A Statistical Investigation and Forecasting of Gold Prices through Box-Jenkins ARIMA in Pakistan

Habib Ullah Khan
Department of Statistics, University of Malakand, Pakistan
Email: habibullah858@uom.edu.pk

Muhammad Seyab Khan (Corresponding Author)
Department of Statistics, University of Malakand, Pakistan
Email: seyabk567@gmail.com

Atta Ullah
Department of Statistics, University of Malakand, Pakistan
Email: attaullaharsalan34@gmail.com

Abstract
Gold plays a significant role in the financial market, monetary policy, and various industries, including jewelry, pharmaceuticals, and electronics; however, its price is often volatile. This study focuses on forecasting daily gold prices of Pakistan using various ARIMA models to identify the most suitable approach for accurate gold price prediction. Gold forecasting is a critical and intriguing area of study, and the Box-Jenkins methodology, which employs ARIMA models, is highly effective for analyzing and forecasting time series with diverse patterns of variation. The ARIMA modeling process involves defining, estimating, diagnosing, and forecasting stages, making it well-suited for this application. Based on the analysis, the ARIMA (4, 1, 4) model was the best fit for the dataset, evaluated using SIGMASQ, AIC, and SC criteria, which yielded the smallest values and were confirmed through the Box-Jenkins. Using the ARIMA (4, 1, 4) model, the study successfully forecasted daily gold prices of Pakistan from October 2028 to June 2022.
Keywords: Time Series Analysis, Box-Jenkins methodology, Forecasting, Gold Forecasting, ARIMA Models

Introduction
The proverb "All that glitters is not gold" reminds us that not everything that appears attractive is valuable in reality. In Pakistan, gold has captured the interest of people from all treads of life as a significant asset option. Those investing in gold typically have two primary goals: first, to act as a hedge against inflation, as returns on gold investments often align with inflation rates over time; and second, to diversify their investment portfolios, helping to mitigate risk and reduce overall volatility. How people invest in gold has evolved significantly. Traditionally, individuals bought gold in the form of jewelry. However, modern investment options now include purchasing gold

coins and bars available through scheduled banks and investing in Gold Exchange Traded Funds (ETFs). Gold ETFs are financial instruments similar to mutual funds, investing directly in gold and listed on stock exchanges for trading. Other investment avenues include Gold Fund of Funds, which allows investment without the need for a Demat account, and Equity-based Gold Funds, where investments are made in companies involved in gold mining, extraction, and marketing rather than directly in gold itself. The significance of gold in Pakistan has shifted over time. Historically, thousands of years ago, gold was a universally accepted medium of exchange among various kingdoms and empires. Today, Pakistan has a strong demand for gold, particularly for jewelry fabrication, placing it among the top gold-importing countries due to limited local gold mining production. Demand for gold often surges during key occasions, such as wedding seasons, post-harvest periods, and festivals, while it tends to decline during the monsoon season. Currently, the National Stock Exchange has introduced various instruments linked to gold investments, making it easier for investors to buy gold through exchanges. This system reduces the additional costs associated with purchasing jewelry, which can negatively impact investor returns. Therefore, investors need to stay informed about fluctuations in gold prices to make informed decisions. This paper aims to provide insights into forecasting gold prices using a time-series ARIMA model, tailored to the context of Pakistan.

Literature review

This study highlights how critical it is to select the best forecasting models in edict to increase prediction accuracy, which is essential for preserving investor trust and making well-informed decisions, Shafie et al. (2022).

The paper analyzed gold prices in Pakistan from January 1995 to January 2019 using the Markov Regime Switching Autoregressive (MRS-AR) model, revealing significant heteroscedasticity and non-stationary behavior. It highlighted the importance of regime changes for understanding price fluctuations and demonstrated that the MRS-AR model can effectively forecast in complex data scenarios. The authors advocated for the use of regime-switching models across various forecasting applications due to their flexibility and resilience, Qasim et al. (2021).

The study examined how effective ARIMA models are in forecasting daily gold prices in Malaysia from 2003 to 2014, using symmetric GARCH-type models. It investigated three types of innovation distributions: generalized error, t, and Gaussian. The findings revealed that the best model for predicting the dynamics of gold prices was the ARIMA (0,1,0) combined with a standard GARCH(1,1) model featuring t-distributed innovations, Yaziz & Zakaria (2019).

Their research demonstrates how well ARIMA captures time-series patterns and uses past data to predict future trends. Autocorrelation functions (ACF), partial autocorrelation functions (PACF), and the Augmented Dickey-Fuller (ADF) test for stationarity are among the diagnostic tests used to determine the parameters of the ARIMA model: autoregressive order (p), differencing order (d), and moving average order (q). Additionally, by adding non-linear relationships to the data, they used SVM to improve prediction accuracy. Their results highlight how crucial it is to combine machine learning and statistical methods for reliable financial forecasting Makala and Li (2021).

The paper examined the patterns of international oil price fluctuations using the ARIMA-GARCH model, highlighting the combination of GARCH's capacity to model volatility and ARIMA's linear forecasting capabilities. This hybrid approach offers a strong framework for financial forecasting by efficiently capturing both the trend and variance in time-series data. The study emphasizes the importance of combining statistical models to handle complex market dynamics, particularly for commodities with significant price volatility Xiang (2022).

According to this study, ARIMA (1, 1, 2) is best suited for short-term forecasts because actual and predicted values closely match and post-sample forecasts show an upward trend. ARIMA (1, 1, 1) was chosen for another study's Box-Jenkins methodology because of its predictive accuracy over 38 years of data. Strong explanatory power was demonstrated by these models' consistently high R2 values (>0.99) and low error metrics (MAE, RMSE), Singh and Kumar (2017).

Using historical data from November 2003 to January 2014, the paper investigated the use of ARIMA models for predicting gold prices in India. The study showed how well ARIMA models capture price trends and fluctuations, providing investors with insightful information to help them make the best possible buying and selling decisions. Tripathy demonstrated the model's usefulness in reducing the risks connected with gold investments by examining historical price movements, especially in a volatile economic climate impacted by political and inflationary factors, Tripathy (2017).

Yakean looked at the factors that affect gold prices in Thailand, emphasizing a number of market and economic variables. The study most likely looked at the effects of macroeconomic factors on gold prices, including inflation, interest rates, and exchange rates. Gaining an understanding of these elements is essential to creating forecasting models like ARIMA that work well in Pakistan and other markets. Researchers can gain a better understanding of how comparable factors may impact gold prices in other nations by examining the factors that influence gold prices in Thailand. This analysis can then be used to develop and implement models, Yakean (2022).

The study demonstrated the integration of statistical and machine learning techniques for time series analysis by presenting a hybrid forecasting model that combines SARIMA and Artificial Neural Networks (ANN) to predict water levels in the Red Hills Reservoir. By identifying both linear and nonlinear patterns in data, hybrid models provide better accuracy and dependability than standalone methods, as the study showed. With the integration of ARIMA with cutting-edge techniques to improve predictive performance, this methodology shows the potential for applying similar hybrid approaches to gold price forecasting in Pakistan, Azad et al. (2022).

Seasonal Autoregressive Integrated Moving Average (SARIMA) and Nonlinear Autoregressive Neural Network (NAR) models were combined in Wang et al.'s traffic flow prediction method. Accuracy and reliability are increased by this hybrid approach, which successfully combines the linear fitting capabilities of SARIMA with the nonlinear interpretation and memory functions of NAR. With high accuracy rates (up to 92%) and good adaptation to seasonal variations and outside disturbances like epidemics, the study showed that the SARIMA-NAR combination model outperforms standalone models in terms of predictive performance. According to the paper, this study demonstrates the potential of integrating statistical and machine learning models for time series forecasting, providing insights that could be used to predict the price of gold in Pakistan through hybrid approaches based on ARIMA, Wang et al. (2022).When choosing between the ARIMA and ARFIMA frameworks for economic and financial time series modeling, they stress the importance of long-memory verification. Their analysis shows that whether a series needs fractional integration (ARFIMA) or integer differencing (ARIMA) to address non-stationarity depends on the estimation of the fractional differencing parameter. This means that in order to detect long-range dependence in gold price forecasting, preliminary tests such as R/S analysis, ADF, and PP tests should be carried out. ARIMA models may perform poorly if there is significant fractional integration, Ismail and Al-Gounmeein (2022).They examine how SARIMA (Seasonal ARIMA) models are used to predict traffic accidents, highlighting how crucial it is to account for seasonality in time series forecasting. Their research demonstrates how well the model manages seasonal and non-seasonal components, guaranteeing precise predictions in datasets with recurring patterns. Given that gold prices frequently show seasonal trends influenced by cultural, economic, and international factors, this method is pertinent to forecasting gold prices in Pakistan. In order to choose the best model configuration, the authors also emphasize the importance of assessing model performance using metrics like AIC and BIC. This realization highlights how SARIMA models or comparable extensions could be used to improve ARIMA-based gold price forecasts in Pakistan's dynamic market environment, Deretic et al. (2022).

Aimer and Lusta use a Markov-switching VAR approach, which takes regime changes in economic conditions into account, to examine the relationship between exchange rates and oil prices under uncertainty. Their research highlights how crucial it is to record structural alterations in time series data, especially for commodities that are impacted by macroeconomic variables. Since both commodities are susceptible to changes in the global economy and politics, the methodology offers insights into modeling gold prices even though its focus is on oil prices. By addressing sudden changes in market trends, which are typical of gold price movements brought on by external shocks, regime-switching models combined with ARIMA could improve forecasting accuracy, Aimer & Lusta (2021).The dynamics of food price volatility and its effects on household welfare in Nigeria are examined by the study, which stresses the significance of comprehending commodity price fluctuations. Their research emphasizes how time series models can be used to forecast and analyze price movements in order to reduce economic risks Chigozirim et al. (2021).

In order to forecast CO_2 emissions in Saudi Arabia, the authors compare forecasting techniques such as ARIMA models, Holt-Winters Exponential Smoothing, and Artificial Neural Networks (ANN). Their research contrasts ANN's capacity to detect nonlinear patterns with ARIMA's proficiency in simulating linear trends in time series data. In order to optimize ARIMA models in forecasting applications, they also stress how crucial it is to choose the right model parameters using statistical metrics like AIC and BIC, Alam & AlArjani (2021).

The current study uses monthly data from January 2000 to June 2013 to forecast South African gold sales using the Box-Jenkins methodology. Following the procedures of model identification, estimation, diagnostic checking, and forecasting, the study concludes that a seasonal ARIMA$(4,1,4)\times(0,1,1)_{12}$ model is the best fit in terms of prediction accuracy, with a mean absolute percentage error (MAPE) of 11%. Their results show how well ARIMA models predict gold sales and emphasize how useful they are for figuring out market trends and organizing operations in resource economics, Tsoku et al. (2017).

Autoregressive Distributed Lag (ARDL) model to examine how certain macroeconomic factors affect Nigeria's economic growth. Their research shows the intricate connections between macroeconomic variables like interest rates, inflation rates, and exchange rates and how they affect economic expansion. The methodology stresses the significance of comprehending how macroeconomic variables can impact economic outcomes, even though their focus is not on gold prices, Adenomon & Ojo (2020).

The market efficiency of gold exchange-traded funds (GETFs) in India, which concentrates on determining whether the market displays weak-form efficiency. They use parametric serial correlation tests, augmented Dickey-Fuller unit root tests, and non-parametric runs tests to assess the randomness of price movements using daily returns data from five GETFs that were traded on the Indian Stock Exchange between March 2010 and August 2015. Their results indicate inefficiencies in the price discovery mechanism, indicating that the Indian GETF market is not weak-form efficient Nargunam & Anuradha (2017).

Using the Markov-Switching ARCH model, examines exchange rate volatility during the COVID-19 pandemic, emphasizing the structural changes and increased uncertainty in financial markets. According to the study, there were noticeable regime shifts during this time, and exchange rate volatility increased dramatically as a result of the pandemic. These results underscore the importance of accounting for regime-switching dynamics and external shocks when forecasting time series. At the same time, exchange rates are the main focus of the study, Koç (2021).

Methodology

Statistics provide numerical estimates that facilitate planning, they are essential for forecasting and decision-making. One important statistical technique is time series analysis, which examines phenomena over time to make minimally inaccurate predictions. It is crucial for sales forecasting and applied fields, requiring stationary data for precise forecasts, where certain statistical characteristics define stationarity.

Time Series

A time series is an assemblage of data points that emerge successively over a given duration. Cross-sectional data, on the other hand, documents a particular point in time. A time series in investing tracks the movement of the chosen data points, like as prices, over a predefined period of time by recording them at regular intervals. There is no need to provide a minimum or maximum duration. A time series can be about any variable that varies over time.

Time Series Data

The goal of the study is to examine the determinants of time series forecasting of daily gold prices in Pakistan. For the empirical analysis of the gold prices, secondary data will be used. The data regarding the daily basis gold prices of Pakistan will be obtained from http://goldpricez.com/gold/history/pkr/months-3 for the period January 1st, 2016, to July 17th, 2022, using the Eviews12 software package for analysis of the data.

Components of Time Series

Many Economic phenomena are influenced by both direct and indirect forces, which throughout time lead to many changes. Time series analysis helps to understand the behavior of these events by looking at how they have changed over time. When analyzing time series data, statisticians find four basic components. The first element is the Trend Component (T), which describes the data's long-term trajectory and displays a steady rise or fall over time. Compared to other components, this change becomes apparent after a longer time. The second is the Seasonal Component (S), which stands for consistent, foreseeable changes that take place at predetermined intervals (quarterly, monthly, etc.) and are frequently brought on by outside variables that have an impact on the data. The third component in a time series is cyclic movements, which are long-term oscillations. These oscillations usually last five to twelve years or more, and most of them are observed in economic data. These oscillations are associated with well-known business cycles. These cyclic movements can be studied if a long set of data free from unpredictable perturbations is available. These are sudden changes that are unlikely to occur again in a time series. Seasonal, cyclic, or trend movements are unable to account for these aspects of a time series. Uneven fluctuations, sometimes known as random or residual components, make up the fourth component. Despite their inadvertent nature, these shifts have the capacity to consistently modify patterns, seasonal variations, and cyclical oscillations in the years to come. Floods, fires, earthquakes, revolutions, epidemics, strikes, and other similar events are the root causes of such irregularities.

Stationary and Non-Stationary Time Series

There are two main kinds of time series data, stationary and non-stationary. There are two types of stationary time series that show fluctuations around a constant mean: weakly stationary time series, in which the mean, variance, and auto covariance are all constant, and strongly stationary time series, in which the joint probability of values does not change over time. Direct analysis of non-stationary time series is challenging because they display cyclical changes or trends that change over time. By using techniques like differencing, which eliminates trends and seasonality, the majority of real world time series can be changed into stationary time series. Techniques like first or second differencing are employed to deal with non-stationarity, and transformations like taking square roots or logarithms can stabilize variance. Seasonal differencing can be used to eliminate seasonal components. Following these modifications, the time series stabilizes, allowing for more precise forecasting and modelling.

Test of Stationary Series

Economic variables are difficult to model since they often follow broad trends, making them non-stationary time series. These time series need to be converted into stationary time series to address this. Stationarity is tested and ensured using a variety of techniques of time series before additional modelling or analysis.

Autocorrelation Function

The autocorrelation function calculates the correlation between neighboring observations in a time series. Its two main purposes are to detect non-randomness in the data and to determine the appropriate time series model when the data is not random. At lag k, the sample's autocorrelation function is calculated.

$$\rho(k) = \frac{\sum_{t=1}^{n-k}\left(Z_t - \bar{Z}\right)\left(Z_{t+k} - \bar{Z}\right)}{\sum_{t=1}^{n}\left(Z_t - \bar{Z}\right)^2} \quad k = 0,1,2,... \tag{1}$$

Where an estimator for (k) is (k). The autocorrelation coefficient between observations from different periods can be calculated using this function, and its value is $-1 \leq (k) \leq 1$; values close to ±1 indicate stronger correlation. The value would be zero in the case of stationarity, meaning that non-autocorrelation coefficients are represented by $(k) = 0$.

$$(k) = \gamma_o \gamma_k \qquad where \;\; k = 0,1,2... \tag{2}$$

Methods and Material

The following methodology will be used for forecasting of Daily Gold Prices of Pakistan.

Models of Time Series in Forecasting

Autoregressive (AR), Moving Average (MA), Mixed (ARMA), and Integrated Mixed (ARIMA) models are time series forecasting techniques that only take into account a variable's historical values, ignoring other explanatory factors. These models are applicable to phenomena that have appropriate time-series information. The study's forecasting will be done using ARMA models, which are the subject of this investigation.

Autoregressive Models

A time series' current value is expressed by the autoregressive model (AR) as a weighted sum of its past values plus a random error. The current value is expressed in terms of weighted past errors by the moving average model (MA). More flexibility in representing time series data is provided by the mixed model (ARMA), which combines the AR and MA models. In order to handle nonstationary time series, ARIMA models are an expansion of ARMA models. AR, MA, and I (integration to make the series stationary) are their three constituent parts. The formula for ARIMA is ARIMA (p, d, q), where p is the AR model's order, q is the MA model's order, and d is the number of differences desired to make the series stationary.

Autoregressive Integrated Moving Average (ARIMA)

The time series must first be transformed into a stationary form in order to use various models efficiently. Usually, differencing is used to accomplish this, in which each value in the series is deducted from the value that came before it $t-1$. A second differencing can be used if the series is still non-stationary following the first one. The "Integrated" (I) component, which denotes the differencing step that aids in achieving stationarity, is an additional component added to the ARIMA model, which operates similarly to the ARMA model. In conclusion, the ARIMA model predicts future observations by combining the lagged values, previous error terms, and the number of differencing steps used to mark the series stationary.

$$Y_t = \varphi_o + \varphi_1 Y_{t-1} + \varphi_2 Y_{t-2} + \ldots + \varphi_p Y_{t-p} + \theta_o + \theta_1 \varepsilon_{t-1} + \theta_2 \varepsilon_{t-2} + \ldots + \theta_p \varepsilon_{t-p} \tag{3}$$

Model Building Stage of Time-Series Data

The following three phases are applied iteratively as part of the Box-Jenkins method:

Step 1: Recognition. A class of basic ARIMA model is chosen based on data plots, autocorrelations, partial autocorrelations, and other information. In essence, this involves determining suitable values for p, d, and q.
Step 2: Making an estimate. According to Box-Jenkins, back casting, maximum likelihood techniques, and other methods are used to estimate the parameters of the chosen model.
Step 3: Examining the diagnosis. By taking into account the autocorrelations of the residual series, the fitted model is examined for shortcomings. Until step three fails to improve the model, these stages are applied iteratively.

Diagnostic Checking

The diagnostic checking of the model is the last stage after it has been fitted. In order to determine whether further structure (high correlation values) can be discovered, the autocorrelation plots of the residuals are examined. The model is deemed acceptable and forecasts are produced if all autocorrelations and partial autocorrelations are modest. The model is re-estimated and the values of p and/or q are modified if some of the autocorrelations are significant.

Best Fitted Models

Mean Square Error (MSE)

The mean square error (MSE) is a more sensitive measure of a substantial forecast error than MAE, and the following formula estimates it:

$$MSE = \sum_{t=1}^{n} (response_t - predicted_t)^2$$

Akaike Information Criterion (AIC)

A mathematical technique for assessing how well a model fits the data it was created from is the Akaike information criterion (AIC). AIC is used in statistics to evaluate potential models and identify the one that best fits the data. The number of independent variables used to construct the model is the basis for calculating AIC. The model's greatest likelihood estimate, or how effectively it replicates the data. The model that uses the fewest number of independent variables to explain the most variation is the best-fit model, according to AIC.

Diagnostic and Graphical Analysis

The last step after fitting a model is to check it for errors. To check for additional structure (high correlation values), the autocorrelation plots of the residuals are examined. The model is deemed sufficient and predictions are produced if all autocorrelations and partial autocorrelations are minimal. The model is re-estimated and the values of p and/or q are modified if some of the autocorrelations are significant.

Results and Discussion

This is about the results and discussion of our research. The study only analyzed a time series of daily Gold Prices in Pakistan. This time series composed of 1385 daily observations from October, 2018 to July, 2022. The time series identify the best fitted ARIMA (p, d, q) model for these data and estimate the model parameters. After that, it is necessary to check the efficiency of this model by using AIC. Then used appropriate ARIMA model for forecasting. In this study statistical software EVIEW 12 is used.

Study of the series stationary

At this stage the time series for original data is drawn to know initially about some Characteristics of this series. Figure 1 represents the characteristics.

Figure 1: Gold Prices in Pakistan (PKR)

The line shows in 1 gold price data from October 2018 to July 2022 in Pakistan show no pattern of stationarity and non-stationarity. It is therefore desirable to draw the autocorrelation function (ACF), and the partial autocorrelation function (PACF) of data and draw confidence interval of (ACF) and (PACF) to detect the stationary or non-stationary of time series.

ACF and PACF of the Gold Price in PKR

Table 1: Correlogram of Gold Price

Included observations: 1388

Autocorrelation	Partial Correlation		AC	PAC	Q-Stat	Prob
		1	0.997	0.997	1381.4	0.000
		2	0.993	0.011	2754.4	0.000
		3	0.990	0.002	4119.1	0.000
		4	0.986	-0.012	5475.4	0.000
		5	0.983	-0.014	6823.0	0.000
		6	0.979	0.012	8162.2	0.000
		7	0.976	-0.006	9493.0	0.000
		8	0.973	0.000	10815.	0.000
		9	0.969	0.004	12130.	0.000
		10	0.966	0.048	13436.	0.000
		11	0.963	0.007	14736.	0.000
		12	0.960	0.011	16029.	0.000
		13	0.957	-0.012	17314.	0.000
		14	0.954	-0.042	18592.	0.000
		15	0.951	0.008	19862.	0.000
		16	0.948	-0.008	21125.	0.000
		17	0.944	0.006	22380.	0.000
		18	0.941	-0.012	23627.	0.000
		19	0.938	-0.008	24866.	0.000
		20	0.934	-0.009	26098.	0.000
		21	0.931	-0.012	27321.	0.000
		22	0.927	-0.007	28535.	0.000
		23	0.924	-0.003	29742.	0.000
		24	0.921	0.001	30941.	0.000
		25	0.917	0.005	32131.	0.000
		26	0.914	-0.021	33314.	0.000
		27	0.910	0.020	34488.	0.000
		28	0.907	-0.012	35654.	0.000
		29	0.903	-0.002	36813.	0.000
		30	0.900	-0.008	37963.	0.000
		31	0.897	0.026	39106.	0.000
		32	0.893	-0.024	40240.	0.000
		33	0.890	0.002	41367.	0.000
		34	0.886	0.022	42487.	0.000
		35	0.883	0.010	43599.	0.000

Table 1 is the Correlogram of Gold Prices. Figure 1 and table 1 indicates random walk behavior. However, the price is showing fluctuation but overall trend is upward. Table 1.1 shows there is high ACFs and PCFs. Through the table 1.1 of the original series of Correlation Coefficients and figures of ACF and PACF, it is noted that there is non-stationary in the data of series. So to transform the gold price series to change it to first difference and tested again for stationarity.

First Difference for Stationarity

Table 2: Shows First Difference For Stationarity

1	-0.053	-0.053	3.8439	0.050
2	-0.012	-0.015	4.0422	0.133
3	0.049	0.048	7.4064	0.060
4	0.017	0.022	7.8197	0.098
5	-0.040	-0.037	10.008	0.075
6	-0.014	-0.020	10.277	0.113
7	-0.021	-0.025	10.865	0.145
8	-0.018	-0.017	11.309	0.185
9	0.003	0.004	11.319	0.254
10	-0.014	-0.013	11.589	0.314
11	-0.051	-0.051	15.189	0.174
12	0.023	0.015	15.912	0.195
13	0.039	0.040	18.078	0.155
14	-0.018	-0.009	18.540	0.183
15	0.017	0.014	18.930	0.217
16	-0.004	-0.012	18.952	0.271
17	0.002	0.000	18.957	0.331
18	-0.009	-0.009	19.069	0.388
19	0.010	0.009	19.208	0.444
20	0.022	0.027	19.902	0.464
21	0.003	0.006	19.914	0.527
22	-0.008	-0.010	19.999	0.583
23	-0.013	-0.015	20.255	0.626
24	-0.056	-0.056	24.686	0.423
25	0.017	0.011	25.086	0.458
26	-0.041	-0.038	27.454	0.386
27	-0.001	0.002	27.455	0.439
28	-0.008	-0.010	27.557	0.488
29	-0.009	-0.011	27.662	0.536
30	0.025	0.026	28.584	0.540
31	-0.000	0.001	28.584	0.591
32	0.041	0.039	30.916	0.521
33	-0.016	-0.019	31.280	0.553
34	0.034	0.027	32.925	0.520
35	0.007	0.003	32.989	0.566
36	-0.014	-0.009	33.255	0.600

Table 2 shows the ACFs and PACFs for first difference data of the series, the ACFs and PACFs are statistically not significant.

Table 3: ADF test results for the original series (Gold Prices in Pakistan)

ADF – Test	t -statistic	Test Critical Values 5%	Prob.
With constant	-39.16281	**−2.863423**	**0.0000**
With constant and trend	**−39.15893**	**−3.413063**	**0.0000**

The root unit test also show that the data is stationary. Depicts Augmented Dickey-Fuller Test; p value for direct values is -39.21308 so is rejected H0 at 5% level of significance. The Gold Price series is stationary at first difference.

Table 4: Comparison of a set of values of AIC, SIC, SIGMASQ

Models	**AR**	**MA**	**SIGMASQ**	**AIC**	**SC**
ARIMA (4, 1, 4)	4	4	6719.548	11.65643	11.67153
ARIMA (2, 1, 2)	2	2	6724.734	11.65720	11.67229
ARIMA (2, 1, 3)	2	3	6712.056	11.65531	11.67041
ARIMA (3, 1, 2)	3	2	6712.566	11.65539	11.67048
ARIMA (3, 1, 4)	3	4	6709.891	11.65499	11.67009
ARIMA (3, 1, 3)	3	3	6711.305	11.65520	11.66085
ARIMA (1, 1, 1)	1	1	6710.079	11.65501	11.67011
ARIMA (1, 1, 3)	1	3	6693.358	11.65252	11.66762

From the table 4, the minimum values for SIGMASQ, AIC and SC are given under the model ARIMA (4, 1, 4). Thus, the ARIMA (4, 1, 4) is the most appropriate model for the Gold Prices in Pakistan.

Variable	Coefficient	Std. Error	t-Statistic	Prob.
C	4.883260	2.231723	2.188112	0.0288
AR(1)	-0.053252	0.019124	-2.784553	0.0054
MA(3)	0.051518	0.027032	1.905801	0.0569
SIGMASQ	6693.358	123.0243	54.40681	0.0000
R-squared	0.005325	Mean dependent var		4.881903
Adjusted R-squared	0.003168	S.D. dependent var		82.06125
S.E. of regression	81.93117	Akaike info criterion		11.65252
Sum squared resid	9283688.	Schwarz criterion		11.66762
Log likelihood	-8077.025	Hannan-Quinn criter.		11.65817
F-statistic	2.468073	Durbin-Watson stat		2.001128
Prob(F-statistic)	0.060530			
Inverted AR Roots		-.05		
Inverted MA Roots	.19+.32i	.19-.32i	zz -.37	

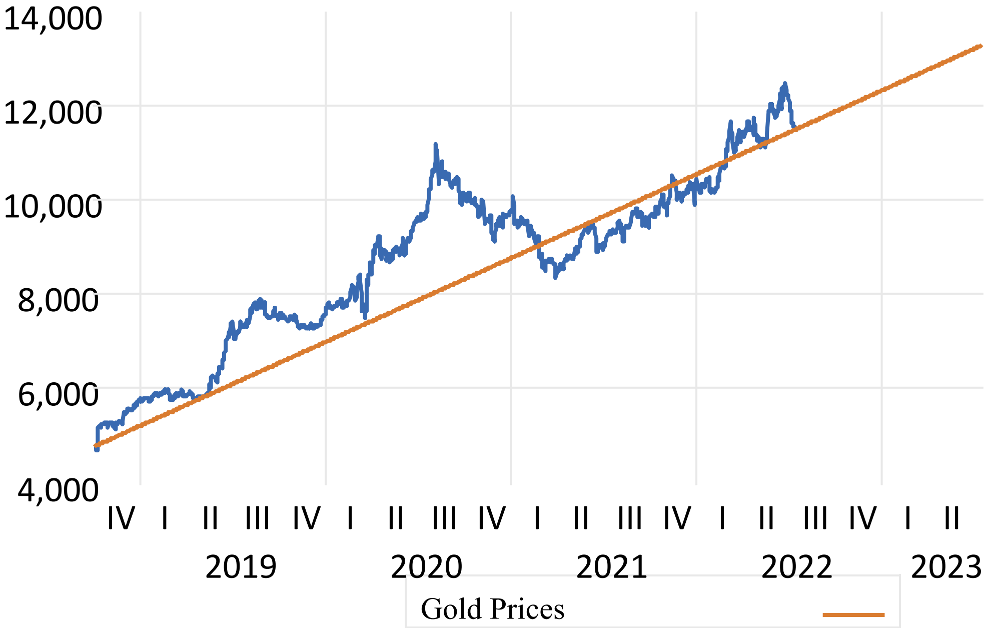
Estimates of model parameters ARIMA (4, 1, 4)

Table 5: Show estimates of model parameters ARIMA (4, 1, 4)

ARIMA (4, 1, 4) is the best model because it has the lowest AIC and SIC. After identifying the most suitable model which is ARIMA (4, 1, 4), we estimate the specific model parameters by Least-squares method.

Combined Graph of Gold Price and Gold Price Forecast

Figure 6: Shows combined graph of Gold Price and Gold Price Forecast in Pakistan



The above figure shows that the data series up to 2022 is the actual data set and the on word is the forecast data for the future which shows rise in the prices.

Table 7: Shows the Actual Prices, Forecast Prices, and Difference

Date	Actual Prices	Forecast Prices	Differen ce	Date	Actual Prices	Forecast Prices	Differen ce
19-04-22	11522	11061.71	-460.29	03-06-22	11766.5	11280.77	-485.73
20-04-22	11522	11066.58	-455.42	04-06-22	11774.4	11285.64	-488.76
21-04-22	11537.5	11071.45	-466.05	05-06-22	11860.7	11290.5	-570.2
22-04-22	11367.2	11076.32	-290.88	06-06-22	11737.4	11295.37	-442.03
23-04-22	11661.9	11081.19	-580.71	07-06-22	11773.8	11300.24	-473.56
24-04-22	11719.4	11086.05	-633.35	08-06-22	11773.8	11305.11	-468.69
25-04-22	11631.2	11090.92	-540.28	09-06-22	11792.3	11309.98	-482.32
26-04-22	11579.7	11095.79	-483.91	10-06-22	11865.6	11314.84	-550.76
27-04-22	11579.7	11100.66	-479.04	11-06-22	11906.4	11319.71	-586.69
28-04-22	11389.6	11105.53	-284.07	12-06-22	11982.8	11324.58	-658.22
29-04-22	11343.6	11110.39	-233.21	13-06-22	12140.2	11329.45	-810.75
30-04-22	11263.8	11115.26	-148.54	14-06-22	12134.3	11334.32	-799.98
01-05-22	11310.5	11120.13	-190.37	15-06-22	12236.4	11339.18	-897.22

Table 7 shows the three months' comparisons between the actual data and the forecast data and also gives the difference between the Actual and forecast.

Summary and Conclusion

The results of using ARIMA models to analyze Pakistan's daily gold prices are shown in this chapter. The forecast covers the time frame from 2018 to 2022. There is no discernible pattern of stationarity in the gold price data from October 2018 to July 2022. The autocorrelation function (ACF) and partial autocorrelation function (PACF) were employed to evaluate this. A random walk pattern in the data is displayed in Table 1 and Figure 1. The gold price data shows fluctuations but an overall upward trend. Table 1 shows high ACF and PACF values, indicating non-stationarity. After applying first differencing, Table 2 shows that the ACF and PACF are now stationary, and the Augmented Dickey-Fuller test confirms stationarity with a p-value of -39.21308. The best model based on the lowest AIC and SIC values is ARIMA (4, 1, 4), as shown in Table 4. This model was estimated using the Least Squares method. Figure 1.6 presents actual data up to 2022 and forecasted values afterwards, showing a rising trend. Table 7 compares actual and forecast data for three months, along with the forecast errors.

Further, the estimated coefficients for the best-fit ARIMA (4,1,4) model show that the constant term and the AR(1) term are statistically significant at the 5% level, while the MA(3) term is marginally significant. The model yields a reasonable Durbin-Watson statistic of approximately 2.00, indicating no autocorrelation in the residuals. In a differenced time series, a high R-squared value is not expected, so its relatively low value is not a concern. Visually comparing historical gold prices with the forecasted values generated by the ARIMA (4,1,4) model reveals that the forecasted prices align closely with the actual trend, supporting the model's forecasting ability. The forecast suggests a continued upward trend in gold prices in Pakistan. However, it consistently underestimates actual prices by 300 to 900 PKR. While the forecast accurately captures the direction and trend, the forecast errors tend to increase over time, which is common in time series forecasting. These discrepancies emphasize the model's limitations and indicate potential opportunities for improvement, such as exploring hybrid modeling techniques like ARIMA-GARCH or ARIMA-ANN.

Concluding that Gold prices in Pakistan showed an overall rising trend from 2018 to 2022, and forecasts indicate that this upward trend is likely to continue, reflecting increasing demand and economic factors influencing gold as a safe investment.

References

[1]. Shafie, N. A., Rahman, H. A. A., Azizi, M. F. R., & Saruand, M. F. (2022). Forecasting of Malaysia gold price with exponential smoothing. *Journal of science and technology*, 14(1), 44–50.

[2]. Qasim, T. B., Iqbal, G. Z., Hassan, M. U., & Ali, H. (2021). Application of Markov Regime Switching Autoregressive Model to Gold Prices in Pakistan. *Review of Economics and Development Studies*, *7*(3), 309-323.

[3]. Yaziz, S. R., & Zakaria, R. (2019, November). ARIMA and Symmetric GARCH-type Models in Forecasting Malaysia Gold Price. In *Journal of Physics: Conference Series* (Vol. 1366, No. 1, p. 012126). IOP Publishing.

[4]. Makala, D., & Li, Z. (2021, February). Prediction of gold price with ARIMA and SVM. In *Journal of Physics: Conference Series* (Vol. 1767, No. 1, p. 012022). IOP Publishing.

[5]. Xiang, Y. (2022). Using ARIMA-GARCH Model to Analyze Fluctuation Law of International Oil Price. *Mathematical Problems in Engineering*, *2022*.

[6]. Singh, K., & Kumar, A. (2017). Autoregressive Integrated Moving Average Model for Gold Price Forecasting: Evidence from the Indian Market. *Indian Journal of Research in Capital Markets*, *4*(3), 33.

[7]. Tripathy, N. (2017). Forecasting the gold price with auto regressive integrated moving average model. *International Journal of Economics and Financial Issues,* 7(4), 324-329.

[8]. Yakean, S. (2022). Determinants of Gold Price in Thailand. *IUP Journal of Applied Finance*, *28*(1).

[9]. Azad, A. S., Sokkalingam, R., Daud, H., Adhikary, S. K., Khurshid, H., Mazlan, S. N. A., & Rabbani, M. B. A. (2022). Water Level Prediction through Hybrid SARIMA and ANN Models Based on Time Series Analysis: Red Hills Reservoir Case Study. *Sustainability*, *14*(3), 1843.

[10]. Wang, Y., Jia, R., Dai, F., & Ye, Y. (2022). Traffic Flow Prediction Method Based on Seasonal Characteristics and SARIMA-NAR Model. *Applied Sciences*, *12*(4), 2190.

[11]. Ismail, M. T., & Al-Gounmeein, R. S. (2022). Overview of Long Memory for Economic and Financial Time Series Dataset and Related Time Series Models: A Review Study. *IAENG International Journal of Applied Mathematics*, *52*(2), 1-9.

[12]. Deretić, N., Stanimirović, D., Awadh, M. A., Vujanović, N., & Djukić, A. (2022). SARIMA modelling approach for forecasting of traffic accidents. *Sustainability*, *14*(8), 4403.

[13]. Aimer, N. M., & Lusta, A. A. (2021). Exchange rates and oil price under uncertainty and regime switching: A Markov-switching VAR approach. *Economic Journal of Emerging Markets,* 200-215.

[14]. Chigozirim, O. N., Okore, N. P., Ukeh, O. O., & Mba, A. N. (2021). Dynamics of Food Price Volatility and Households' Welfare in Nigeria. Agris *On-Line Papers in Economics and Informatics,* 13(4), 49–60.

[15]. Alam, T., & AlArjani, A. (2021, September). Forecasting CO 2 Emissions in Saudi Arabia Using Artificial Neural Network, Holt-Winters Exponential Smoothing, and Autoregressive Integrated Moving Average Models. *In 2021 International Conference on Technology and Policy in Energy and Electric Power (ICT-PEP)* (pp. 125-129). IEEE.

[16]. Tsoku, J. T., Phukuntsi, N., & Metsileng, D. (2017). Gold sales forecasting: The Box-Jenkins methodology. *Risk Governance & Control: Financial Markets & Institutions*, 7(1), 54–60.

[17]. Adenomon, M. O., & Ojo, R. O. (2020). Autoregressive Distributed Lag Modeling of the Effects of Some Macroeconomic Variables on Economic Growth in Nigeria. *Folia Oeconomica Stetinensia*, 20(2), 1–19.

[18]. Nargunam, R., & Anuradha, N. (2017). Market efficiency of gold exchange-traded funds in India. *Financial Innovation,* 3(1).

[19]. Koç, H. (2021). Exchange Rate Volatility in the Covid-19 Period: An Analysis Using the Markov-Switching ARCH Model. *Ekoist: Journal of Econometrics and Statistics*, *0*(35), 205–220.