Advance Journal of Econometrics and Finance
Vol-3, Issue-2, 2025

Advance Journal of Econometrics and Finance

Online ISSN

Print ISSN

https://ajeaf.com/index.php/Journal/About

[redacted]

Name of Publisher: SCHOLAR CRAFT EDUCATION & RESEARCH HUB

Review Type: Double Blind Peer Review

Journal Frequency: Quarterly Research Journal

Forecasting the KSE-100 index during Novel Coronavirus (COVID-19) Using Time Series Models

Nosheen Faiz[1], Muhammad Ali[2] Muhammad Ahmad[3], Soofia Iftikhar[4], Naz Gul[5], Muhammad Hamraz[*6]

Muhammad Ali
Higher Education Department Khyber Pakhtunkhwa.
Muhammad Ahmad
Department of Statistics Abdul Wali Khan University, Mardan, Mardan Pakistan.
Soofia Iftikhar
Department of Statistics, Shaheed Benazir Bhutto Women University, Peshawar, Pakistan.
Naz Gul
Department of Statistics Abdul Wali Khan University, Mardan, Mardan Pakistan.
Nosheen Faiz
Department of Statistics Abdul Wali Khan University, Mardan, Mardan Pakistan.
Muhammad Hamraz*
*Department of Statistics Abdul Wali Khan University, Mardan, Mardan Pakistan.
***Email ID:** mhamraz@awkum.edu.pk

Abstract

Forecasting stock market behavior is challenging due to its nonlinear, complex, and volatile nature, making it a key area of interest for financial experts and data scientists. Accurate predictions provide valuable guidance for investors in decision-making. The COVID-19 pandemic significantly impacted global economies, causing substantial declines in major stock markets. This study applies various time series forecasting models to predict the KSE-100 Index closing prices over short- and medium-term horizons (7, 15, and 30 days) and identifies the most accurate model based on out-of-sample performance. Using daily KSE-100 data from January 2 to December 31, 2020, the study also examines correlations with COVID-19 epidemiological variables. Empirical results indicate a negative correlation between daily COVID-19 confirmed cases and KSE-100 closing prices (-0.179, 95% CI: -0.333 to -0.053) and between daily COVID-19 deaths and KSE-100 closing prices (-0.188, 95% CI: -0.325 to -0.044). For the 30-day forecast, the drift method outperformed others in MASE, MAPE, MAE, and RMSE, while the simple exponential method was most accurate for 15-day forecasts, and the Holt method performed best for 7-day forecasts. These findings highlight the impact of the pandemic on market behavior and provide insights for investors and policymakers.

Keywords: KSE-100, COVID-19, ARIMA, PSX, Error trend seasonal, Neural network autoregressive

Introduction

The stock market plays a crucial role in every nation's economy, but forecasting financial time series such as stock prices remains challenging due to market fluctuations and multiple influencing factors [1]. Accurate predictions help investors plan and optimize their investments [2]. The outbreak of COVID-19 has further impacted global and local economies, causing significant revenue losses and market volatility [3]. This study focuses on forecasting the KSE-100 Index during the COVID-19 pandemic using different time series models and selecting the best model based on forecasting accuracy.

Time series data, collected at regular intervals, allows analysis of variable changes over time [4]. Forecasting predicts future values based on past observations and is widely used in fields such as finance, economics, biology, and astronomy. In financial markets, time series forecasting aids investors in making informed investment decisions. The KSE-100 Index, established in 1991, tracks the performance of 100 leading companies on the Pakistan Stock Exchange (PSX), representing 70–80% of the total free-float capitalization [5][6]. Over the years, the KSE-100 has experienced highs and lows, including sharp declines during the global financial crisis and the COVID-19 pandemic, where it fell by around 62% from January to March 2020 [7].

COVID-19, caused by the SARS-CoV-2 virus, emerged in Wuhan, China in December 2019 and rapidly became a global pandemic [8-9]. Its spread disrupted businesses, supply chains, and economies worldwide, leading to reduced investment activity and market losses [3] [11]. In Pakistan, the first COVID-19 case was reported on 26th February 2020, with subsequent lockdowns significantly affecting economic activity [12] [14] [16].

The novelty of this research lies in analyzing the relationship between COVID-19 variables—daily confirmed cases and deaths—and the daily closing prices of the KSE-100 Index, while simultaneously employing advanced univariate time series models to forecast the index over multiple horizons (7, 15, and 30 days). This dual approach bridges epidemiological data and financial market analysis, providing valuable insights for investors and policymakers during pandemic-induced uncertainty.

The paper is organized as follows: Section 1 presents the introduction. Section 2 reviews relevant literature on stock market forecasting. Section 3 outlines the research methodology and forecasting techniques. Section 4 presents results and discussion, and Section 5 concludes the study.

2. Literature Review

The stock market, as a dynamic indicator of economic activity, has garnered significant attention for predictive modeling. Various time series and machine learning models have been employed to forecast indices, including the KSE-100 Index. This section reviews methodologies and explores the influence of exogenous factors, particularly the COVID-19 pandemic. Traditional models such as ARIMA, first introduced by Box and Jenkins, have been widely used for stock market forecasting due to their ability to capture linear dependencies [17]. However, the financial market's nonlinear and chaotic nature has prompted researchers to incorporate hybrid models that integrate ARIMA with artificial intelligence techniques like artificial neural networks (ANNs) [18-19]. For instance, Makridakis et al. emphasized the value of combining statistical methods with machine learning to improve forecasting accuracy [20]. Machine learning and deep learning models have significantly advanced forecasting capabilities. Long short-term memory (LSTM) networks, which excel at capturing sequential dependencies, have been successfully applied in stock market predictions. Fischer and Krauss reported that LSTMs outperform traditional methods in predicting financial indices [21]. Similarly, Hussain et al. explored hybrid models combining LSTM with ARIMA, demonstrating superior performance in financial time series forecasting [22]. Shah et al. utilized convolutional neural networks (CNNs), which showed promising results in identifying intricate patterns in stock market data [23]. The COVID-19 pandemic introduced unprecedented volatility into global financial markets, prompting researchers to study its impact. Haroon and Rizvi analyzed pandemic-related news' effect on stock market volatility, highlighting significant correlations [24]. Ashraf conducted a broader analysis of the pandemic's influence on stock market behavior, concluding that COVID-19 case trends negatively impacted market performance across the globe [25]. Zaremba et al. explored how government containment measures, such as lockdowns, influenced stock market liquidity [26].

The method of neural network has been used by [27] to forecast the prices of different agriculture commodities. It is evident from empirical findings that the proposed NN model performance is better than the others. Similarly, the ANN has been utilized by [28] to forecast crude oil (WTI, Brent), Henry Hub natural gas, and New York Harbor No. 2 heating oil. By optimizing model configurations, it achieves high accuracy, with relative root mean square errors of 1.95%, 1.80%, 9.51%, and 20.35%, respectively. The results support technical analysis and policy decisions in energy markets. The study conducted by [29] suggested nonlinear autoregressive neural networks to forecast carbon emission allowance prices based on daily closing prices from the China Guangdong Carbon Emission Exchange (2013–2021). By optimizing model settings, accurate and stable predictions are achieved. The findings support technical and policy analyses, aiding stakeholders in understanding energy costs and planning for a green transition. Authors in [30] predict thermal coal futures trading volumes on the Zhengzhou Commodity Exchange (2016–2020) using nonlinear autoregressive neural networks. By optimizing model settings, it achieves accurate and stable predictions with minimal errors, covering up to the 99.273rd quantile of observed volumes.

In a study proposed by [31] used neural networks to forecast house prices across 100 major Chinese cities (2010–2019), marking the first machine learning-based analysis with such wide coverage. A simple model with four delays and three hidden neurons achieves stable performance, with an average relative root mean square error of 1%. The results support standalone or combined forecasting for market trends and policy analysis and offer a deployable framework for broader applications. Furthermore, Gaussian regression models have been widely used to predict different time series variables such as corn prices [32], steel price indices [33], pre-owned housing price index [34], regional steel price indices for east China [35], and palladium price prediction [36]. Despite the growing body of literature, the direct correlation between COVID-19-specific variables, such as daily confirmed cases and deaths, and stock market indices has not been extensively explored. This study aims to address this gap by analyzing the relationship between these variables and the daily closing prices of the KSE-100 Index. The novelty of this approach lies in integrating epidemiological data with financial forecasting techniques, thereby enhancing our understanding of stock market behavior during crises. In addition to correlation analysis, the study employs a multi-horizon forecasting framework (7, 15, and 30 days) to predict the KSE-100 Index, further distinguishing it from previous research. By adopting this dual focus, the study provides a comprehensive perspective on stock market dynamics in the context of global crises.

3. Methodology

This study provides a detailed overview of the data used in the study, as well as the forecasting methods employed to predict the KSE-100 closing prices. The statistical software, along with the fpp package in R, has been utilized to generate the numerical results. The methods applied in this study are outlined below, with their descriptions provided in the following sections. To achieve the stated objectives, a correlation test was conducted to determine if there is a relationship between COVID-19 confirmed cases and KSE-100 closing prices, as well as between COVID-19 confirmed deaths and KSE-100 closing prices. Forecasts for three different time periods, 30 days, 15 days, and 7 days, were generated using various time series forecasting methods, including the mean method, naïve method, drift method, simple exponential smoothing (SES), Holt's method, error trend seasonal (ETS), and autoregressive integrated moving average (ARIMA) during the pandemic. The accuracy of the models for each forecast horizon was evaluated using statistical metrics such as root mean square error (RMSE), mean absolute error (MAE), mean absolute percentage error (MAPE), and mean absolute scaled error (MASE).

3.1. Data Consideration

This study relies on secondary data comprising daily closing prices of the KSE-100 index, downloaded from Yahoo Finance (https://finance.yahoo.com), covering the period from January 1, 2020, to December 31, 2020. Any missing values in the closing prices were removed, resulting in a total of 251 observations. Three data sets were used for training and testing: the first set included 221 observations for training and the remaining 30 for testing; the second set included 236 observations for training and 15 for testing; and the third set included 244 observations for training and 7 for testing. The models discussed in Section 3.1 were fitted to the training sets, and parameter estimation was conducted using automated techniques. In-sample accuracy was assessed to determine which model best fit the data. Forecasts were then generated using the fitted models and evaluated against the test sets for out-of-sample accuracy, assessing which model provided better predictions of the KSE-100 closing prices for 30-step, 15-step, and 7-step forecasts. Data on COVID-19 was sourced from Kaggle (https://www.kaggle.com), specifically for Pakistan, covering the period from February 26, 2020, to December 31, 2020. The study examines the relationship between the daily KSE-100 closing price index and the daily confirmed COVID-19 cases, as well as daily confirmed deaths. Pearson's correlation coefficient was used to assess the relationships between the KSE-100 closing prices and both daily confirmed cases and deaths due to COVID-19. Some observations of daily confirmed cases and deaths were removed since the Pakistan Stock Exchange is closed on weekends, and data for those days is unavailable. This adjustment ensured consistency in data lengths by removing the corresponding observations of daily confirmed cases and deaths.

3.2. Univariate forecasting methods

In this study, the following univariate time series methods are used to predict the KSE-100 index during the period of the pandemic of COVID-19.

3.2.1 Mean method

Mean method or average method, the forecasts are equal to the average or mean of the historical values of the time series, supposing $y_1, y_2, \ldots, y_T$ be the historical values of a time series, then the forecast is given as

$$\widehat{Y}_{T+h|_T} = \bar{y} = (y_1 + y_2 + \ldots + y_T)/T \qquad (1)$$

Where the number of steps you want to forecast is represented by h, and all of the forecasted values will be equal to the mean of the historical values.

3.2.2 Naïve Method

In the naïve method, all the forecasts are equal to the last observed value of the time series. The h-step forecast is given as

$$\widehat{Y}_{T+h|_T} = y_T \qquad (2)$$

This method works well for some financial and economic data with respect to time. The naïve method utilizes the information from the most recent value of the series, therefore having no effect of the previous values in the series.

3.2.3 Drift Method

It is a modification of the naïve method, which allows the forecast to decrease or increase over time. Drift is the change, which is the mean change observed in the series. The equation for the h-step forecast is given as

$$\widehat{Y}_{T+h|_T} = y_T + \frac{h}{T-1}\sum_{t=2}^{T}(y_t - y_{t-1}) = y_T + \frac{h}{T-1}(y_T - y_1) \qquad (3)$$

Drawing a line from the last value to the first and then extrapolating it into the future is analogous to this.

3.2.4 Simple Exponential Smoothing Method

The most current values of the series are given larger weights in a weighted average forecasting method called "simple exponential smoothing" or "single exponential smoothing". The weights are exponentially lower as we move farther back in time. The one-step forecast by simple exponential is given as:

$$\widehat{Y}_{T+1|_T} = \alpha y_T + \alpha(1-\alpha)y_{T-1} + \alpha(1-\alpha)^2 y_{T-2} + \dots \qquad (4)$$

Where α is the smoothing parameter whose range is in between 0 and 1. Small values of α (i.e., closer to zero) will result in more weightage to the past values and large values of α (i.e., closer to one) will result in less weightage to the past values of the series.

3.2.5. The Holt-Winter Linear Trend Forecasting Procedure

An expansion of the standard exponential smoothing technique, the Holt method adds an additional parameter to capture the trend component in the data. This two-parameter model, alpha and beta used to represent the slope estimate at time t and represents the level estimate of the series at time t. In the Holt method, we have the following equations, which are written as;

Forecast equation: $\widehat{Y}_{T+1|_T} = l_t + hb_t$ (5)

Level equation: $l_t = \alpha y_t + (1-\alpha)(l_{t-1} + b_{t-1})$ (6)

Trend equation: $b_t = \beta^*(l_t - l_{t-1}) + (1-\beta^*)b_{t-1}$ (7)

Both α and $\beta^* = \alpha\beta$ have values ranging from zero to one. We have to estimate α, β ,l_0 and b_0. Estimation is done by optimization tools the values for α, β ,l_0 and b_0 are chosen from the observed data in such a manner that minimizes MSE.

3.2.6. Error Trend and Seasonal (ETS) Model

The ETS method is suitable for time series with trend and seasonal components and is a special case of exponential smoothing, often framed as a state-space model. It uses a three-letter code to denote error, trend, and seasonality, allowing for about 30 model variations (Table 3.1). Model parameters are typically estimated via maximum likelihood and optimized with sample smoothing to ensure stable forecasts.

Table 1. Different components of the ETS model

Trend Type	**Seasonal Type**		
	N	**A**	**M**
N	NN	NA	NM
A	AN	AA	AN
A_d	A_dN	A_dA	A_dM
M	MN	MA	MM

M_d	M_dN	M_dA	M_dM

In the above Table 1 the multiplicative, and damped multiplicative components is denoted by M and M_d, additive and damped additive by A, and A_d, and if there is no component present in time series then it is denoted by N. The best model among 30 different combinations will be selected on the basis of the well-known model selection criteria such as Akaike's information criteria (AIC) and Bayesian information criteria (BIC). Different combinations of ETS models will be fitted to the time series, the best model among these combinations will be that model having minimum values of AIC and BIC. The numerical formulas of these two model selection criteria is as follows [39, 40].

$$AIC = -2 * \ln(l) + 2 * p \tag{8}$$

$$BIC = -2 * \ln(l) + 2 * \ln(n) * p \tag{9}$$

3.2.7 Autoregressive Integrated Moving Average Model (ARIMA)

This method was first time proposed by Box and Jenkins [41] and after its first introduction it became the most widely used technique for predicting the univariate time series. Unlike the regression models this model is completely data driven and only requires the historical time series data whose past or lag values can be used as independent variables whereas the current values is considered as the dependent variable. It comprises the following three parts: AR (Autoregressive**),** I (Integrated) and MA (Moving Average)

AR part of the model tells the number of previous values or lags of the observed time series denoted by *'p'* present in the equation. The *'I'* part tell the number of times the series was differenced to make it stationary, and denoted by *'d'*, and the *MA* part of the model tells about the number of previous forecast errors or lags of forecast errors represented by *'q'*. If Y_t denotes a univariate time series*,* then the general form of *ARIMA (p, d, q)* can be expressed as under:

$$Y_t = \beta + \alpha_1 Y_{t-1} + \alpha_2 Y_{t-2} + \dots + \alpha_p Y_{t-p} + e_t - \varphi_1 e_{t-1} - \varphi_2 e_{t-2} - \dots - \varphi_q e_{t-q} \tag{10}$$

$$Y_t = c + \sum_{i=1}^{p} \alpha_i y_{t-i} + e_t - \sum_{j=1}^{q} \varphi_j e_{t-j} \tag{11}$$

The coefficients of the *MA* term denoted by $\varphi's$ are negative in the above equation, which is due to the proposal presented by Box and Jenkins, followed by many researchers. The technique for recognizing the values of *p, d,* and *q* for a time series is the autocorrelation functions (ACFs), and partial autocorrelation functions (PACFs). Understanding the plots of ACFs and PACFs are important in the sense that from these plots one can decide to choose the best candidate *ARIMA* model. Furthermore, from these plots are helpful in determining the order of AR and MA terms included in the model. It is worth mentioning here that the order of the ARIMA model does not depends directly on ACFs, and PACFs plots but it's just give the clue that what possible order of the model should be? Therefore, it is important that the readers knows about the ACFs and PACFs plots, both ACFs and PACFs functions are described in a separate section 4.1.3.

3.2.8 Neural Network Auto-Regression (NNAR)

The main idea of NN is taken from the human brain, and hence capable of solving complex problems in an efficient manner as compared to other ML algorithms. The performance of this novel ML algorithm depends upon its parameter tuning, and sometimes the inaccurate tuning produce poor prediction results as compared to the univariate time series models. When using the NN model for univariate time series and comparing it with other time series models the best choice to build the architect of this model is to use the lagged values of the actual time series as inputs, technically such model is then known as neural network autoregressin (NNAR). In this study a single hidden layer NN model using lagged values as input to the model for the purpose to predict the nonlinear, and nonstationary GDP growth rate of China, India, and Pakistan is implemented. A flow diagram of NNAR model having *p* lagged inputs and *k* nodes can be seen in the following Figure1.

Advance Journal of Econometrics and Finance
Vol-3, Issue-2, 2025

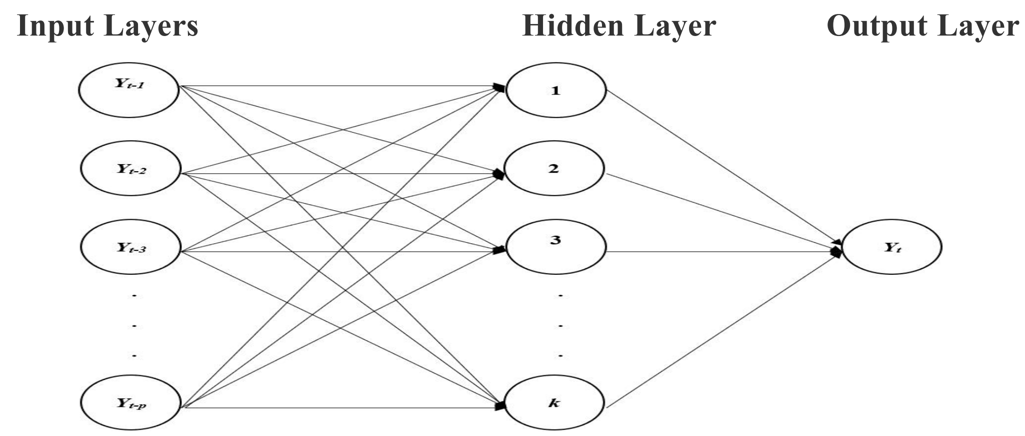


Figure 1: NNAR model with *p* lagged values as input layers

3.3. Accuracy Measures

Different statistical metrics has been used by the researchers to check the prediction accuracy of different time series and machine learning methods [42, 43, 44]. In this work, we used seven different univariate time series methods and a single machine learning algorithm which is known as NNAR to predict the nonlinear and nonstationary daily closing prices of KSE-100 index of Pakistan stock market. The best model out of these eight methods will be selected on the basis of the minimum values of the four statistical metrics such as RMSE, MAE, MAPE, and MASE. The mathematical formulas of these performance metrics are as follows:

$$RMSE = \sqrt{\frac{1}{N}\sum_{t=1}^{N}\left(A_t - P_{h,t}\right)^2} \tag{12}$$

$$MAE = \frac{1}{N}\sum_{t=1}^{N}\left|A_t - P_{h,t}\right| \tag{13}$$

$$MAPE = \frac{1}{N}\sum_{t=1}^{N}\left|\frac{A_t - P_{t,h}}{A_t}\right| \tag{14}$$

$$\text{MASE} = \text{mean}(|q_i|) \tag{15}$$

Where q_i in equation (27) is known as scaled error and defined as follows:

$$q_i = \frac{e_i}{\frac{1}{n}\sum_{i=2}^{n}|A_i - A_{i-1}|}$$

Furthermore, A_t and P_t denote actual and predicted values. The above defined metrics are the most widely used methods to check the forecast accuracy of different univariate time series methods to predict the one step ahead forecasts. Among these three performance metrics the MAPE is very common because of its simple mathematical structure and easy way to interpret its values. Apart from RMSE the other three performance metrics such as RMSE, MAE, and MASE have also been widely used in the previous studies, while understanding the numerical values of RMSE is not an easy task [45].

4. Results & Discussions

In this section of the paper we will analyze and discuss different results obtained throughout this study.

4.1.Figures of the Data

Figure 2 shows the daily KSE-100 closing prices in 2020, highlighting significant fluctuations due to the COVID-19 pandemic. The index began the year at 41,400 but dropped sharply in April amid global uncertainty, economic disruptions, and panic selling. A partial recovery occurred in the second quarter as restrictions eased and stimulus measures boosted investor confidence. Despite this, volatility persisted throughout the year, influenced by pandemic developments, vaccine progress, and macroeconomic factors. By year-end, the index improved but had not fully returned to pre-pandemic levels, reflecting a slow and cautious recovery.

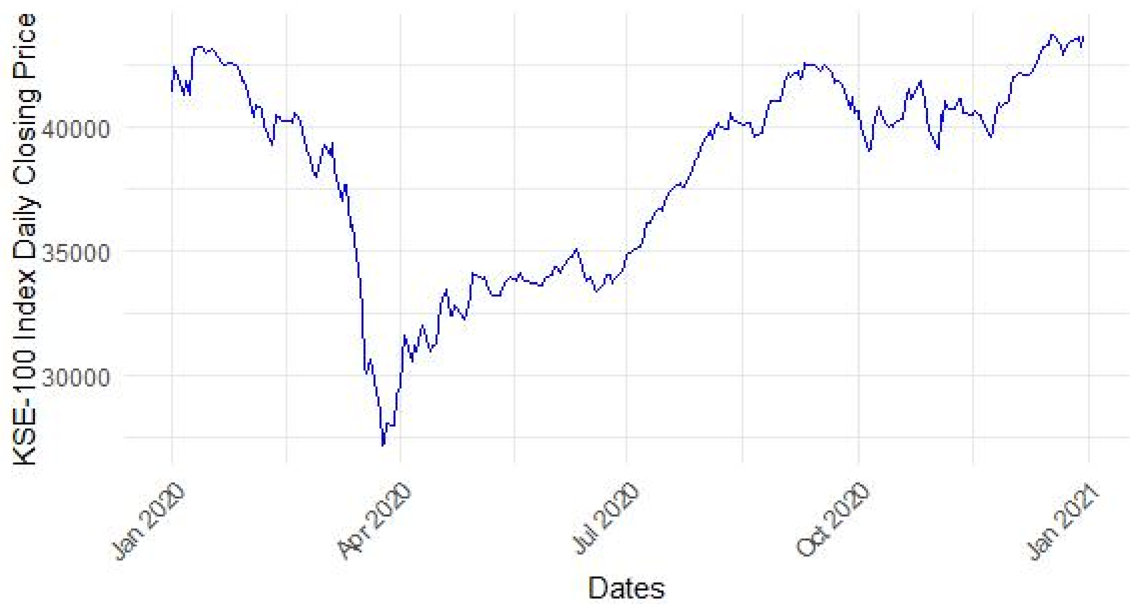


Figure 2: Daily closing prices of KSE-100 index in the time window January 01, to December 31 2020

Government and central bank interventions, including interest rate cuts by the State Bank of Pakistan, supported the market, but recovery was uneven due to policy uncertainty, inflation concerns, and fluctuating investor sentiment. Early pandemic fear gradually shifted to cautious optimism, with some investors viewing the market as undervalued, contributing to partial recovery. Figures 3 and 4 show that daily COVID-19 cases and deaths followed chaotic, nonlinear, and nonstationary trends. Daily cases were initially low due to limited travel, minimal community spread, and underreporting, then surged from April to June during the first wave, peaking as mobility increased and lockdown measures relaxed. Cases declined from July to September, reflecting the impact of public health interventions and seasonal factors, but a second wave emerged from October to December, driven by eased restrictions and gatherings, before showing signs of decline by year-end due to renewed containment efforts and increased public compliance.

Figure 3: Daily confirmed cases from COVID-19 in the time window February 26, to December 31 2020.

The trajectory of daily COVID-19 deaths in Pakistan, shown in Figure 4, reflects the timing and effectiveness of government interventions, including lockdowns, travel restrictions, and mass testing, as well as public adherence to preventive measures like social distancing and mask usage. Early in the pandemic, deaths were low due to limited local transmission, underreporting, and delayed symptom onset. As cases surged, daily deaths rose correspondingly, peaking at 153 on June 20, highlighting the strain on healthcare resources and the impact of widespread community transmission. The subsequent decline underscores the importance of timely interventions, public cooperation, and healthcare capacity in managing the crisis. Overall, these trends emphasize the need for continuous monitoring, adaptive policies, and data-driven strategies tailored to the local context to mitigate the human toll and prepare for future pandemics.

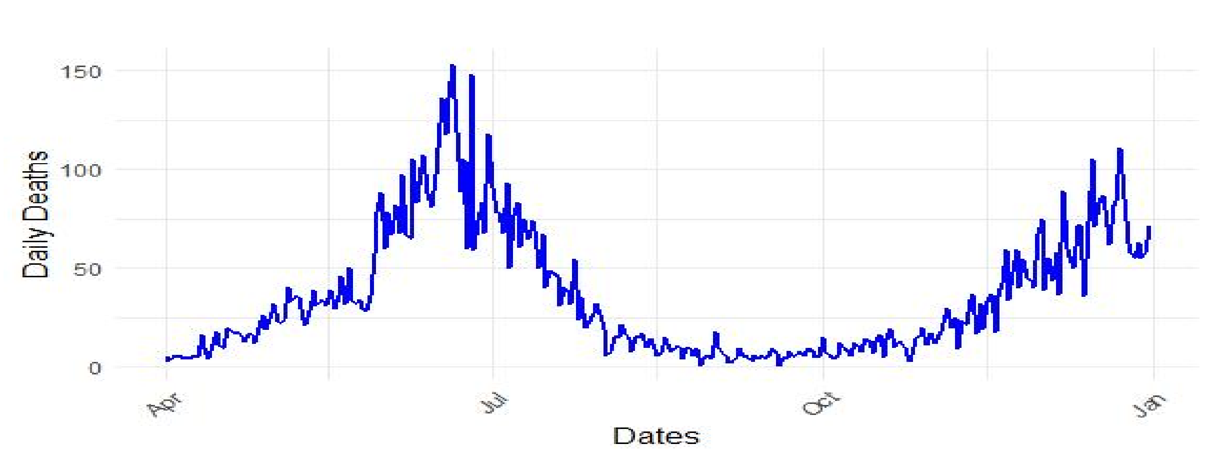


Figure 4: Daily deaths from COVID-19 in the time window February 26, to December 2020.

Following the June peak, daily COVID-19 deaths steadily declined alongside confirmed cases, reflecting the effectiveness of public health measures, increased testing, and public compliance with preventive practices. By December 2020, deaths had significantly decreased, highlighting the combined impact of interventions and partial population immunity. The trends emphasize the importance of timely measures, healthcare capacity, early detection, and sustained public engagement to mitigate fatalities and manage future pandemic waves.

4.2 Descriptive Statistics of the Data

The descriptive statistics of the KSE-100 index daily closing prices are also presented in the following table 2.

Table 2: Descriptive Statistics of the KSE-100 index daily closing prices

Count	Min	Max	Range	Skewness	Kurtosis
251	27229	43767	16537.89	-0.71	-0.61
Mean	**S.E Mean**	**St. Dev**	**Q1**	**Median**	**Q3**
37777	256.391	4061.998	34376	40030	41391

Table 2 summarizes the descriptive statistics of 251 daily KSE-100 closing prices. Prices ranged from 27,229 to 43,767, with a mean of 37,777 (S.E. = 256.39) and a standard deviation of 4,062, indicating moderate variability. The median was 40,030, with Q1 = 34,376 and Q3 = 41,391. The distribution is slightly left-skewed (skewness = −0.71) and platykurtic (kurtosis = −0.61), suggesting more frequent lower prices and fewer extreme fluctuations than a normal distribution. Overall, the market showed moderate stability with a slight tendency toward lower daily closing prices.

4.3 Autocorrelation and Partial Autocorrelation Plots

The ACF and PACF plots identify correlations between the KSE-100 index and its lagged values, helping determine the MA and AR orders for ARIMA modeling. The ACF shows strong correlations at the first few lags (lag 1 > 0.8, lag 2 > 0.6), which gradually weaken but remain significant, suggesting short-term dependence and possible cyclical patterns. Correlations beyond lag 20 fall within the confidence interval, indicating they are not statistically significant.

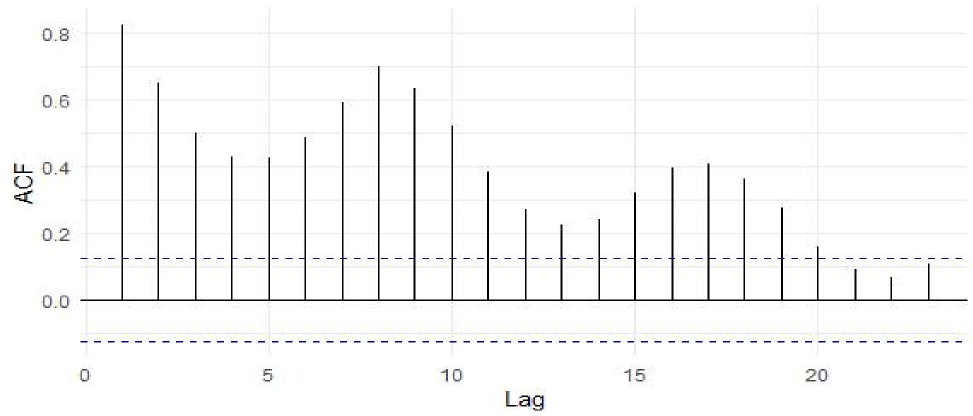
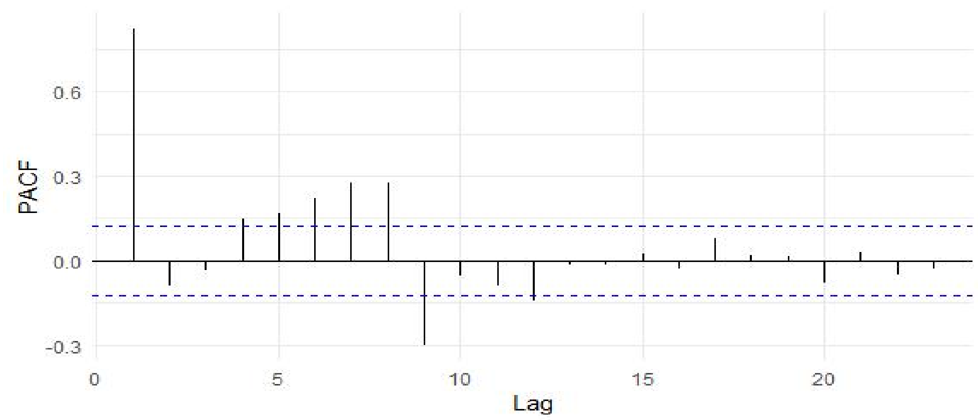


Figure 5: Autocorrelation function of KSE-100 index daily closing price in the time window January 1st, to December 2020.

The ACF shows significant correlations up to lag 20, suggesting an MA order of 20, after which correlations are insignificant. The PACF indicates a strong positive correlation at lag 1, weak or insignificant correlations at lags 2-7, and significant correlations at lags 3 and 4, implying these should be included in the AR structure. Lag 8 shows a slight negative correlation, while lags 9-20 are insignificant, indicating they do not contribute meaningfully to the series' autoregressive dynamics.



.

Figure 6: Partial autocorrelation function of KSE-100 index daily closing price in the time window January 1st, to December 2020.

Based on this analysis, the order of the AR terms in the ARIMA model can be inferred. Since there is significant autocorrelation at lag 1, and also at lags 3 and 4, it suggests an AR (4) model. This means that the time series can be best modeled using autoregressive terms up to lag 4, with the first four lags contributing significantly to the model. The remaining lags do not appear to add valuable information for the AR component.

4.4. Result of the Correlation Test

A correlation and regression analysis (Table 3) reveals a statistically significant negative relationship between daily confirmed cases and KSE-100 closing prices. The t-value (−2.7) and p-value ($0.0076 < 0.05$) indicate a meaningful inverse effect, with the estimated coefficient (−0.1973) showing that each unit increase in confirmed cases is associated with an average decrease of approximately 0.20 units in the KSE-100 index.

Table 3: Correlation test for KSE-100 closing prices and daily confirmed cases from COVID-19

t value	df	p value	Estimated Coefficient	95% Lower C.L	95% Upper C.L
-2.7	180	0.0076	-0.197292	-0.33318604	-0.05336797

This shows a clear inverse relationship. The 95% confidence interval for the coefficient ranges from -0.33318604 to -0.05336797. Since this range does not include zero, it confirms the statistical significance of the predictor's effect. The confidence interval also suggests that the true effect of the predictor is likely to fall within this range. Similarly, the correlation test is carried out using the KSE-100 index daily closing prices and daily deaths from COVID-19. The results of this test are presented in the following table 4.

Table 4: Correlation Test for KSE 100 closing prices and Confirmed deaths

t value	df	p value	Estimated Coefficient	95% Lower C.L	95% Upper C.L
-2.5714	180	0.01094	-0.188234	-0.32479668	-0.04398264

The t-value of −2.5714 indicates a statistically meaningful negative relationship between confirmed deaths and KSE-100 closing prices. With 180 degrees of freedom and a p-value of 0.01094 (< 0.05), the null hypothesis of no association is rejected. The estimated coefficient (−0.1882) suggests that each unit increase in confirmed deaths is associated with an average decrease of approximately 0.19 units in KSE-100 prices. The 95% confidence interval (−0.3248, −0.0440) excludes zero, further confirming the statistical significance of this inverse relationship.

4.5. Forecast Accuracy of the Different Models

This section evaluates model accuracy on training and testing data across 7-, 15-, and 30-day forecast horizons using MAE, RMSE, MAPE, and MASE. Results show that different univariate time series models perform best at different horizons, while the NNAR model consistently underperforms. These findings highlight that classical methods can outperform machine learning approaches and that model selection depends on the underlying data trends.

Table 5: In-sample accuracy for 30-days ahead forecast

Methods	MAE	RMSE	MAPE	MASE
Mean	3639.471	4109.484	10.10388	9.259615
Naïve	393.0478	571.5677	1.10164	1
Drift	393.6985	571.4969	1.103639	1.001656
SES	391.1247	570.0586	1.096176	0.995107
HW	404.6796	569.4153	1.137267	1.029594
ETS(A,A_d,N)	391.1494	557.4302	1.094528	0.99517
ARIMA(3,1,3)	**387.0274**	**553.1583**	**1.082163**	**0.984683**
NNAR	391.981	689.987	1.999	1.651

A 30-day-ahead forecast was initially considered; however, the conventional 80/20 train–test split was avoided due to market instability during the COVID-19 pandemic. Instead, short-term forecasts of 30, 15, and 7 days were generated to better support investor decision-making. Results in Table 5 show that, for the training data, the ARIMA model outperforms all others, achieving the lowest MAE, RMSE, MAPE, and MASE, followed by the ETS model, while the mean model performs the worst.

Table 6: Out-sample accuracy for 30-days ahead forecast

Methods	MAE	RMSE	MAPE	MASE
Mean	3279.919	3332.742	8.057916	8.344836
Naïve	513.4707	621.585	1.270707	1.306382
Drift	**476.4481**	**595.2408**	**1.176114**	**1.212189**
SES	513.4559	621.5712	1.270669	1.306345
HW	526.8812	632.6795	1.304443	1.340502
ETS	738.9372	891.0127	1.836584	1.880019
ARIMA(3,1,3)	926.6694	1057.892	2.300071	2.357651
NNAR	656.786	897.876	2.230	2.879

For the 30-day-ahead test forecast, the drift model outperforms all others, achieving the lowest RMSE, MAE, MAPE, and MASE, while the NNAR model underperforms relative to classical methods. As these models are data-driven rather than assumption-based, different models yield the highest accuracy for the training and testing datasets, as reflected by the minimum error measures. In the same way the following Table 7 presents the in-sample accuracy results for a 15-day-ahead forecast using various forecasting methods. The evaluation metrics include MAE, RMSE, MAPE, and MASE. These metrics are crucial for assessing the performance of each model in terms of forecasting accuracy.

Table 7: In-sample accuracy for 15-days ahead forecast

Methods	MAE	RMSE	MAPE	MASE
Mean	3574.372	4052.612	9.913099	9.059355
Naïve	394.5504	568.4149	1.095579	1
Drift	395.6905	568.1702	1.099066	1.00289
SES	392.7626	567.0263	1.09055	0.995469
HW	405.7904	566.7253	1.129715	1.028488
ETS(A,A_d,N)	392.5583	554.8245	1.088205	0.994951
ARIMA(3,1,3)	**389.0669**	**551.1355**	**1.078134**	**0.986102**
NNAR	398.768	602.781	1.769	1.098

ARIMA (3,1, 3) achieves the best in-sample performance with the lowest error measures, followed by ETS and SES. Holt, Naïve, and Drift show moderate accuracy, while the Mean and NNAR models perform worst, indicating that classical univariate models outperform the neural network approach for this forecast horizon.

Table 8: Out-sample Accuracy for 15-days ahead forecast

Methods	MAE	RMSE	MAPE	MASE
Mean	3021.676	3045.644	7.435968	7.658529
Naïve	1478.924	1527.299	3.634917	3.748378
Drift	1612.333	1650.086	3.964453	4.086507
SES	**1478.846**	**1527.224**	**3.634725**	**3.748181**
HW	2690.567	2750.177	6.627815	6.819324
ETS(A,A_d,N)	2320.207	2356.45	5.709002	5.880636
ARIMA(3,1,3)	2403.777	2461.915	5.911708	6.092446
NNAR	2398.098	2576.098	6.098	5.098

Simple exponential smoothing provides better forecasts than all other models, including the single machine learning NNAR model, as indicated by its minimum values of MAE (1478.846), RMSE (1527.224), MAPE (3.634), and MASE (3.748). The investigational results presented in Table 7 demonstrate that the second most efficient model in terms of forecast accuracy is the Drift model.

The following table 9 presents the in-sample accuracy for a 7-step ahead forecast of the KSE-100 index, with various forecasting methods evaluated using MAE, RMSE, MAPE, and MASE.

Table 9: In-sample accuracy for 7-step ahead forecast

Methods	MAE	RMSE	MAPE	MASE
Mean	3539.441	4022.612	9.811122	8.902195
Naïve	397.592	570.9554	1.098239	1
Drift	398.2241	570.8825	1.10014	1.00159
SES	395.8561	569.612	1.093385	0.995634
HW	408.6304	571.3494	1.131127	1.027763
ETS(A,A_d,N)	**395.7337**	**561.0335**	**1.091177**	**0.995326**

ARIMA(0,1,0)	395.909	569.6083	1.093508	0.995767
NNAR	409.808	587.9887	1.987	1.208

The Mean model performs worst, while Naïve and Drift show moderate accuracy. SES improves short-term forecasts, with ETS and ARIMA delivering the best performance. In contrast, the NNAR model yields the highest errors, indicating that classical univariate models are more effective for this forecast horizon. The last case of this forecast competition is for the 7-day ahead forecast for test data. The following Table 10 shows the investigational results of different models in terms of the numerical values of MAE, RMSE, MAPE, and MASE.

Table 10: Out-sample accuracy for 7-days ahead forecast

Methods	MAE	RMSE	MAPE	MASE
Mean	2680.419	2700.732	6.633015	6.741633
Naïve	219.4079	382.7555	0.549843	0.551842
Drift	199.5577	353.2897	0.500151	0.501916
SES	219.435	382.7874	0.549911	0.55191
HW	**194.6447**	**341.7864**	**0.487732**	**0.489559**
ETS(A,A_d,N)	201.6063	318.5844	0.504136	0.507068
ARIMA(0,1,0)	219.4079	382.7555	0.549843	0.551842
NNAR	220.987	390.192	0.6767	0.7657

The Mean model performs worst, while Naïve, SES, Holt, and ARIMA show reasonable accuracy. The Drift model achieves the lowest errors and is the most effective for the 7-day-ahead forecast. In contrast, the NNAR model underperforms classical time series methods for this horizon.

5. Conclusion

This study examined the importance of forecasting stock market movements during periods of extreme uncertainty, with particular emphasis on the KSE-100 index of the Pakistan Stock Exchange during the COVID-19 pandemic. The primary objectives were to compare the predictive performance of eight forecasting methods for daily KSE-100 closing prices and to analyze the relationship between market movements and COVID-19 indicators. Empirical results reveal a weak but statistically significant negative correlation between daily COVID-19 confirmed cases and deaths and the KSE-100 closing prices, highlighting the adverse impact of pandemic dynamics on market behavior. Seven classical univariate time series models and one machine learning–based approach were evaluated using four accuracy measures: RMSE, MAE, MAPE, and MASE. Forecasting performance was assessed across three horizons (30-, 15-, and 7-day-ahead) using separate training and testing datasets. For the 30-day-ahead horizon, the Drift model and ARIMA exhibited superior performance for training and testing datasets, respectively. In the 15-day-ahead forecasts, ARIMA(3,1,3) and Simple Exponential Smoothing (SES) consistently achieved the lowest error values, followed by ETS and the Naïve model. For short-term (7-day-ahead) forecasts, ETS and Holt-Winters models outperformed other approaches, demonstrating strong short-term predictive capability.

Overall, classical univariate time series models consistently outperformed the NNAR machine learning model across all forecast horizons. This suggests that, despite the typically volatile nature of financial markets, the KSE-100 index exhibited structured patterns during the pandemic period, likely driven by its association with epidemiological variables. These findings indicate that univariate time series models remain effective forecasting tools when market behavior is influenced by identifiable external shocks.

The main limitation of this study is the exclusive reliance on univariate forecasting approaches, with limited use of machine learning techniques. Future research will extend this work by incorporating advanced machine learning models, such as support vector machines and artificial neural networks, along with relevant exogenous variables, to further enhance forecasting accuracy and robustness under complex market conditions.

References

[1]. N. M. Mahreen Mahmud, "Volatility Dynamics in an Emerging Economy: Case of Karachi Stock Exchange," Economic Research-Ekonomska Istraživanja , vol. 24, no. 4, pp. 51-64, 2011.

[2]. O. A. K. A. Ayodele A. Adebiyi., "Stock Price Prediction Using the ARIMA Model," 2014 UKSim-AMSS 16th International Conference on Computer Modelling and Simulation, Cambridge, pp. 106-112, 2014.

[3]. J. L. W. R. Mohsin Shafi, "Impact of COVID-19 pandemic on micro, small, and medium-sized Enterprises operating in Pakistan," Research in Globalization, vol. 2, no. ISSN 2590-051X, 2020.

[4]. K.-S. C. Jonathan D. Cryer, Time Series Analysis With Applications in R,Second Edition, Springer-Verlag New York, 2008.

[5]. N. U. K. Shahan Akhtar, "Modeling Volatilty on the Karachi Stock Exchange," Journal of Asia Buisness Studies, vol. 10, no. 3, pp. 253-275, 2016.

[6]. Pakistan Stock Exchange Limited, "Company Profile," 2019. [Online]. Available: https://www.psx.com.pk/psx/exchange/profile/about-us.

[7]. Malik, "Market," Buisness Recorder, 2020 Apr 2020. [Online]. Available: https://www.brecorder.com/news/592525/psx-remains-calm-despite-covid-19-hits-market-worse-than-2008-crisis.

[8]. F. A. R. M. F. I. Z. Ghufran Ahmad, "Evaluating Data-Driven Forecasting Methods for Predicting SARS-CoV2 Cases: Evidence From 173 Countries," medRxiv, 2020.

[9]. World Health Organization, "Coronavirus disease (COVID-19)," 2020. [Online]. Available: https://www.who.int/emergencies/diseases/novel-coronavirus-2019/question-and-answers-hub/q-a-detail/coronavirus-disease-covid-19.

[10]. M. C. Velavan TP, "The COVID-19 epidemic," Trop Med Int Health, vol. 25, no. 3, pp. 278-280, 2020.

[11]. O. S. G. Mert Topcu, "The impact of COVID-19 on emerging stock markets," Finance Research Letters, vol. 36, no. 1544-6123, 2020.

[12]. Ali, "Pakistan confirms first two cases of coronavirus, govt says “no need to panic”," 27 February 2020. [Online]. Available: https://www.dawn.com/news/1536792.

[13]. The Express Tribune, "K-P reports second COVID-19 fatality," 18 March 2020. [Online]. Available: https://tribune.com.pk/story/2178813/1-pakistan-reports-first-covid-19-death-g-b.

[14]. The Express Tribune, "LIVE: Govt decides to extend 'lockdown' until April 14 after NCC meeting on COVID-19," 1 April 2020. [Online]. Available: https://tribune.com.pk/story/2188441/1-live-ndma-establish-12-labs-test-covid-19-tally-crosses-2000.

[15]. DAWN, "Cabinet approves easing lockdown after May 9," 06 May 2020. [Online]. Available: https://www.dawn.com/news/1554927.

[16]. The Tribune, "Covid-19: Pakistan records highest single-day increase, tally at 1,39,230," 14 June 2020. [Online]. Available: https://www.tribuneindia.com/news/world/covid-19-pakistan-records-highest-single-day-increase-tally-at-1-39-230-99026.

[17]. G. E. Box and G. M. Jenkins, Time Series Analysis: Forecasting and Control, San Francisco, CA, USA: Holden-Day, 1976.

[18]. G. P. Zhang, "Time series forecasting using a hybrid ARIMA and neural network model," Neurocomputing, vol. 50, pp. 159–175, 2003.

[19]. S. H. Yu, "A hybrid model integrating ARIMA with artificial neural networks for a time series forecasting," Appl. Soft Comput., vol. 11, no. 2, pp. 231–238, 2011.

[20]. S. Makridakis, E. Spiliotis, and V. Assimakopoulos, "Statistical and machine learning forecasting methods: Concerns and ways forward," PLoS One, vol. 13, no. 3, p. e0194889, 2018.

[21]. T. Fischer and C. Krauss, "Deep learning with long short-term memory networks for financial market predictions," Eur. J. Oper. Res., vol. 270, no. 2, pp. 654–669, 2018.

[22]. M. Hussain, J. Li, M. N. Ali, and M. N. Amjad, "Predictive analysis of financial data using hybrid ARIMA and LSTM models," J. Comput. Sci., vol. 36, pp. 101034, 2020.

[23]. D. Shah, H. Isah, and F. Zulkernine, "Stock market analysis: A review and taxonomy of prediction techniques," Int. J. Financial Stud., vol. 7, no. 2, pp. 26, 2019.

[24]. O. Haroon and S. A. R. Rizvi, "COVID-19: Media coverage and financial markets behavior—A sectoral inquiry," J. Behav. Exp. Finance, vol. 27, pp. 100343, 2020.

[25]. B. N. Ashraf, "Stock markets’ reaction to COVID-19: Cases or fatalities?" Res. Int. Bus. Finance, vol. 54, pp. 101249, 2020.

[26]. Zaremba, R. Kizys, D. Y. Aharon, and E. Demir, "COVID-19, government policy responses, and stock market liquidity around the world: A note," Res. Int. Bus. Finance, vol. 56, pp. 101359, 2020.

[27]. Jin, B. and Xu, X. (2024), "Wholesale price forecasts of green grams using the neural network", Asian Journal of Economics and Banking, Vol. ahead-of-print No. ahead-of-print. https://doi.org/10.1108/AJEB-01-2024-0007.

[28]. Jin, B. and Xu, X., 2024. Price forecasting through neural networks for crude oil, heating oil, and natural gas. Measurement: Energy, 1(1), p.100001

[29]. Jin, B. and Xu, X., 2016. Carbon emission allowance price forecasting for China Guangdong carbon emission exchange via the neural network. Global Finance Review. 2024; 6 (1): 3491.

[30]. Jin, B., Xu, X. and Zhang, Y., 2024. Thermal coal futures trading volume predictions through the neural network. Journal of Modelling in Management

[31]. Xu, X. and Zhang, Y., 2021. House price forecasting with neural networks. Intelligent Systems with Applications, 12, p.200052

[32]. Jin, B. and Xu, X., 2024. Forecasting wholesale prices of yellow corn through the Gaussian process regression. Neural Computing and Applications, 36(15), pp.8693-8710.

[33]. Jin, B. and Xu, X., 2024. Predictions of steel price indices through machine learning for the regional northeast Chinese market. Neural Computing and Applications, 36(33), pp.20863-20882.

[34]. Jin, B. and Xu, X., 2024. Pre-owned housing price index forecasts using Gaussian process regressions. Journal of Modelling in Management.

[35]. Jin, B. and Xu, X., 2024. Machine learning predictions of regional steel price indices for east China. Ironmaking & Steelmaking, p.03019233241254891.

[36]. Jin, B. and Xu, X., 2024. Palladium Price Predictions via Machine Learning. Materials Circular Economy, 6(1), p.32.

[37]. Gerlach, S. and Peng W. (2006) Output Gaps and Inflation in Mainland China. China Economic Review, 17, 210-225. http://dx.doi.org/10.1016/j.chieco.2005.09.002.

[38]. Zhao, Y. (2006) The Establishment and Analysis of the Times Series Model of China's GDP. Journal of Xi'an Institute of Finance and Economics, 3, 11-14

[39]. Marcellino, M., Stock, J. H. & Watson, M. V. (2006). A comparison of direct and iterated multistep AR methods for forecasting macroeconomic time series, Journal of Econometrics

[40]. Binner, J. M., Elger, T., Nilsson, B & Tepper, J. A. (2005). Tools for non-linear time series forecasting in economics- an empirical comparison of regime switching vector autoregressive models and recurrent neural networks, Advances in Econometrics, 19, pp. 71-92

[41]. Zhao, X.D. (2008) Estimation of Output Gap in China Based on Phillips Curve. The Journal of World Economy, 1, 57- 64

[42]. Xu, X. and Zhang, Y., 2022. Second-hand house price index forecasting with neural networks. Journal of Property Research, 39(3), pp.215-236.

[43]. Xu, X. and Zhang, Y., 2022. Soybean and soybean oil price forecasting through the nonlinear autoregressive neural network (NARNN) and NARNN with exogenous inputs (NARNN–X). Intelligent Systems with Applications, 13, p.200061.

[44]. Xu, X. and Zhang, Y., 2023. Coking coal futures price index forecasting with the neural network. Mineral Economics, 36(2), pp.349-359.

[45]. Ang, A., Bekaert, G. & Wei M. (2007). Do macro variables, asset markets, or surveys forecast inflation better? Journal of Monetary Economics, Volume 54, Issue 4, May 2007, pp. 1163-1212.