provide several examples of models that can be estimated using the Kalman filter. Overall, [1] provides a thorough introduction to the Kalman filter and its use in time series analysis and forecasting.

It is also worth noting classic publications dedicated to economic analysis, which consider time series [4, 6, 7].

Methods such as Simple Moving Average (SMA), Weighted Moving Average (WMA), Exponential Moving Average (EMA), Exponentially Weighted Moving Average (EWMA), and the Kalman filter are widely used in time series analysis, signal processing, and financial analysis.

Each method has its strengths and weaknesses, and the choice of method depends on the specific application and data characteristics.

1. Problem statement

Conduct a study of the effectiveness of time series analysis methods using a known exponential function with added random noise. Mean Squared Error (MSE) will be used as the effectiveness criterion.

2. Main results

Time series analysis is a powerful tool for analyzing data that changes over time. One of the most common applications of time series analysis is in finance, where it is used to detect trends and patterns in financial data and to forecast future values.

One way to study the effectiveness of time series analysis methods is by modeling data using a known exponential function with added random noise. This approach allows the main trend in the data to be controlled while introducing the variability characteristic of real data.

To model data using a known exponential function with random noise, the following formula can be used:

$$y(t) = A \cdot e^{(-k \cdot t)} + \varepsilon(t) \quad (1)$$

where A and k are constants that determine the shape of the exponential curve; t is the time index; $\varepsilon(t)$ is a random variable that represents the random noise.

The exponential function (1) at $A = 100$ and $k = 0.00098$ and the random noise imposed on it are shown in Fig. 1.

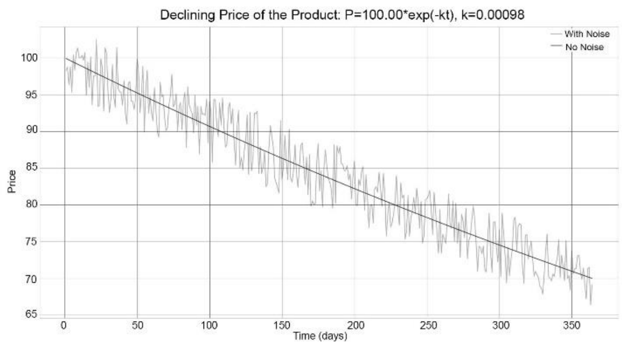

Fig. 1. The exponential function (1) at A = 100 and k = 0.00098 and the random noise imposed on it

Simple Moving Average (SMA) is a commonly used technique for analyzing time series data that helps to identify trends in a dataset over a specific period of time. It is a type of moving average that calculates the average value of a set of data points over a certain time period and is useful for smoothing out short-term fluctuations in the data.

$$SMA = \frac{\sum_{i=1}^{n} D_i}{p} \quad (2)$$

where *Di* is a data set; *p* is the length of smoothing or the period of the SMA (the number of values included in the calculation of the moving average).

To calculate the SMA, the sum of a set of data points over a certain period of time is divided by the number of data points in that period. For example, if we are interested in a 10-day SMA, we sum up the data point values for the last 10 days and divide that sum by 10.

SMA can be used to detect trends in a data set over time. When the SMA increases over time, it indicates that the underlying data points are also increasing. Conversely, when the SMA decreases over time, it indicates that the underlying data points are decreasing.

The Simple Moving Average can be configured for different time horizons depending on the specific analysis.

One of the drawbacks of SMA is that it gives equal weight to all data points in the moving average, regardless of how fresh they are. This can lead to a lag in detecting changes in the trend, as SMA may take some time to adapt to new information.

Also, the Simple Moving Average is not ideal for time series data with irregular structure. The smoothing effect of the Moving Average can hide important data features, such as sharp changes or extreme values..

SMA is less effective for long-term trends, as it can overestimate short-term fluctuations and mask the main long-term trend.

The graphical interpretation of the results of the SMA algorithm is shown in Fig. 2.

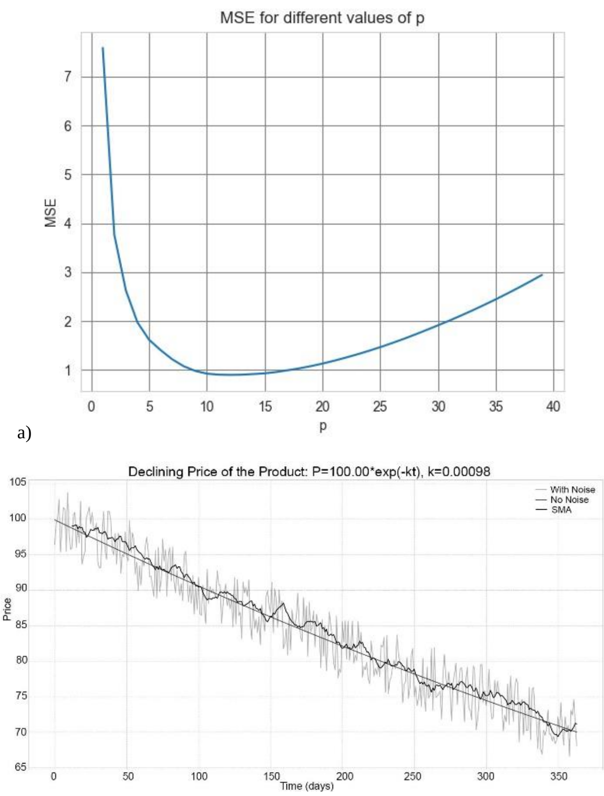

Fig. 2. a) Dependence of the root-mean-square error on the period of SMA; b) SMA at the minimum value of the root-mean-square error (MSE): p = 12; MSE = 0.90

Weighted Moving Average (WMA) is a method of analyzing time series data that is used to analyze and forecast trends in a series of data over time.

It works by computing the average value of a set of data points over a certain period, where each data point is given a weight based on its importance or relevance to the analysis. Weights are usually assigned in such a way that new data is given greater weight than old data, reflecting the fact that newer data is generally more relevant for predicting future trends.

$$WMA = \frac{\sum_{i=1}^{n} D_i \cdot p_i}{\sum_{i=1}^{n} p_i} \quad (3)$$

where *Di* represents the data for the *i*-period (with *i = 1* being the current period); *pi* represents the weight values for the *i*-periods.

Weighted Moving Average (WMA) gives more weight to recent data points, making it more sensitive to changes in data than Simple Moving Average. This allows for better trend detection and forecasting of future values.

WMA can be customized by changing the weight of each data point in the moving average, providing greater flexibility in analysis.

WMA can be useful for smoothing time-series data with irregular structure or volatility, as weights can be adjusted to emphasize certain parts of the data.

However, WMA is prone to lag in detecting changes in trend, as the weights given to recent data points may not be enough to immediately incorporate new information.

The graphical interpretation of the results of the WMA algorithm is shown in Fig. 3.

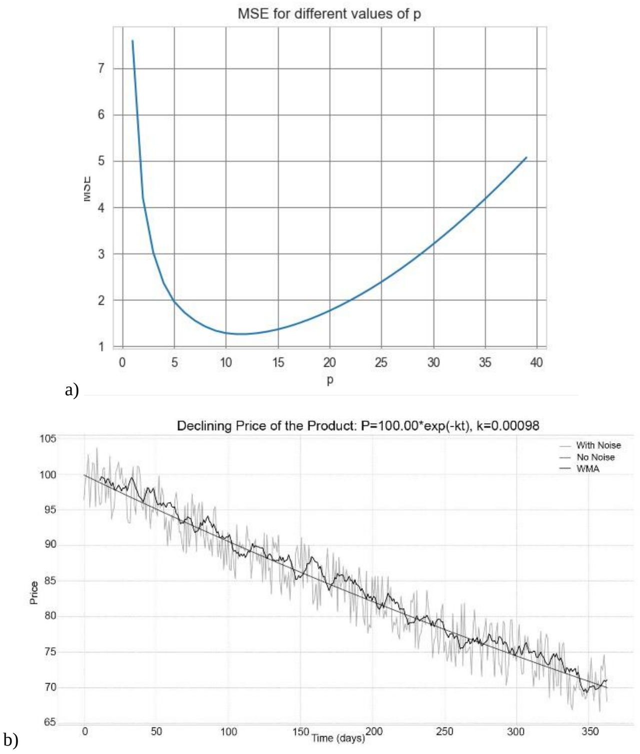

Fig. 3. a) Dependence of the root-mean-square error on weight values; b) WMA at the minimum value of the root-mean-square error (MSE): p = 11; MSE = 1.26

Exponential Moving Average (EMA) is a widely used method for analyzing time series data and predicting trends over time.

Like other Moving Average methods, EMA involves calculating the average value of a set of data points over a certain period. However, EMA gives more weight to the latest data points, using a weighting coefficient that exponentially decreases. This means that EMA responds better to recent changes in the data series and can provide a more accurate forecast of future trends.

$$EMA = \frac{EMA_{i-1} \cdot (p-1) + 2 \cdot D_i}{p+1} \quad (4)$$

where D_i is the value in the i-th period; p is the calculation period; EMA_{i-1} is the value of the previous period's EMA.

Exponential Moving Average (EMA) reacts better to changes in data than Simple Moving Average (SMA) and Weighted Moving Average (WMA). This is because EMA gives more weight to the latest data points and the weight decreases exponentially as the data points become older.

EMA can be customized by adjusting the smoothing factor, providing greater flexibility in analysis.

EMA is less prone to lag than SMA and WMA because it responds more quickly to changes in data.

However, EMA may be more sensitive to outliers than SMA and WMA because outliers have a greater impact on EMA than on other averages.

EMA is still subject to lag when detecting changes in trends because the smoothing factor may not be sufficient to immediately incorporate new information.

The graphic interpretation of the results of the EMA algorithm is shown in Fig. 4.

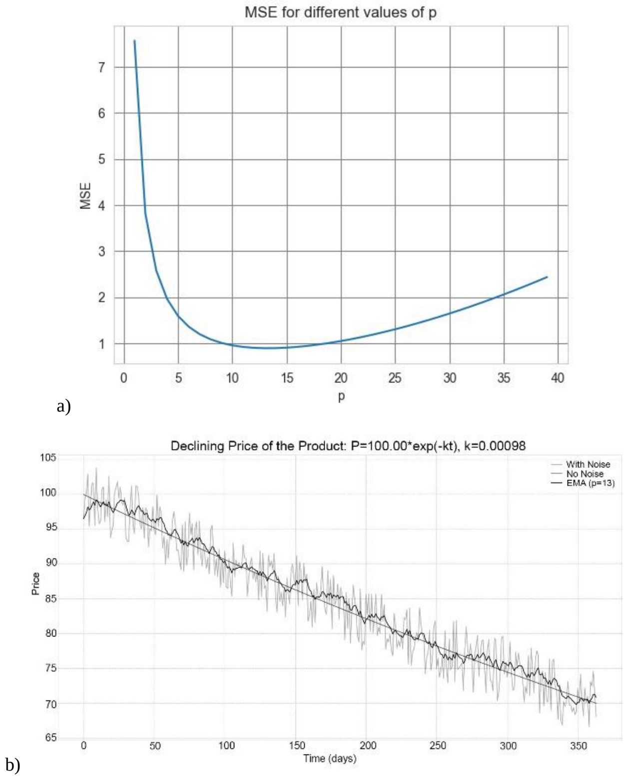

Fig. 4. a) Dependence of the root-mean-square error on the smoothing coefficient value; b) EMA at the minimum value of the root-mean-square error (MSE): p = 13; MSE = 0.89

Exponentially Weighted Moving Average (EWMA) is a time series analysis method used to analyze and forecast trends in data series over time.

EWMA is a variation of the traditional moving average method that gives more weight to recent data points than earlier ones, using a weight coefficient that exponentially decreases. The weight coefficient is usually a smoothing parameter chosen based on the characteristics of the analyzed data series.

$$EWMA = p \cdot r_t + (1-p) \cdot EWMA_{t-1} \quad (5)$$

where p is the weight coefficient; r is the data value in the current period.

EWMA provides greater flexibility in analysis than simple moving average and weighted moving average because it can be adjusted by changing the smoothing coefficient.

However, the exponentially weighted moving average may be more sensitive to outliers than simple moving average and weighted moving average, as outliers have a greater impact on the exponential average than on other averages.

A graphical interpretation of the results of the EWMA algorithm is shown in Fig. 5.

The Kalman filter is a mathematical algorithm used to estimate the state of a system based on a series of measurements. It is widely used in many fields, including finance, engineering, and science, to analyze numerical series and extract significant information from them.

In essence, the Kalman filter works by combining the forecast of the system state based on previous measurements with a new measurement of the system to obtain an updated estimate of the system state. The filter is designed to take into account any uncertainties or errors in measurements, as well as any noise or other sources of variability in the system.

The Kalman filter is particularly useful for analyzing time series data, as it can be used to estimate the underlying trend or pattern in the data, even if the data is noisy or contains outliers.

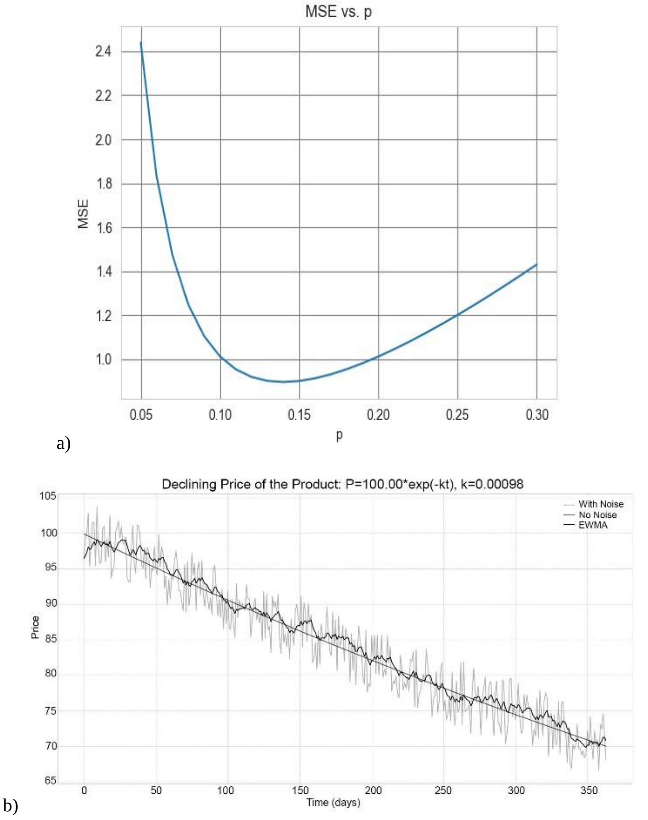

Fig. 5. a) Dependence of the root-mean-square error on the value of the weight coefficient; b) EWMA at the minimum value of the root-mean-square error (MSE): p = 0.14; MSE = 0.89

However, the Kalman filter can require significant computation, especially for large datasets or complex models. The Kalman filter relies on assumptions about the statistical properties of time series data, such as the distribution of noise, which may not always be accurate.

A graphical interpretation of the results of the algorithm using the Kalman filter is shown in Fig. 6.

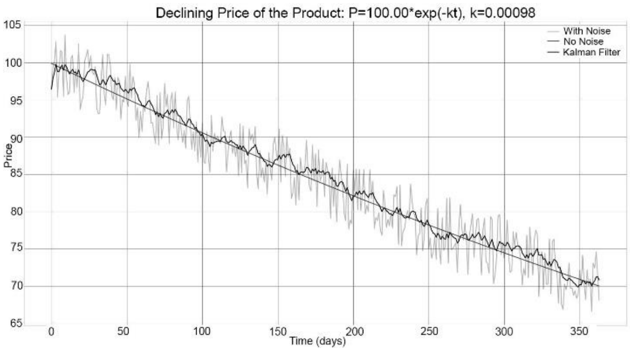

Fig. 6. The results of the Kalman filter at the minimum value of the root-mean-square error (MSE): a = 0.009, b = 0.418, c = 1.000, MSE = 0.8502

To build the dependencies, Python 3.10 scripts were written. For graphical interpretation of the results, the Matplotlib library was used, along with NumPy for numerical operations.

3. Conclusion

Thus, methods such as Simple Moving Average (SMA), Weighted Moving Average (WMA), Exponential Moving Average (EMA), Exponentially Weighted Moving Average (EWMA), and Kalman filter are widely used in time series analysis, signal processing, and financial analysis. Each method has its strengths and weaknesses, and the choice of method depends on the specific application and characteristics of the data. It is important for researchers and practitioners to understand the properties and limitations of these methods to make informed decisions when analyzing time series data.

References

[1] Brockwell P. J., Davis R. A.: Introduction to time series and forecasting. Springer, 2016.
[2] Gardner E. S.: Exponential smoothing: The state of the art. Journal of forecasting 4(1), 1985, 1–28.
[3] Hyndman R. J., Athanasopoulos G.: Forecasting: principles and practice. Otexts, 2018.
[4] Lewis C. D.: Industrial and business forecasting methods. Butterworth, Kent 1982.
[5] Ruppert D., Wand M. P., Carroll R. J.: Semiparametric regression. Cambridge University Press, 2018.
[6] Terry J. W., Parramore K.: Quantitative Methods for Finance. International Thomson Business Press, 1997.
[7] Wynn R. F., Holden K.: An introduction to applied econometric analysis. Macmillan, London 1974.

Ph.D. Volodymyr Lotysh
e-mail: auvp@lntu.edu.ua

Lutsk National Technical University, PhD. (technical). Research interests: open-source software applied for simulations of problems using distributed platforms. Author of nearly 80 publications in this research.

http://orcid.org/0000-0003-0899-8015

Ph.D. Larysa Gumeniuk
e-mail: l.gumeniuk@lntu.edu.ua

Lutsk National Technical University, PhD. (technical), Department of Automation and Computer-Integrated Technologies.
Research interests: modeling of reliability and safety of the automated control systems.
Has more than 60 publications in this area.

http://orcid.org/0000-0002-7678-7060

Ph.D. Pavlo Humeniuk
e-mail: l.gumeniuk@lntu.edu.ua

Lutsk National Technical University, PhD. (technical), Department of Automation and Computer-Integrated Technologies.
Research interests: programming, robotics.

http://orcid.org/0000-0002-6251-8548