AL KUT JOURNAL OF ECONOMIC AND ADMINISTRATIVE SCIENCES
Publisher: College of Economics and Management - Wasit University

Hybrid Forecasting of Financial Time Series Using Adaptive Regression Splines and Exponential Smoothing

Lamiaa Abdul-Jabbar Dawod[1] , Balsam Mustafa Shafeeq[2], MUSTAFA ABDULJABBAR DAWOOD[3],

Waleed Ahmed Hassen Al-Nuaami[4]

[1]Middle technical university /College of Health and medical Techniques/ Department of Optics Techniques, Iraq

lamya.abd@mtu.edu.iq

[2] Technical College of Management –Baghdad, Middle Technical University, Iraq

balsammustafa95@mtu.edu.iq

[3] University of Baghdad / college of administration and economics/Statistics Department,

mostafa.abduljabar1201a@coadec.uobaghdad.edu.iq

[4] Department of Biology, College of Education for Pure Sciences, University of Diyala,

purecomp.waleed.hassan@uodiyala.edu.iq

Abstract:

The purpose of forecasting models premised on time series (TS) is to generate predictive models that provide a closer approximation of future data, with an acceptable error. A proposed framework is developed within the framework of this study as a hybrid model to improve the forecasting accuracy of the traditional TS models based on the Multivariate Adaptive Regression Splines (MARSplines) technique. The integration targets to take advantage of the flexibility and adaptation of MARSplines in the TS. In order to compare the performance of the suggested model, historical data over a six-year period of the Borsa Istanbul were used as a case study. The results indicate that the combined regression-based method gives a significant increase in the predictive capabilities and practical importance of the time series model.

Keywords: Hybrid Forecasting Model, Financial Time Series Prediction, MARSplines, Exponential Smoothing, Nonlinear Time Series Analysis.

1. Introduction:

Stock exchange markets (SEMs) are used in a wide range of countries and constitute principal platforms of investment efforts and observation of the price variations in sectors. As a rule, such markets are fair access systems which distribute information on trading in real-time worldwide financial networks. The SEMs operate depending on their local time zones and during the timeframes, prices of assets might change massively as a result of the high turnover during this period. The value that is gained earlier during the start of a trading tide is known as the opening value and the last one at the close of a market is known as the end-of-day value (Bundoo, 2011). The subsequent time series of end-of-day values on different time frames (daily, weekly, monthly or yearly, etc.) one can track and assess the time-related trends and volatility in the market developments.

The values of the stock markets are prone to continuous changes during the trading day because of a constant purchase and sale process, or on a larger scale political and social happening (Goonatilake & Herath, 2007). Therefore, stock exchange data tends to undergo volatile fluctuations that are the results of such external and non-economic factors (Rai et al., 2023). The ability to tell the emergent trends, making proper timing of investment decisions and forecast market reactions are some of such competences required by an investor who wants to maximize his or her decisions. Nonetheless, such a modeling of such complex dynamics is quite problematic, owing to inherently non-parametric and noisy financial time series data (Bhowmik, 2013). This fluidity is an indicator of the need to have more advanced modeling structures which supplement the conventional structures with state-of-the-art advanced adaptive equation sets that can capture such inconsistencies (Aggarwal et al., 1999). The basis of Time Series (TS) forecasting models rests this assumption, which states that there is continuity of underlying patterns or trends observed in past to the present and future. Put in simple terms, it follows that the historical records contain meaningful information that can be used to predict the future behavior to some given extent (Chowdhury et al., 2016). The algorithms of time series commonly used in data mining are especially useful when applied to temporal data in order to help discover the statistical relationships of this kind and ultimately make good forecasts on the basis of historical patterns (Ghous et al., 2023).

This paper presents a new time scale forecasting algorithm which conceptually models data distributions in time. The practical aspect of the experiment is dedicated to the continuous scale work on the publicly available market data. This study has two main contributions to the literature of the study as follows: (i) As seen in recent years, numerous highly accurate analysis models were published; however, most of these methods are relying on the use of very complex machine learning approaches and require significant amounts of intervention on part of the person developing the model. This has the effect that such models are frequently built-up merely by trial-and-error. By contrast, high levels of predictive accuracy are attained with minimal exposure to researcher intervention or the amount training data needed using the proposed technique. (ii) In addition, the model supports an analysis of datasets in different scopes, regardless of their dimensionality or size, without the need to have a strong knowledge of machine learning. The ability to make good predictions of future data is especially vital in the field of finance where informed and timely decision-making relies mostly on predictions. TS models and other statistical methods that use machine learning can be used as a solid method of projecting future trends using past statistics (Wirawan, 2023). This paper takes into consideration and studies six years of market data of Borsa Istanbul. After some experimental tests, a hybrid algorithm was developed with the objective of making an improvement in the forecasting accuracy of the standard TS models and helping overcome the drawbacks experienced in the use of traditional predictive methods.

2. Research Background:

Multivariate Adaptive Regression Splines (MARSplines) is an adaptive regression spline data mining technique, known to be very flexible and computationally efficient in modeling either continuous or binary categorical data (Friedman, 1991). Under this method, piecewise linear regression models are built and reflect real life non-linear relations in data. Among the main advantages of the MARSplines method, it is possible to distinguish the fact that the method enables to select and use only the most meaningful set of predictor variables, which simplifies the model and excludes the quantification of interactions between redundant variables (Hastie et al., 2009). When it comes to an economic environment, the performance of the expert projections and overall market expectations with the realities played a significant role in following the decisions of the investors (Fama, 1970). In where case, the capability to make high accuracy prediction concerning the future behavior of the market is of high relevance towards determining the preferences of direct investments. In order to meet this requirement, a great number of statistical forecasting models have been designed to assist in facilitating the business world in the economics and finance domains (Tsay, 2005). Time-Series based predictive models particularly are very good tools among them. In one of the studies (Rahmon & Samson, 2023) used exponential smoothing methods namely Holt double and Holt Winters triple to predict weekly closing prices of creative industry firms which were listed on the. The study also evaluated how well this method can predict with a hybrid of regression modeling technique adopted in the study and its comparison with the traditional time series techniques.

Similarly,(Zhang, 2003) developed a hybrid forecasting model that combined TS methods with artificial neural networks (ANNs) to develop forecasts of stock prices of some companies in Turkey. The principle behind their strategy is to incorporate the model of TS that yields the strongest correlation with the dependent variable which is by the name stock prices into ANN model with the view of improving the precision of predictive capability. The paper provides the comparative study of various forecasting methods, and it is shown that the best success is obtained by applying neural networks to the time series elements that contain most information (Khashei & Bijari, 2011).

Kao and Chiu (2020) proposed a new chart pattern recognition model which combines Multivariate Adaptive Regression Splines (MARSplines) with recurrent neural networks (RNN) as a model called MARS-RNN model. This hybrid structure forms a very effective tool to identify the latent patterns in the process control charts and hence assist in constructing strategic decision-making structures. The model was compared in its performance to that of other options such as Random Forest (RF), and ANN as well as other hybrid compositions using MARSplines (Liaw & Wiener, 2002). It was empirically found that the suggested model outperformed (with respect to both choice of the relevant) input variables and prediction accuracy (Goodfellow et al., 2023).

Another hybrid modelling framework that (Aragão, 2024) suggest is the one that combines regression methods with the time series models. In this context, a number of regressions processes, the most known of which is the MARSplines algorithm is incorporated to standard time series modelling with the aim to enhance the accuracy of forecasts. The general focus is on the tracking of the epidemic trajectories accompanied by the forecast of the future dynamics, thus ensuring the capability to be ready in the case of a potential outbreak situation (Okumuş et al., 2019). The authors compare the predictive power of MARSplines-based hybrid model to the Back Propagation Neural Network (BPNN). According to their findings, the former always tends to make more accurate predictions as compared to the latter. BPNN demonstrates its well performance in predictive tasks of diverse autonomous activities; still, comparative exploration with hybrid models will be irreplaceable to the analytical discipline. In this regard, (Rumelhart et al., 1986) evaluated the performance, between BPNN and MARSplines, when estimating stock index valuation. It was shown that, in the course of working in a time-series model, MARSplines could achieve the highest predictive accuracies, compared with BPNN.

Time-series models have proved really useful in systems that move in a cyclical behavior and are repeated after some intervals of time (Box et al., 2015). One of the major strengths of these models is the tool kit of analytic frameworks which can be easily applied to temporal data. However, the major limitation is that they tend to under-represent the parabolic oscillations inherent in nonlinear data, especially in the signal dominant high -frequency or irregular phenomena (Tong, 1990). MARSplines is one example of explicit nonlinear regression solution that has gained popularity and credibility due to its strong capacity in terms of predictive accuracy, the trait that is intrinsic to MARS flexible and adaptative architecture. Specifically, MARSplines has already been proven to be efficient in handling large-scale data or, in other words, with a set of a great number of independent variables both in terms of computer performance and practical applicability in high-dimensional studies (Hastie et al., 2009). Another similar research in the field of meteorology was carried out to simulate the snowfall pattern in an area of Iran. There were two hybrid approaches adopted; (1) Support Vector Machines (SVM) and MARSplines and (2) Random Forest (RF) and TS. The comparison has shown that RF-TS hybrid model offered better predictive abilities than SVM-MARSplines model (Javan & Movaghari, 2022)

The MARSplines method is based on a series of linearly related basis functions which are created in successive order. The precise value of basic functions as well as the parameters within each stage (or spline) are auto calculated by the data structure (Khosravi et al., 2023). At each stage, the data is recursively subdivided into subregions, and in these subregions' regression models are formed by applying basis-function derivatives intended at approximating the behaviour of the continuous variables (Vafakhah et al., 2022).

In the sector of health, (TS) theorized models are being used on data which have a direct correlation to human health, which is a variable that demands modeling analysis noteworthy of sensitivity, as well as, precision (Youssef Ali Amer et al., 2021). The periodicity and periodicity of the most crucial physiological signals strong increases the forecast ability of TS models (Gupta et al., 2020).Conventionally, vital sign surveillance is performed at a predetermined threshold band and the automated alert modules that act on values outside the predetermined parameters are considered the normal clinical procedure (Hyland et al., 2020). Provided that the time evolution of them is queried via the TS techniques, it becomes possible to predict the future states of alarms (Baker et al., 2020). Moreover, any disturbances of therapeutic intervention or drug alterations that are made before the achievement of these critical levels can prevent the occurrence of poor clinical results. Such transition phases are usually marked by visible peaks or inflection points representing critical junctures of development in the course of a medical condition. This means that the TS modeling framework provides a useful platform of assessing the effectiveness of interventions, and of monitoring the dynamics during the course of evolution of key processes over time (Youssef Ali Amer et al., 2020).

By the end of the model constructed with the help of MARSplines algorithm, there appears visual model of the relationship structure and the levels of association between the independent variables (Hastie et al., 2009). This feature, which makes MARSplines stand out among other modeling methods, is that it has increased interpretability and high comprehensibility, particularly, when working with complex multivariable models (Adiguzel & Cengiz, 2023).

Over the last few years, Multivariate time-Series (MTS) combined regressions models have in fact been widely used in the framework of deep learning (DL)-based data-mining applications, across the fields of finance, healthcare, and others (Lim & Zohren, 2021). As an example, (Mode & Hoque, 2020) suggested a hybrid system, which combines DL approaches with MTS-based regression to maximize the predictive power of temporal patterns in the cybersecurity-related data analysis problem.

Time series forecasting (TSF) models are designed to anticipate future observations by systematically identifying and modeling recurring patterns embedded in historical data. Informed, forward looking decision-making is hardly achievable without the support of such predictive tools. (Ilić, 2017) hybrid framework enhances traditional TS analysis by incorporating a linear regression approach namely, the Error-Based Linear Regression (EBLR) scheme which refines forecasts through the integration of regression techniques. EBLR so that the accuracy of classic TS models can be improved. To be very specific, EBLR uses regression tree structures to recreate and improve the initial TS residuals hence enhancing the predictive accuracy of different iterations (Okumuş et al., 2019). This is done in two stages. Phase 1 consists of preparing base-line TS method to come up with the initial forecasts. The residuals of phase 1 are converted to a regression tree in phase 2 that provides a more accurate list of predictions .

The methodology of the study and details of the applied methods are explained in Section 3. In Section 4, the application of the stated methods and methodology on data is given. In the last section, the main findings and results obtained from the research are expressed.

3. Methodology

This paper proposes a hybrid regression model consisting of the additive exponential smoothing algorithm from time series analysis and Multivariate Adaptive Regression Splines (MARSplines). The main goal is to improve the forecasting accuracy of traditional time series models by taking advantage of flexibility and adaptability features that can be incorporated into any regression part of the model.

3.1. Adaptive Exponential Smoothing for Time-Based Market Patterns

A panel time series usually exhibits similar dynamics across well-defined, repetitive temporal intervals. This type of cyclical recurrence is referred to as the seasonal component of the model. Regardless of how often such patterns recur, the corresponding durations are regarded as distinct periods as long as they maintain chronological order. Exponential smoothing methods give smaller weights to observations from earlier periods as the time horizon extends (Ghorbani & Pamucar, 2026). In this structure, the most recent observations have the best weight in determining the forecast. The seasonal effect at a given time span under this model is given by the following formulation. The model is defined as:

$$Z_\tau = \lambda_\tau + \kappa_\tau \cdot \tau + \phi_{\tau,q} + \epsilon_\tau \tag{1}$$

Where:

Z_τ : observed financial value at time τ,

λ_τ : smoothed average level,

κ_τ: slope (trend) coefficient,

$\phi_{\tau,q}$: seasonal index for period q,

ϵ_τ : residual random noise term.

When no trend or seasonality is present, $\kappa_\tau = 0$ or $\phi_{\tau,q} = 0$, respectively.

The recursive estimation of components is reformulated as:

$$\Lambda_\tau(i) = \alpha_1 \cdot \left(Z_\tau(i) - \Phi_{\tau-q}(i)\right) + (1-\alpha_1) \cdot \left(\Lambda_{\tau-1}(i) + \Theta_{\tau-1}(i)\right) \tag{2}$$

$$\Theta_\tau(i) = \alpha_2 \cdot \left(\Lambda_\tau(i) - \Lambda_{\tau-1}(i)\right) + (1-\alpha_2) \cdot \Theta_{\tau-1}(i) \tag{3}$$

$$\Phi_\tau(i) = \alpha_3 \cdot \left(Z_\tau(i) - \Lambda_\tau(i)\right) + (1-\alpha_3) \cdot \Phi_{\tau-q}(i) \tag{4}$$

Where:

$\Lambda_\tau(i)$: level component for country i,

$\Theta_\tau(i)$: dynamic trend for time τ,

$\Phi_\tau(i)$: adjusted seasonal component,

$\alpha_1, \alpha_2, \alpha_3$: smoothing constants for level, trend, and seasonality, respectively,

q: seasonal period length (e.g., 12 for monthly data, 4 for quarterly),

$i = 1,2, \dots ,6$: index of countries in Borsa Istanbul panel.

The h step-ahead forecast is computed with:

$$\hat{Z}_{\tau+h}(i) = \Lambda_\tau(i) + h \cdot \Theta_\tau(i) + \Phi_{\tau+h-q}(i) \quad (5)$$

Where:

$\hat{Z}_{\tau+h}(i)$: predicted value at horizon $\tau + h$,

$h \cdot \Theta_\tau(i)$: cumulative trend contribution,

$\Phi_{\tau+h-q}(i)$: seasonal estimate aligned with future cycle.

3.2.Flexible Regression Modeling with Adaptive Splines

The MARSplines methodology was first introduced by (Friedman, 1991) and since then, with the growth of statistical computing, has gained significant popularity. Whereas standard regression approaches fall under the known functional forms, the MARSplines allows for modeling complex relationships between variables using a nonparametric structure. Rather than fitting one global equation, localized regression segments are based on basis functions. It proceeds in two steps; first splitting the data into balanced subsets, and second building separate regression structures for each subset by introducing spline functions into them. Eventually, these partial models are merged into one comprehensive predictive model. MARSplines is best known for its great performance on data having nonlinear distribution, where an ordinary linear model breaks down (Stanimirović et al., 2024). By way of its piecewise-defined, bidirectional functions, the method catches finesse patterning and delivers rather exact estimations of continuous target variables under conditions of high dimensionality or when there are mixed-type predictors.

$$\hat{y} = \psi(\mathbf{X}) = a_0 + \sum_{j=1}^{Q} a_j \cdot \varphi_j(\mathbf{X}) \quad (6)$$

Where:

$\hat{y}$: the predicted value of the response variable,

x: the vector of predictor (independent) variables,

a_0: the intercept term of the model,

a_j: the regression coefficient associated with the j^{th} basis function,

$\varphi_j(\mathbf{X})$: the basis functions, typically defined as piecewise linear splines,

Q: the total number of selected basis functions in the final model.

In datasets that have seasonal or cyclical structures, the basic functions are adjusted by MARSplines iteratively and dynamically over several runs of the model to reduce residual error according to a least-squares criterion. The adaptive modeling approach allows the algorithm itself to determine automatically not only which predictors are most relevant but also any interaction effects among them. MARSplines is a nonparametric method and so it is very flexible, but such flexibility may lead to overfitting; thus, a pruning mechanism is introduced for removing further simplifying the model by eliminating basis functions so as to generalize the model better and hence making MARSplines an appropriate robust variable selection methodology (Aksoy et al., 2022) .The MARSplines variable selection process is a two-stage stepwise approach that starts with a constant term. During the first stage, or forward pass, all basis functions search among possible spaces to find those combinations of predictors and knot locations which reduce the prediction error; this goes on until model fit reaches some pre-set level. In the second stage or backward pass, it drops those basis functions which are not contributing significantly toward improving goodness of fit to make the model parsimonious (Al-Nuaami et al., 2024).

3.3.Merging Exponential Smoothing with MARSplines Techniques

The hybrid modeling structure uses a two-stage integration in such a way that predictors adjusted by exponential smoothing are then included in the MARSplines regression model. At this stage, every predictor is updated for every run based on the time-series mechanism of exponential smoothing. Firstly, major parameters concerning data smoothings are identified as level (α), trend (γ), and seasonality (δ). These values guide the derivation of corresponding smoothed components level L_t, slope T_t, and seasonal index S_t. These smoothed values are then mapped onto the components of the traditional time-series notation: μ_t is approximated by L_t, the slope term β_t by T_t,, and the seasonal term $S_{t,p}$ by S_t. The resulting formulation is similar to the formulation of seasonal exponential smoothing, which was described above. This iteration process of update mechanism is applied to all observations in the dataset. Once the time-series estimations y_t are established for all time points, these values substitute the original predictor matrix X in the MARSplines equation (Stanimirović et al., 2024). The hybrid function at this stage becomes:

$$y = f(y_t) = B_0 + \sum_{t=1}^{T} \sum_{i=1}^{S} B_i \phi_i(y_t) \qquad (7)$$

Where:

- B_0 denotes the constant term,
- $\phi_i(y_t)$ represents the reformulated basis functions,
- B_i are the corresponding coefficients,
- y_t are the input values derived from the smoothed time-series estimations,
- The nested summation iterates over both time points T and the number of basic functions S.

Following this, the extended mathematical representation of the integrated model considering all components derived from smoothing (trend, level, seasonal, and residual error) can be written as:

$$y = f(y_t) = B_0 + \sum_{t=1}^{T} \sum_{i=1}^{S} B_i \cdot \phi_i\left(\hat{\mu}_t + \hat{\beta}_t t + \hat{S}_{t,p} + \hat{a}_t\right) \qquad (8)$$

Where:

- $\phi_i(\cdot)$ denotes the adapted basis functions of the MARSplines model,
- $\hat{\mu}_t, \hat{\beta}_t, \hat{S}_{t,p}$ and $\hat{a}_t$ represent the smoothed estimators of mean level, trend, seasonal component, and noise, respectively, derived from the exponential smoothing time-series algorithm.

- The structure reflects a layered model in which each regressor y_t input is refined through the smoothing-based estimation before entering the spline-based learning structure.

The data used in this analysis were downloaded from Kaggle.com, an open-source website with very rich statistical datasets and analytical tools. This particular dataset is under the title 'Uniqlo Stock Price' and contains all trading-related information for Uniqlo shares traded at the Borsa Istanbul on a daily basis: opening price, closing price, maximum and minimum transaction values within a day, volume of transactions, and total daily monetary value of share transactions included as variable stock trading. This variable shows the aggregate financial value of buying and selling Uniqlo shares that take place within one trading day. Also, the Date is the only categorical feature that shows the exact trading date of each observation. The period covered in this data set extends from January 2, 2019 through January 10, 2024. There are a total of 1,233 observations within it wherein the first 1,226 records have been used to train the model and the last seven have been kept aside to validate as well as test its performance. The dataset falls under the creative commons zero (CC0 1.0 Universal) license category hence is freely usable for academic as well as analytical purposes. To better assess the model's predictive ability, data with volatility and a tendency toward odd swings was purposely picked rather than data that shows steady linear or arithmetic patterns over consecutive days. In this regard, stock market data came to be seen as apt since it is by nature dynamic information. The choice of utilizing data from the Borsa Istanbul has been arbitrary and does not reflect any aimed reasoning. In the current investigation, a new algorithm is proposed based on uniting a time series-based forecasting model and multivariate adaptive regression technique. The performance of the proposed hybrid algorithm is evaluated by analyzing a case based on its data belonging to Borsa Istanbul. In the integrated structure, the Stock Trading variable is assigned the dependent variable, whereas all the other features, except a date are used as the independent predictors. A measure of how close the model forecast is compared with the observed actual of testing subset is used to test the predictive accuracy of the model. In particular, the assessment is based on the mean absolute residuals of the true values and predicted values of the dependent variable. Additive exponential smoothing time-series analysis algorithm is the first forecasting model to be used in this study. In this context stock trading is taken as the target variable. The model works with a set of predetermined parameters: the number of seasonal periods (t) will be 12, which is associated with a 12-month cycle. The level smoothing parameter is assigned as

$= 0.03$, the seasonal smoothing parameter as $\delta = 0{,}03$, and the trend smoothing parameter as $\gamma = 0{,}03$.

The second forecasting model is also built in accordance with the recently elaborated integrated algorithm and performed in three steps. During the first stage, the lock parameters are just the same independent variables (Open, Close, High, Low, and Volume) in the additive exponential smoothing time-series analysis which are considered separately as the parameters. Namely, the time-series forecasting procedure conducted on the dependent variable in the first model now becomes used in obtaining the respective values of the independent variables in the testing set. The second stage involves the development of MARSplines model based on the training dataset as well as the predictors that are already specified. The mathematical expression that is obtained at this step is subsequently treated as the reference function to be used to obtain the following estimations (Al-Nuaami et al., 2024). Lastly, at the third step, estimated values of the independent variables obtained at the first stage would be entered into the MARSplines function computed at the second stage and it would be possible to recompute all of the outcomes of dependent variable. Absolute correlations of the predicted and actual are therefore calculated to determine the performance of the model.

4. Findings

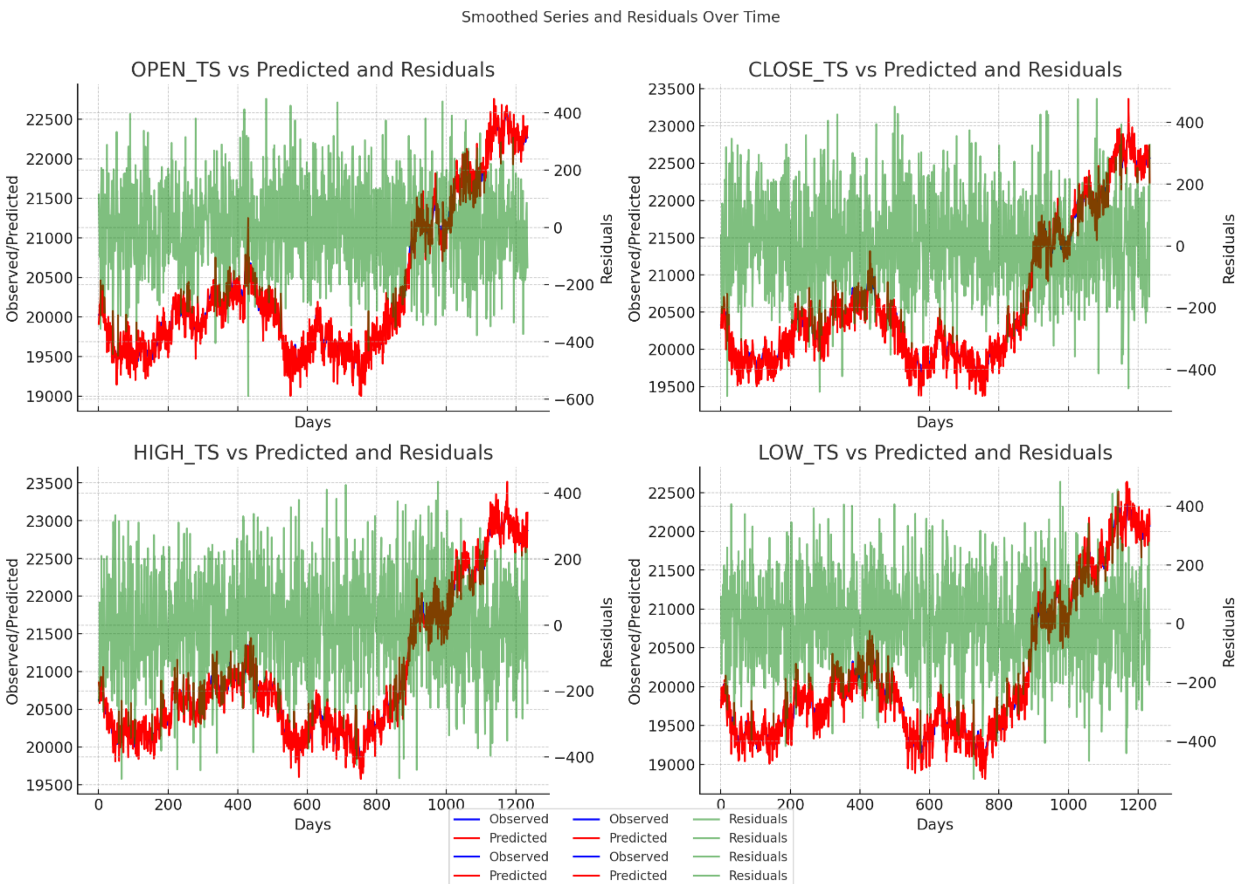
Based on the control (training) dataset, an exponential smoothing time-series predictive model is created, in which stock trading is considered as a dependent variable. The values set as parameters, in which the model has been constructed, are consistent with the parameters set at the very beginning. Based on this, all the entries in the testing set are re-calculated by applying this model formulation. The predicted values obtained from the model are presented in Table 1 (Notation: $E^{+N} = 10^{N}$ and $E^{-N} = 10^{-N}$), while the corresponding visualization of the model's output is shown in Fig 1.

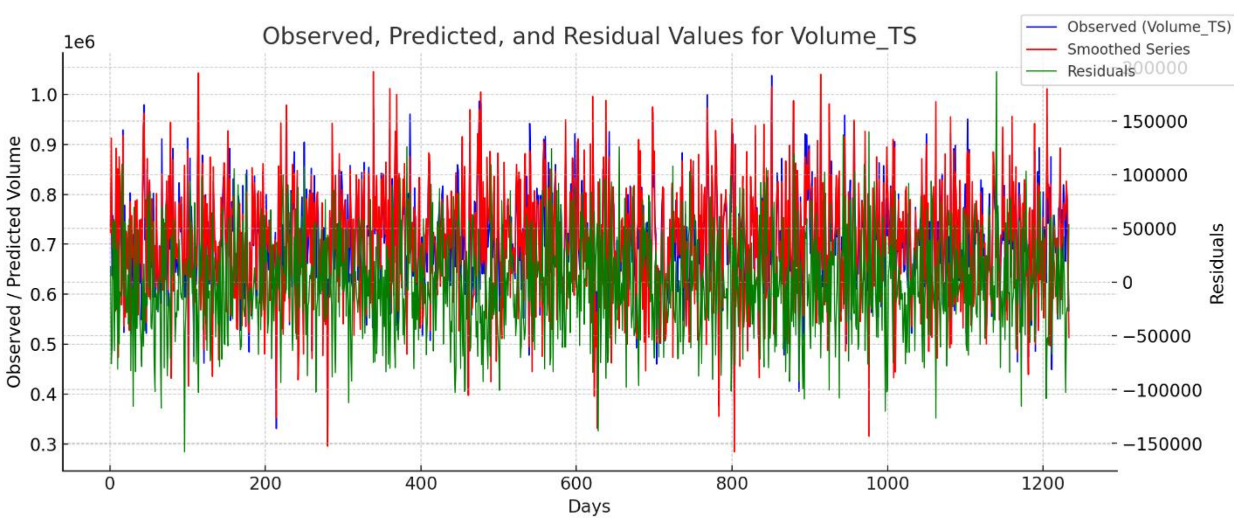
Table 1. Time Series-Based Observed, Forecasted, and Residual Values of Stock Indicators

TEST	OPEN	HIGH	LOW	CLOSE	VOLUME
OBSERVED	5.33E+4	5.44E+4	3.33E+4	5.43E+4	5.55E+5

	3.33E+4	6.22E+4	5.35E+4	4.54E+4	4.44E+5
	4.55E+4	4.22E+4	4.88E+4	4.79E+4	2.56E+6
	4.66E+4	5.98E+4	5.88E+4	5.66E+4	3.25E+6
	4.88E+4	4.77E+4	4.99E+4	4.77E+4	4.72E+5
	5.84E+4	4.55E+4	5.89E+4	5.89E+4	6.07E+5
	4.55E+4	5.90E+4	4.64E+4	4.90E+4	3.32E+6
(TIME SERIES) PREDICTED	5.41E+4	3.32E+4	5.55E+4	5.77E+4	4.07E+5
	3.39E+4	5.52E+4	3.23E+4	5.77E+4	4.47E+5
	5.59E+4	3.41E+4	4.25E+4	3.66E+4	5.68E+5
	5.58E+4	4.31E+4	6.64E+4	5.22E+4	3.27E+5
	5.56E+4	5.59E+4	3.32E+4	3.35E+4	5.86E+5
RESIDUALS	-4.04E+2	3.33E+0	-3.78E+1	6.99 E+2	2.45E+5
	5.52E+2	2.66E+2	2.19E+0	-2.38E+2	-2.08E+4
	-3.15E+3	-3.15E+3	-3.66E+3	-1.00E+3	6.68E+5
	-2.55E+3	-3.33E+3	-4.66E+3	-3.68E+3	3.29E+4
	-3.67E+3	-2.65E+3	-2.55E+3	-2.43E+3	2.42E+5
	-3.94E+3	-4.05E+3	-3.32E+3	-4.45E+3	5.00E+5

The updated model of the time series is developed based on the re-computed values of the factors (Open, Close, High, Low, and Volume) in order to estimate the dependent variable. The results of the predicted outcomes according to the time series approach are indicated in table 1. In addition to these, the values obtained during testing dataset plus the predicted value corresponding to the testing dataset and the residuals (difference in between observed and predicted) will also be placed within the same table. Moreover, Table 2 is graphically shown in Figure 1 in detail. Variables with comparable value ranges are presented together. The left y-axis represents the scale of observed and predicted values, while the right y-axis indicates the magnitude of residuals. The x-axis denotes the sequence of days, ranging from 1 to 1233. In the graphs, blue lines illustrate observed values, red lines indicate predicted values, and green lines display residuals. For a meaningful interpretation, the absolute and relative positions of these lines should be assessed independently. A closer alignment between the blue and red lines suggests higher model accuracy. Furthermore, the distribution of green lines around the baseline x = 0 reflects the model’s overall effectiveness. To assess the independence of residuals, scatterplots were generated with respect to the time variable (date). A non-random pattern in the distribution would indicate a violation of the independence assumption. However, the scatterplots produced for each variable revealed no such patterns, confirming that the residuals satisfy the independence condition across all variables.


Smoothed Series and Residuals Over Time
OPEN_TS vs Predicted and Residuals
CLOSE_TS vs Predicted and Residuals
HIGH_TS vs Predicted and Residuals
LOW_TS vs Predicted and Residuals
Observed/Predicted
Residuals
Days
Observed
Predicted
Residuals


Observed, Predicted, and Residual Values for Volume_TS
Observed (Volume_TS)
Smoothed Series
Residuals
1e6
Observed / Predicted Volume
Residuals
Days

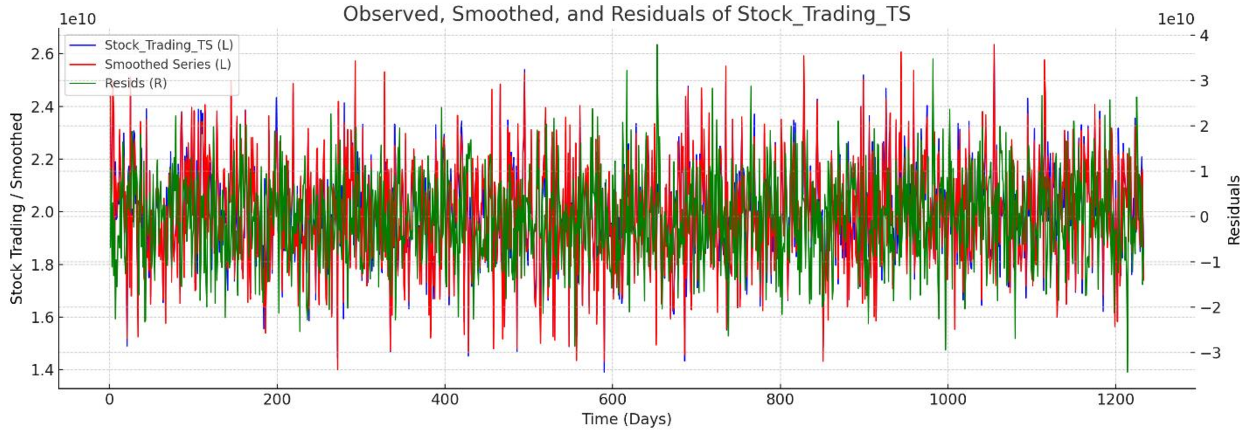


Fig 1. Time Series Prediction

During the next step, the variables generated using time-series smoothing model will be included into the MARSplines model to forecast the dependent variable. This model step is only applied to the training set. In such a way, one can evaluate the predictive performance of the model by carrying out its comparative analysis to conduct iterative refinements targeted towards making the models accurate. On the other hand, the dependent variable is concealed in the testing process and predications are formed by taking the most optimum model that has been constructed using the training data. That is, we do not tune or add any models to a test set, we only apply a finalized model. Considering that the dataset covers a period of six years, the seasonality effects might exist. So, the seasonal component in the model will explicitly have one period (mainly depending upon where the modeling activity is done), where the number of periods will be 12 as it will be representing the 12 months of the year. Still, the size of the testing set is too small to train a new regression model based on it: it contains only seven observations. Thus, the test set is then simply subjected to the mathematical structure of the most successful model of MARSplines generated during the training stage. Table 3 summarizes the details of this model and Appendix contains the Predictive Model Markup Language (PMML) representation of this model.

Table 2. Evaluation of Predictive Accuracy in Stock Trading via Two Analytical Models

TESTING SET	Observed_Stock	Time_Series_Stock	MARSplines_Stock
Case 1	3.85E+10	3.52E+10	1.43E+10

Case 2	4.68E+10	3.46E+10	3.54E+10
Case 3	4.36E+10	2.63E+10	3.54E+10
Case 4	3.67E+10	2.44E+10	4.78E+10
Case 5	3.52E+10	3.46E+10	3.46E+10
Case 6	2.61E+10	2.63E+10	3.76E+10
Case 7	4.32E+10	3.71E+10	1.44E+10

Table 3. MARSplines Regression Model Summary (Alternative Format)

		Regression Statistics
Dependent Variable	Stock Trading_TS	4.55E+10
Independent Variables	Open_TS, High_TS, Low_TS, Close_TS, Volume_TS	2.67E+10
Number of Terms	16	1.55E+10
Number of Basis Functions	12	6.42E+10
Order of Interactions	3	2.66E+6
Penalty	1	2.23E+9
Threshold	0.0005	7.32E+1
GCV Error	1.17E+19	4.32E+1

During testing the results of the dependent variable of seven cases are calculated using the function based on the MARSplines model in the form (Aksoy et al., 2022). In particular, the model that is created with the training data is applied to the testing dataset. The stock trading variable has estimated values using the given as the number referred to this formula during consideration as shown in Table 2. Let 's denote the dependent and independent variables in the MARSplines model as follows:

Stock trading data: Volume (x_1), High (x_2), Low (x_3), Open (x_4), with time- series indicator (Y). Clarify specific request for analysis or action. Accordingly, the MARSplines regression model that represents the relationship between the dependent and independent variables is expressed by the following mathematical formulation:

$$\begin{aligned} Y = {} & 8.63 \times 10^{10} \\ & -2.16 \times 10^{4} \cdot max(0, 3.18 \times 10^{6} - x_1) \\ & +3.14 \times 10^{6} \cdot max(0, x_2 - 2.32 \times 10^{4}) \\ & -2.82 \times 10^{6} \cdot max(0, 4.29 \times 10^{4} - x_2) \\ & +3.43 \times 10^{6} \cdot max(0, x_3 - 3.30 \times 10^{4}) \\ & +2.53 \times 10^{4} \cdot max(0, x_1 - 3.32 \times 10^{6}) \\ & -2.23 \times 10^{6} \cdot max(0, x_4 - 3.23 \times 10^{4}) \\ & +2.32 \times 10^{4} \cdot max(0, 3.73 \times 10^{4} - x_4) \\ & -3.23 \times 10^{6} \cdot max(0, x_3 - 3.42 \times 10^{4}) \\ & +3.23 \times 10^{6} \cdot max(0, x_4 - 2.34 \times 10^{4}) \\ & -2.36 \times 10^{6} \cdot max(0, x_3 - 2.19 \times 10^{4}) \\ & +8.12 \times 10^{5} \cdot max(0, x_2 2.68 \times 10^{4}) \\ & -6.32 \times 10^{5} \cdot max(0, x_3 - 4.32 \times 10^{4}) \end{aligned} \tag{9}$$

In Table 2, original (observed), the time-series predicted, and stock trading values are presented, which are calculated with the MARSplines function. In the following procedure, a model of a time series is fitted on the values of calculations based on the function found with MARSplines model. The level of accuracy of the combined method MARSplines and time-series algorithms is determined by the similarity rate of the observed and the predicted value. To this end, in the first place, the absolute correlation of predicted and observed data is compared. The success rate of prediction of the model in question is obtained in the following step.

Table 4. Correlation Between Predicted and Observed Values

Prediction Model	Absolute Correlation with Observed Values	P-Value
Times Series Model	0.523	0.062
Integrated Model	0.723	0.011

The coefficient between predicted values and observed values is a significant measure of how well the models of forecasting perform. Table 4 contains the absolute correlation coefficient of the two modeling strategies. Absolute value of the correlation between the predicted result of the MARSplines model and the observed is 0,723 compared to 0,523 of the time-series models. The values that are significant statistically are also marked in the table with red color. As per the findings, correlation coefficient linked to the MARSplines model is statistically significant at the 5 percent significance level, whereas that of time-series model is not statistically significant.

Table 5. Multiple Regression

Summary Statistics			
Multiple R	0.808	F(2,4)	8.32
Multiple R^2	0.724	P-value	0.021
Adjusted R^2	0.635	Std.Error (Estim)	6.67E

Table 6. Multiple Regression Parameter Estimations

Reg Summary						
N=7	b	Std.Error	b	Std.Error	T	P-value
Intercept			2.41E*11	4.95*10	5.09	0.007
Stock	0.61	0.621	8.06	4.36	1.34	0.361

MARS	-1.71	0.621	-.15.3	5.57	-2.93	0.044

During the second step of analysis, a multiple regression model is built in order to determine the predictive accuracy. In here, the close relation of the expected values and actual data is studied. The model assumes that original stock is a dependent variable, whereas MARSplines stock and time-series stock are independent ones. Table 5 contains a summary of the statistical performance of the model, whereas Table 6 contains an account of the parameter estimates. The results show that MARSplines Stock plays a significant role in the explanatory effect of the model as opposed to time-series stock which is not statistically significant.

Conclusion:

The research gives the new integrated prediction algorithm that can better represent data distribution over time. The emerging issue with explicit effect is the proliferation of user generated content over the past years which requires the use of goal seeking machine learning to gain relevant prediction success based on the outside influence. In respect of the presented research, the first important contribution of the proposed approach is that the results of the forecasting it produces are precise enough with large amounts of data and it requires minimum researcher intervention and an equivalent amount of training data to be used. Aligning with these requirements, the algorithm increases the forecasting ability of conventional time-series approaches and can be applied to datasets of diverse dimensions and fields without having integral proficiency in the field of machine learning. A nonlinear and very intricate dataset had to be chosen in order to put the efficacy of the proposed method to the test, i.e., data on the stock market which is prone to erratic oscillations, thus very difficult to learn. The empirical test was carried out on five years publicly available information on Borsa Istanbul. A time-series model of exponential smoothing was then followed to forecast values in the future. Thereafter, MARSplines model was used and it used predictors based on the time-series estimates. The dataset produced on the basis of MARSplines produced a higher absolute correlation using factual values (0.723) in contrast to the baseline time-seq model (0.523). Moreover, it was shown that the statistics of the predictive outputs of the MARSplines model are significant, which proves its outstanding activity compared to the generally accepted time-series modelling methods.

References:

Adiguzel, M. B., & Cengiz, M. A. (2023). Model selection in multivariate adaptive regressions splines (MARS) using alternative information criteria. *Heliyon*, *9*(9).

Aggarwal, R., Inclan, C., & Leal, R. (1999). Volatility in emerging stock markets. *Journal of Financial and Quantitative Analysis*, *34*(1), 33–55.

Aksoy, T., Karakaya, G., & Ghorbani, S. (2022). Application of fuzzy Topsis and Taguchi methods for optimization problems with disruptive risk: A systematic review. *Disruptive Technologies and Eco-Innovation for Sustainable Development*, 229–244.

Al-Nuaami, W. A. H., Dawod, L. A., Kibria, B. M. G., & Ghorbani, S. (2024). Design and implementation of a deep learning model and stochastic model for the forecasting of the monthly lake water level. *Limnological Review*, *24*(3), 217–234.

Aragão, D. P. (2024). *A set of independent variables for time series regression tasks of pandemic scenarios based on Covid-19*.

Baker, S., Xiang, W., & Atkinson, I. (2020). Continuous and automatic mortality risk prediction using vital signs in the intensive care unit: a hybrid neural network approach. *Scientific Reports*, *10*(1), 21282.

Bhowmik, D. (2013). Stock market volatility: An evaluation. *International Journal of Scientific and Research Publications*, *3*(10), 1–17.

Box, G. E. P., Jenkins, G. M., Reinsel, G. C., & Ljung, G. M. (2015). *Time series analysis: forecasting and control*. John Wiley & Sons.

Bundoo, S. K. (2011). *Asset price developments in an emerging stock market: The case of Mauritius*. African Economic Research Consortium Nairobi.

Chowdhury, A., Uddin, M., & Anderson, K. (2016). *Volatility spillovers and time-zone effect: New evidence from emerging markets across three different time zones*. Working Paper. Available online: https://pdfs. semanticscholar. org/d68c ….

Fama, E. F. (1970). Efficient capital markets: A review of theory and empirical work. *The Journal of Finance*, *25*(2), 383–417.

Friedman, J. H. (1991). *Estimating functions of mixed ordinal and categorical variables using adaptive splines*.

Ghorbani, S., & Pamucar, D. (2026). Remote Sensing-Based Evaluation of Lake Area Dynamics: A Quantitative Assessment for Environmental Management in Turkey. *Spectrum of Operational Research*, *3*(1), 352–358.

Ghous, H., Malik, M. H., Mahrukh, A., & Zaffar, A. M. (2023). Exchange stock price prediction using time series data: A survey. *Pakistan Journal of Humanities and Social Sciences*, *11*(2), 1110–1124.

Goodfellow, I., Bengio, Y., & Courville, A. (2023). Mit press: Cambridge, ma, usa, 2016. *Deep Learning.[Google Scholar]*.

Goonatilake, R., & Herath, S. (2007). The volatility of the stock market and news. *International Research Journal of Finance and Economics*, *3*(11), 53–65.

Gupta, P., Malhotra, P., Narwariya, J., Vig, L., & Shroff, G. (2020). Transfer learning for clinical time series analysis using deep neural networks. *Journal of Healthcare Informatics Research*, *4*(2), 112–137.

Hastie, T., Tibshirani, R., Friedman, J. H., & Friedman, J. H. (2009). *The elements of statistical learning: data mining, inference, and prediction* (Vol. 2). Springer.

Hyland, S. L., Faltys, M., Hüser, M., Lyu, X., Gumbsch, T., Esteban, C., Bock, C., Horn, M., Moor, M., & Rieck, B. (2020). Early prediction of circulatory failure in the intensive care unit using machine learning. *Nature Medicine*, *26*(3), 364–373.

Ilić, M. D. (2017). Toward a unified multi-layered modeling and simulation paradigm for electric energy systems. *2017 North American Power Symposium (NAPS)*, 1–6.

Javan, K., & Movaghari, A. R. (2022). Assessment of climate change impacts on extreme precipitation events in Lake Urmia Basin, Iran. *Desert*, *27*(1), 13–33.

Kao, L.-J., & Chiu, C. C. (2020). Application of integrated recurrent neural network with multivariate adaptive regression splines on SPC-EPC process. *Journal of Manufacturing Systems*, *57*, 109–118.

Khashei, M., & Bijari, M. (2011). A novel hybridization of artificial neural networks and ARIMA models for time series forecasting. *Applied Soft Computing*, *11*(2), 2664–2675.

Khosravi, K., Golkarian, A., Omidvar, E., Hatamiafkoueieh, J., & Shirali, M. (2023). Snow water equivalent prediction in a mountainous area using hybrid bagging machine learning approaches. *Acta Geophysica*, *71*(2), 1015–1031.

Liaw, A., & Wiener, M. (2002). Classification and regression by randomForest. *R News*, *2*(3), 18–22.

Lim, B., & Zohren, S. (2021). Time-series forecasting with deep learning: a survey. *Philosophical Transactions of the Royal Society A*, *379*(2194), 20200209.

Mode, G. R., & Hoque, K. A. (2020). Adversarial examples in deep learning for multivariate time series regression. *2020 Ieee Applied Imagery Pattern Recognition Workshop (Aipr)*, 1–10.

Okumus, H. S., Ghorbani, S., & Karatepe, S. (2019). A study on relationship between financial performance and supply chain in the accepted companies in Borsa Istanbul. Uncertain Supply Chain Management, 7(3), 417–426. https://doi.org/10.5267/j.uscm.2018.10.005

Rahmon, I., & Samson, E. (2023). *Stock Market Prediction Using Machine Learning Algorithms*.

Rai, A., Luwang, S. R., Nurujjaman, M., Hens, C., Kuila, P., & Debnath, K. (2023). Detection and forecasting of extreme events in stock price triggered by fundamental, technical, and external factors. *Chaos, Solitons & Fractals*, *173*, 113716.

Rumelhart, D. E., McClelland, J. L., & Group, P. D. P. R. (1986). *Parallel distributed processing, volume 1: Explorations in the microstructure of cognition: Foundations*. The MIT press.

Stanimirović, P. S., Stupina, A. A., Ghorbani, S., & JABBARI, K. H. (2024). Investigating the cost stickiness behavior of organizations after the economic recession caused by the COVID-19 pandemic. *Journal of Infrastructure, Policy and Development*, *8*(7), 3864.

Tong, H. (1990). *Non-linear time series: a dynamical system approach*. Oxford university press.

Tsay, R. S. (2005). *Analysis of financial time series*. John wiley & sons.

Vafakhah, M., Nasiri Khiavi, A., Janizadeh, S., & Ganjkhanlo, H. (2022). Evaluating different machine learning algorithms for snow water equivalent prediction. *Earth Science Informatics*, *15*(4), 2431–2445.

Wirawan, P. (2023). Leveraging Predictive Analytics in Financing Decision-Making for Comparative Analysis and Optimization. *Advances in Management & Financial Reporting*, *1*(3), 157–169.

Youssef Ali Amer, A., Wouters, F., Vranken, J., de Korte-de Boer, D., Smit-Fun, V., Duflot, P., Beaupain, M.-H., Vandervoort, P., Luca, S., & Aerts, J.-M. (2020). Vital signs prediction and early warning score calculation based on continuous monitoring of hospitalised patients using wearable technology. *Sensors*, *20*(22), 6593.

Youssef Ali Amer, A., Wouters, F., Vranken, J., Dreesen, P., de Korte-de Boer, D., van Rosmalen, F., van Bussel, B. C. T., Smit-Fun, V., Duflot, P., & Guiot, J. (2021). Vital signs prediction for COVID-19 patients in ICU. *Sensors*, *21*(23), 8131.

Zhang, G. P. (2003). Time series forecasting using a hybrid ARIMA and neural network model. *Neurocomputing*, *50*, 159–175.